CT Meeting 2020 Proceedings

The 6th International Meeting on

Image Formation in X-Ray Computed Tomography

August 3 - August 7, 2020, Regensburg (virtual only), Germany

www.ct-meeting.org

Conference Chair:

Marc Kachelrieß (German Cancer Research Center, DKFZ)

DECTRIS detecting the future

SIEMENS Healthineers

varian

ziehm imaging

The 6th International Conference on Image Formation in X-Ray Computed Tomography is supported by:

Dectris

Siemens Healthineers

Varian Medical Systems

Ziehm Imaging

We thank our sponsors for their important and valuable contributions!

We would like to thank the members of the scientific committee for their great work by taking the opportunity to mention them in alphabetical order:

Scientific Committee

Name	**Institution**
Adam Wang	Stanford University, USA
Bruno De Man	GE Research, USA
Cristian T. Badea	Duke University, USA
Cyril Riddell	GE Healthcare, France
Emil Y. Sidky	University of Chicago, USA
Frédéric Noo	University of Utah School of Medicine, USA
Ge Wang	Rensselaer Polytechnic Institute, USA
Guang-Hong Chen	University of Wisconsin-Madison, USA
Günter Lauritsch	Siemens Healthineers, Germany
Hengyong Yu	University of Massachusetts Lowell, USA
Hiroyuki Kudo	University of Tsukuba, Japan
Hongbing Lu	Fourth Military Medical University, China
Ioannis Sechopoulos	Radboud University Medical Center, The Netherlands
J. Webster Stayman	Johns Hopkins School of Medicine, USA
Jeffrey A. Fessler	University of Michigan, USA
Jeffrey H. Siewerdsen	Johns Hopkins University School of Medicine, USA
Jerome Liang	Stony Brook University, USA
Jing Wang	University of Texas Southwestern Medical Center, USA
Jingyan Xu	Johns Hopkins University, USA
Johan Nuyts	KU Leuven, Belgium
John Boone	University of California, Davis, USA
Jongduk Baek	Yonsei University, South Korea
Karl Stierstorfer	Siemens Healthineers, Germany
Katsuyuki Taguchi	Johns Hopkins University School of Medicine, USA
Ke Li	University of Wisconsin-Madison, USA
Ken Sauer	University of Notre Dame, USA

Kirsten Boedeker	Canon Medical Systems Corporation
Lifeng Yu	Mayo Clinic, USA
Mats Danielsson	KTH Royal Institute of Technology, Sweden
Michael McNitt-Gray	David Geffen School of Medicine at UCLA, USA
Norbert Pelc	Stanford University, USA
Pascal Paysan	Varian Medical Systems Imaging Lab, Switzerland
Peter B. Noël	University of Pennsylvania, USA
Rongping Zeng	U.S. Food and Drug Administration, USA
Scott Hsieh	Mayo Clinic, USA
Srinivasan Vedantham	University of Arizona, USA
Stanislav Žabić	UIH America, Inc., USA
Stefan Sawall	German Cancer Research Center, Germany
Taly Gilat-Schmidt	Marquette University, USA
Thomas Flohr	Siemens Healthineers, Germany
Thomas Koehler	Philips Research Hamburg, Germany
Thomas Koenig	Ziehm Imaging, Germany
Tobias Lasser	Technical University of Munich, Germany
Wenli Wang	Avant Tomography Consulting LLC, USA
Xiaochuan Pan	University of Chicago, USA
Xun Jia	University of Texas Southwestern Medical Center, USA
Yuxiang Xing	Tsinghua University, China
Zhicong Yu	Accuray, USA
Zhou Yu	Canon Medical Research Institute, USA
Zhye Yin	GE Research, USA

Organization Committee

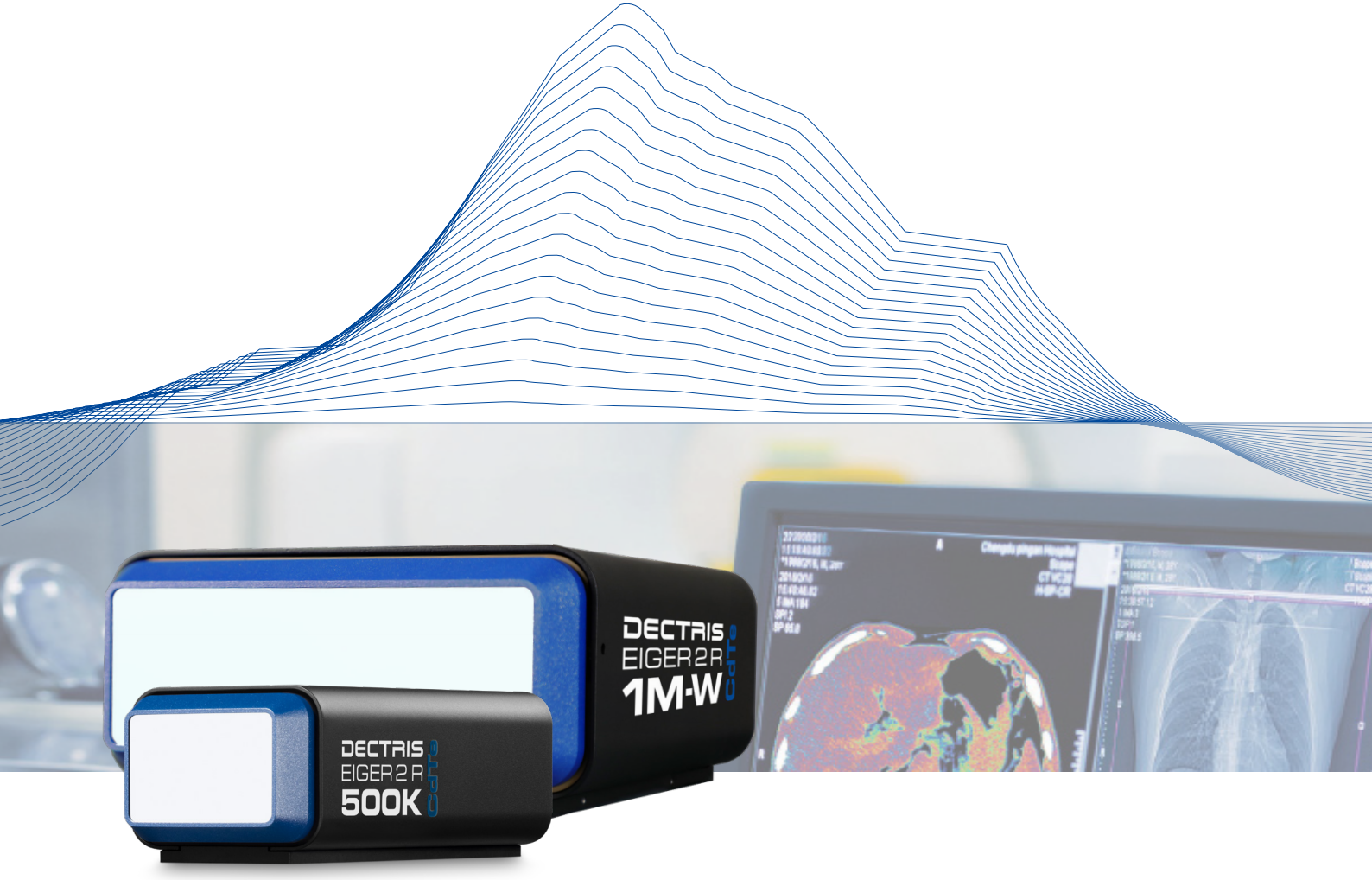
Name	Institute
Andrea Kachelrieß	SHPCI
Carlo Amato	DKFZ
Jan Kuntz	DKFZ, SHPCI
Joscha Maier	DKFZ, SHPCI
Julien Erath	DKFZ, SHPCI
Laura Klein	DKFZ
Marc Kachelrieß	DKFZ, SHPCI
Markus Susenburger	DKFZ, SHPCI
Philip Trapp	DKFZ, SHPCI
Stefan Sawall	DKFZ, SHPCI
Tim Vöth	DKFZ, SHPCI

DKFZ = German Cancer Research Center
SHPCI = Society of High Performance Computational Imaging

Schedule CT Meeting 2020

	Monday *August 3*	**Tuesday** *August 4*	**Wednesday** *August 5*	**Thursday** *August 6*	**Friday** *August 7*
10:00 - 11:40 CEST	**Test Session**	**Image Reconstruction** Opening remarks + 5 talks *(Page 30)*	**Deep Learning** Basic short course: Deep Learning + 3 talks *(Page 224)*	**Photon Counting** Basic short course: Photon Counting + 3 talks *(Page 422)*	
		Coffee	**Coffee**	**Coffee**	
12:00 - 13:40 CEST		**Dual and Multi-Energy Imaging** 5 talks *(Page 50)*	**Motion Estimation and Compensation** 5 talks *(Page 236)*	**Breast CT** Introduction: Breast CT + 4 talks *(Page 434)*	**Observer** Basic short course: Observer + 3 talks *(Page 618)*
		Lunch	**Lunch**	**Lunch**	**Lunch**
14:40 - 16:20 CEST		**Poster Session 1** Poster FF + 35 posters *(Page 70)*	**Poster Session 2** Poster FF + 36 posters *(Page 256)*	**Poster Session 3** Poster FF + 36 posters *(Page 450)*	**Novel CT Technologies** 4 talks + Fully 3D 2021 & CT Meeting 2022 previews *(Page 630)*
		Coffee	**Coffee**	**Coffee**	
16:40 - 18:20 CEST	**Test Session**	**Diagnostic and Interventional** Plenary: Intelligent clinical CT + 3 talks *(Page 210)*	**Artifact and Noise Reduction** 5 talks *(Page 400)*	**Security and Phase Contrast CT** 5 talks *(Page 594)*	

DECTRIS
detecting the future



EIGER2 R CdTe

Detector Series

Join us for the new era of spectral computed tomography

- Hybrid Photon Counting Technology
- Spectral imaging with multiple thresholds
- High resolution with 75 µm pixel size
- High absorption efficiency up to 120 keV with CdTe sensor
- High count rate up to 10^9 photons/sec/mm^2
- Excellent count rate stability

sales@dectris.com | www.dectris.com

Tuesday, August 4

Oral Session : Image Reconstruction
Time : Tuesday, 09:55 – 11:40
Chairs : J. Webster Stayman and Xiaochuan Pan

Time	Author	Title	Page
09:55 – 10:00	Marc Kachelrieß	*Opening Remarks*	
10:00 – 10:20	Baiyu Chen, Erich Kobler, Aaron Sodickson, Daniel K. Sodickson, Ricardo Otazo	*Benefits of Distributed Multidimensional Undersampling for Compressed Sensing CT*	30
10:20 – 10:40	Jingyan Xu, Frédéric Noo	*A k-Nearest Neighbor Regularizer for Model Based CT Reconstruction*	34
10:40 – 11:00	Yongyi Shi, Yongfeng Gao, Jiaxing Tan, Siming Lu, Xuanqin Mou, Zhengrong Liang	*Learned Texture Prior for Bayesian Reconstruction of Super-Resolution CT*	38
11:00 – 11:20	Xiaojuan Deng, Xing Zhao, Hongwei Li	*Limited-Angle CT Reconstruction with Random Walk Regularization*	42
11:20 – 11:40	Nicolas Gindrier, Rolf Clackdoyle, Simon Rit, Laurent Desbat	*Cone-Beam Reconstruction from n-sin Trajectories with Transversely-Truncated Projections*	46

Coffee Break

Oral Session : Dual and Multi-Energy Imaging
Time : Tuesday, 12:00 – 13:40
Chairs : Katsuyuki Taguchi and Thomas Koenig

Time	Author	Title	Page
12:00 – 12:20	Sen Wang, Yuxiang Xing, Hewei Gao, Marc Kachelrieß, Ao Zheng, Li Zhang	*Experimental Feasibility of a Residual-Modelled Material Decomposition Method for Spectral CT*	50
12:20 – 12:40	Weibin Zhang, Shusen Zhao, Huiying Pan, Xing Zhao	*An Iterative Reconstruction Method for Material Decomposition of Dual Spectral Computed Tomography Based on Monochromatic Images*	54
12:40 – 13:00	Pierre-Antoine Rodesch, Salim Si-Mohamed, Simon Rit	*Spatially Varying Regularization Weights for One-Step Spectral CT with SQS*	58
13:00 – 13:20	Yiqun Q. Ma, Wenying Wang, Matthew Tivnan, Junyuan Li, Minghui Lu, Jin Zhang, Josh Star-Lack, Richard E. Colbeth, Wojciech Zbijewski, J. Webster Stayman	*High-Resolution Model-Based Material Decomposition for Multi-Layer Flat-Panel Detectors*	62
13:20 – 13:40	Yifan Deng, Hewei Gao	*Triple-Energy X-Ray CT Using Spectral Modulator with Flying Focal Spot: Modulator Design and Scatter-Modeled Material Decomposition*	66

Lunch Break

Poster Session : Poster Session 1
Time : Tuesday, 14:40 – 16:20

Time	Author	Title	Page
14:40 – 15:00	**Poster Fast Forward**: 30 s presentation for each poster		
Poster Session 1a (Chair: Rongping Zeng and Jongduk Baek)			
15:00 – 16:20	Daxin Shi	*Data-Domain Decomposition Strategy in Short-Scan Configurations for 2D Dual-Energy CT*	70
15:00 – 16:20	Gengsheng L. Zeng	*A Pre-Filter that Incorporates the Noise Model*	74
15:00 – 16:20	Hao Zhao, Yanyan Liu, Guotao Quan	*Cone Beam Artifacts Correction in Multi-detector Computed Tomography Using Deep Neural Networks*	78
15:00 – 16:20	Guotao Quan, Jiao Tian, Yi Wang	*Cardiac Motion Correction of Computed Tomography (CT) with Spatial Transformer Network*	82
15:00 – 16:20	Daniël M. Pelt, Charlotte Maughan-Jones, Oriol Roche i Morgó, Alessandro Olivo, Charlotte K. Hagen	*Rapid and Flexible High-Resolution Scanning Enabled by Cycloidal Computed Tomography and Convolutional Neural Network (CNN) Based Data Recovery*	86
Poster Session 1b (Chair: Bruno De Man and Zhicong Yu)			
15:00 – 16:20	Joaquim G. Sanctorum, Sam Van Wassenbergh, Van Nguyen, Jan De Beenhouwer, Jan Sijbers, Joris J.J. Dirckx	*Projection-Angle-Dependent Image Intensifier Distortion Correction in High-Speed Tomography*	90
15:00 – 16:20	Jochen Butzer, Annika Döring, Sven Gondrom-Linke, Markus Harbecke	*Dual-Energy in Industrial CT – How Simple Math Can Reveal Structures Hidden by Artefacts*	94
15:00 – 16:20	Genwei Ma, Xing Zhao, Shusen Zhao	*Dual Domain Learning Basis Material Decomposition for Dual-Spectral CT via Attention Mechanism*	98
15:00 – 16:20	Yan Ren, Wenting Long, Huiqiao Xie, Xiangyang Tang	*Improving the Accuracy and Dose Efficiency of Photon-Counting Spectral CT via Source Spectrum Shaping*	102
15:00 – 16:20	Marina Eckermann, Mareike Töpperwien, Anna-Lena Robisch, Franziska van der Meer, Christine Stadelmann, Tim Salditt	*Phase-Contrast X-Ray Tomography of Neuronal Tissue with Laboratory Sources at Submicron Resolution*	106

Poster Session 1c (Chair: Adam Wang and Elias Eulig)

15:00 – 16:20	Jinwoo Kim, Ho Kyung Kim	*Dose Reconstruction in Dental CBCT*	110
15:00 – 16:20	Jeonghyeon Nam, Okkyun Lee	*Fast and Accurate Photon-Counting Detector-Based Material Decomposition via Nearest Neighborhood Method*	114
15:00 – 16:20	Cristóbal Martínez, Jeffrey A. Fessler, Manuel Desco, Mónica Abella	*Segmentation-Free Statistical Method for Polyenergetic X-Ray Computed Tomography with a Calibration Step*	118
15:00 – 16:20	Alexander Katsevich, Seongjin Yoon, Michael Frenkel, Peter Munro, Pascal Paysan, Igor Peterlik, Dieter Seghers, Adam Strzelecki	*A Motion Estimation and Compensation Algorithm for 4D CBCT with Cyclic Deformation Model*	122
15:00 – 16:20	Kihong Son, Doil Kim, Seokhwan Jang, Kyoung-Yong Lee	*Capability of a Novel Head Motion Correction Algorithm for X-Ray CT Imaging*	126

Poster Session 1d (Chair: Hengyong Yu and Ioannis Sechopoulos)

15:00 – 16:20	Junyoung Park, Jiyoung Choi, Duhgoon Lee, Minkook Cho, Kyoung-Yong Lee	*Performance of a Novel Cardiac Motion Correction in CT scans of 0.33 Seconds Rotation*	130
15:00 – 16:20	Danyang Li, Dong Zeng, Sui Li, Zhaoying Bian, Jing Huang, Jianhua Ma	*High-Fidelity CT Reconstruction Based on an Unsupervised Gaussian Mixture Model*	134
15:00 – 16:20	Sui Li, Manman Zhu, Danyang Li, Qi Gao, Zhaoying Bian, Dong Zeng, Jianhua Ma	*Self-Supervised Reconstruction of Cerebral Perfusion CT Images Based on Noise Properties*	138
15:00 – 16:20	Stefan Haninger, Matthias Wieczorek, Wolfgang Wein, Tobias Lasser	*Deep Learning Acceleration of OS-SIRT in X-Ray Cone-Beam Computed Tomography*	142
15:00 – 16:20	Markus Susenburger, Pascal Paysan, Adam Strzelecki, Igor Peterlik, Dieter Seghers, Marc Kachelrieß	*4D Segmentation-Based Anatomy-Constrained Motion-Compensated Reconstruction of On-Board 4D CBCT Scans*	146

Poster Session 1e (Chair: Cyril Riddell and Fréderic Noo)

15:00 – 16:20	Eri Haneda, Bernhard Claus, Jed D. Pack, Bruno De Man	*Projection-Based Cardiac Phase Estimation: Towards ECG-less Prospective Cardiac CT Gating*	150
15:00 – 16:20	Jens Renders, Jan Sijbers, Jan De Beenhouwer	*Adjoint Pairs of Image Warping Operators for Motion Modeling in 4D-CT*	154
15:00 – 16:20	Stanislav Žabić, Liuchun He, Yuan Bao	*A Method for Reduction of Axial Truncation Artifacts*	158
15:00 – 16:20	Oriol Roche i Morgó, Fabio Vittoria, Marco Endrizzi, Alessandro Olivo, Charlotte K. Hagen	*Three Different Ways of Implementing Cycloidal Computed Tomography: a Discussion of Pros and Cons*	162
15:00 – 16:20	Theodor Cheslerean-Boghiu, Franz Pfeiffer, Tobias Lasser	*Fast Task-Driven Acquisition Optimization for Anisotropic X-Ray Dark-field Tomography*	166

Poster Session 1f (Chair: Tobias Lasser and Guang-Hong Chen)

15:00 – 16:20	Nadav Shapira, Rajarshi Chattaraj, Leening P. Liu, Michael Geagan, Daeyeon Lee, Chandra Sehgal, Peter B. Noël	*Quantification of Lipid-Shelled Xenon Microbubbles with Dual-Energy CT*	170
15:00 – 16:20	Mayank Patwari, Ralf Gutjahr, Rainer Raupach, Andreas K. Maier	*Low Dose CT Denoising via Joint Bilateral Filtering and Intelligent Parameter Optimization*	174
15:00 – 16:20	Tobias Funk, Daniel S. Badali, Oleg Konings	*GPU Accelerated Monte Carlo Simulations of X-Rays through CAD Objects*	178
15:00 – 16:20	Matt D. Holbrook, Darin P. Clark, Cristian T. Badea	*Deep Learning Based Distortion Correction and Material Decomposition for Photon Counting CT: A Simulation Study*	182
15:00 – 16:20	Seongjin Yoon, Alexander Katsevich, Michael Frenkel, Ed Morton, William Thompson	*Reduction of Artifacts in Multi-Energy Imaging for a Stationary Gantry CT Scanner*	186

Poster Session 1g (Chair: Jing Wang and Taly Gilat-Schmidt)

15:00 – 16:20	Chengpeng Wu, Hewei Gao, Yuxiang Xing, Zhiqiang Chen, Xiaohua Zhu, Li Zhang	*Flux Adaptation for Task-Based Dose Optimization in Grating-based X-Ray Imaging*	190
15:00 – 16:20	John Boone, Ali Uneri, Jeffrey H. Siewerdsen	*A Modular Phantom for Cone Beam and Whole-Body CT Image Quality Assessment*	194
15:00 – 16:20	Taejun Jang, Sungmin Lee, Kangcheol Kim, Hyuncheol Cho, Hyesun Yun, Jin Keun Seo	*Deep Learning-Based Dental Segmentation and Tooth Identification in Dental Cone-Beam CT*	198
15:00 – 16:20	Kaicong Sun, Zhaowen Gong, Sven Simon	*FDRN: Fast Deformable Image Registration Based on Unsupervised Convolutional Neural Networks*	202
15:00 – 16:20	Hyoung Suk Park, Kiwan Jeon	*Low-Dose CT Image Reconstruction with a Deep Learning Prior*	206

Coffee Break

Oral Session : Diagnostic and Interventional
Time : Tuesday, 16:40 – 18:20
Chairs : Frédéric Noo and Norbert Pelc

Time	Author	Title	Page
16:40 – 17:20	Matthias May	*Plenary lecture: Intelligent Clinical CT*	
17:20 – 17:40	Chumin Zhao, Thomas Weber, Magdalena Herbst, Ludwig Ritschl, Sebastian Vogt, Steffen Kappler, Jeffrey H. Siewerdsen, Wojciech Zbijewski	*Advanced Scan Trajectories for Chest Tomosynthesis on Robotic X-Ray Systems*	210
17:40 – 18:00	Sascha Manuel Huck, George S.K. Fung, Katia Parodi, Karl Stierstorfer	*Fluence Field Modulation in X-Ray CT for ROI Imaging: Optimization of Primary Fluence and Comparison of Attenuator Scenarios*	214
18:00 – 18:20	S. Principi, Adam Wang, A. Maslowski, T. Wareing, P. Jordan, Taly Gilat-Schmidt	*Deterministic Boltzmann Transport Equation Solver for Patient-Specific CT Dose Estimation: Validation for Realistic Scanner Configurations and Patient Models*	218

Get two steps ahead with Dual Source CT

SOMATOM Force

Get two steps ahead in clinical excellence
At the top of our Dual Source CT portfolio, SOMATOM Force enables new levels of image quality, clinical outcomes, and ultimately precision medicine. Examine patients without beta-blockers, with no need for them to hold their breath, and with the lowest possible amount of contrast media. Make clearly quantified therapy evaluations with dose-neutral Dual Energy.

Get two steps ahead in workflow performance
Automated technologies support safe, standardized, and highly performant workflows – allowing for appropriate dose and reproducible precision from the smallest to the tallest patients.

Get two steps ahead in expert leadership
Thinking beyond today, you're connected to the future with an ever-growing expert community and exclusive access to our advanced research environment.

siemens-healthineers.com/somatom-force

SIEMENS Healthineers

The scientific overlay is not that of the individual pictured and is not from a device of Siemens Healthineers. It was modified for better visualization.

International version. Not for distribution or use in the U.S. · 7515 0519

Wednesday, August 5

Oral Session : Deep Learning
Time : Wednesday, 10:00 – 11:40
Chairs : Hongbing Lu and Xun Jia

Time	Author	Title	Page
10:00 – 10:40	Yuxiang Xing	*Basic short course: Deep Learning*	
10:40 – 11:00	Rongping Zeng, Claire Yilin Lin, Qin Li, Jiang Lu, Jeffrey A. Fessler, Kyle J. Myers	*Generalizability Test of a Deep Learning-Based CT Image Denoising Method*	224
11:00 – 11:20	Xikai Yang, Xuehang Zheng, Yong Long, Saiprasad Ravishankar	*Learned Multi-Layer Residual Sparsifying Transform Model for Low-Dose CT Reconstruction*	228
11:20 – 11:40	Muge Du, Hewei Gao, Kaichao Liang, Yinong Liu, Yuxiang Xing	*Unsupervised Domain Adaptation for Practical Metal Artefact Reduction in X-Ray CT*	232

Coffee Break

Oral Session : Motion Estimation and Compensation
Time : Wednesday, 12:00 – 13:40
Chairs : Cristian T. Badea and Pascal Paysan

Time	Author	Title	Page
12:00 – 12:20	A. Sisniega, S. Capostagno, Wojciech Zbijewski, J. Webster Stayman, Clifford R. Weiss, T. Ehtiati, Jeffrey H. Siewerdsen	*Local Motion Estimation for Improved Cone-Beam CT Deformable Motion Compensation*	236
12:20 – 12:40	Igor Peterlik, Adam Strzelecki, Mathias Lehmann, Philippe Messmer, Peter Munro, Pascal Paysan, Mathieu Plamondon, Dieter Seghers	*Motion Resilient Iterative 3D CBCT Reconstruction Based on Gradient Weighting*	240
12:40 – 13:00	Rob Heylen, Johan Nuyts	*Motion Correction for 4D CBCT Reconstruction with TV Regularization*	244
13:00 – 13:20	Johan Nuyts, Roger Fulton	*Iterative FDK Reconstruction for Helical CT of the Head with Rigid Motion Compensation*	248
13:20 – 13:40	Joscha Maier, Sergej Lebedev, Elias Eulig, Stefan Sawall, Eric Fournié, Karl Stierstorfer, Marc Kachelrieß	*Coronary Artery Motion Compensation for Short-Scan Cardiac CT Using a Spatial Transformer Network*	252

Lunch Break

Poster Session : Poster Session 2
Time : Wednesday, 14:40 – 16:20

Time	Author	Title	Page
14:40 – 15:00	**Poster Fast Forward**: 30 s presentation for each poster		
Poster Session 2a (Chair: Scott Hsieh and Karl Stierstorfer)			
15:00 – 16:20	Achim Byl, Laura Klein, Jennifer Hardt, Stefan Sawall, Heinz-Peter Schlemmer, Sarah Heinze, Marc Kachelrieß	*Metal Artifact Reduction in Photon Counting CT using Pseudo-Monochromatic Images*	256
15:00 – 16:20	Wenchao Cao, Peter B. Noël	*Slow Triple kVp Switching CT with Convolutional Neural Network Based Sinogram Completion and Material Decomposition*	260
15:00 – 16:20	Tobias Würfl, Nicole Maaß, Frank Dennerlein, Andreas K. Maier	*Projective Reconstruction of Acquisition Geometry without Point Correspondences*	264
15:00 – 16:20	Yidi Yao, Liang Li, Zhiqiang Chen, Wuyang Liang, Xin Jin	*Multi-Energy Monte Carlo Simulation of the Multi-Sources Static CT System*	268
15:00 – 16:20	Frédéric Jolivet, Clarisse Fournier, Jérôme Lesaint, Andrea Brambilla	*A One-Step Algorithm for Spectral CT with an Application on Multi-Source Inverse Geometry*	272
Poster Session 2b (Chair: Jeffrey H. Siewerdsen and Zhou Yu)			
15:00 – 16:20	Tristan M. Gottschalk, Björn W. Kreher, Andreas K. Maier	*Deep Learning Based Metal Inpainting in the Projection Domain using Additional Neighboring Projection Information*	276
15:00 – 16:20	Carlo Amato, Laura Klein, Joscha Maier, Stefan Sawall, Nicole Gehrke, Danielle Franke, Spyridon Gkoumas, Thomas Thüring, Andreas Briel, Christian Brönnimann, Marc Kachelrieß	*Decomposition of Multiple Contrast Agents in Photon-Counting Micro-CT: Iodine and Bismuth*	280
15:00 – 16:20	Darin P. Clark, Cristian T. Badea	*Convolutional Regularization for Accelerated Spectrotemporal CT Reconstruction*	284
15:00 – 16:20	Zhitao Guo, Hengyong Yu	*Low-Dose CT Denoising with Convolutional Neural Network for Unknown Noise Levels*	288
15:00 – 16:20	T. W. Holmes, C. N. De Cecco, A. Pourmorteza	*Investigating the Utility of Virtual Monoenergetic Images in Non-Contrast Lung CT Scans*	292

Poster Session 2c (Chair: Ke Li and Joscha Maier)

15:00 – 16:20	Zhengrong Liang, Yongfeng Gao, Yongyi Shi, Wei Hou	*An Investigation on Bayesian Image Reconstruction without Freely Adjustable Parameter*	296
15:00 – 16:20	Kai Mei, Nadav Shapira, Joshua Scheuerman, Janet Reddin, Amy E. Perkins, Joel S. Karp, Peter B. Noël	*Ultra-Low Dose CT for PET Attenuation Correction: Preliminary Study*	300
15:00 – 16:20	Kaichao Liang, Li Zhang, Le Shen, Zhiqiang Chen, Yuxiang Xing	*Method of Deep Distribution Learning from Noisy Samples and an Application for Low Dose CT*	304
15:00 – 16:20	Berk Iskender, Yoram Bresler	*X-Ray CT Scatter Correction by a Physics-Motivated DNN with Opposite View Processing*	308
15:00 – 16:20	Gilad Shechter	*Projection Domain Decomposition for Photon Counting CT*	312

Poster Session 2d (Chair: Ken Sauer and J. Webster Stayman)

15:00 – 16:20	Xiao Jiang, Lei Zhu	*Filtered Backprojection Reconstruction for Inverse Geometry CT with Sparsely Distributed Sources*	316
15:00 – 16:20	Donghyeon Lee, Hwanhui Jo, Uijin Jeong, Seungryong Cho	*Single-Scan Dual-Energy CT Image Reconstruction via Normalized Gradient Magnitude Image (NGMI) Regularization*	320
15:00 – 16:20	Tao Zhang, Hewei Gao, Yuxiang Xing, Li Zhang, Zhiqiang Chen	*Prior-Knowledge Based Linogram Reconstruction Neural Network for X-Ray Computed Tomography*	324
15:00 – 16:20	Nargiza Djurabekova, Andrew Goldberg, David Hawkes, Guy Long, Luke Shepherd, Marta M. Betcke	*Segmented Static CBCT as Initialization of Fully Dynamic Tomography Reconstruction of Foot and Ankle*	328
15:00 – 16:20	Erich Kobler, Baiyu Chen, Alexander Effland, Thomas Pock, Daniel K. Sodickson, Ricardo Otazo	*Total Deep Variation for SparseCT Reconstruction*	332

Poster Session 2e (Chair: Hongbing Lu and Katsuyuki Taguchi)

15:00 – 16:20	Erdal Pekel, Martin Dierolf, Franz Pfeiffer, Tobias Lasser	*X-Ray Computed Tomography with a Robotic Sample Holder*	336
15:00 – 16:20	Junyoung Kim, Yoseob Han, Jong Chul Ye	*DBP Domain Deep Learning for CT Reconstruction*	340
15:00 – 16:20	Shaohua Zhi, Marc Kachelrieß, Xuanqin Mou	*A Novel 4D-CBCT Reconstruction Framework via a Joint of Spatiotemporal Dictionary Learning and Prior Constraint*	344
15:00 – 16:20	Richard N. K. Bismark, Oliver Beuing, Georg Rose	*Overcoming Truncation Artifacts Caused by the Patient Table in Polyenergetic Statistical Reconstruction on Clinical C-Arm CT Data*	348
15:00 – 16:20	Mengfei Li, Hao Gao, Yunsong Zhao	*Fused Analytical Iterative Reconstruction for Sparse View Dual Energy Computed Tomography*	352

Poster Session 2f (Chair: Michael McNitt-Gray and Michael Knaup)

15:00 – 16:20	Taejin Kwon, Da-in Choi, Jaehong Hwang, Jaesung Lee, Jihwan Lim, Inje Lee, Seungryong Cho	*Dental Panoramic Image Reconstruction from Cone-Beam CT Projection Data*	356
15:00 – 16:20	Genwei Ma, Yinghui Zhang, Tong Wang, Xing Zhao, Hongwei Li	*An Intelligent Alternating Edge-Preserving Diffusion and Smoothing Neural Network for Limited-Angle Computed Tomography Reconstruction*	360
15:00 – 16:20	Jiayu Duan, Yang Li, Jianmei Cai, Xuanqin Mou	*A New Imaging Modality Based on The Flat Panel-Sources*	364
15:00 – 16:20	Masoud Elhamiasl, Sara Teruel Rivas, Koen Salvo, Edmond Sterpin, Johan Nuyts	*The Effect of CT Dose Reduction on Proton Therapy Dose Calculation*	368
15:00 – 16:20	Sarah Muller, Joscha Maier, Elias Eulig, Michael Knaup, Stefan Sawall, Marc Kachelrieß	*Deep Learning-Based Real-Time Estimation of Patient Dose Distributions for Various Medical CT Scan Protocols*	372

Poster Session 2g (Chair: Thomas Koehler and Xun Jia)

15:00 – 16:20	Shaojie Chang, Qiong Xu, Xi Chen, Xuanqin Mou	*Joint Spectrum Estimation and Image Reconstrution in Spectral CT*	376
15:00 – 16:20	Shabab Bazrafkan, Vincent Van Nieuwenhove, Joris Soons, Jan De Beenhouwer, Jan Sijbers	*Ringing Artefact Removal from Sparse View Tomosynthesis using Deep Neural Networks*	380
15:00 – 16:20	Ao Zheng, Zhiqiang Chen, Kaichao Liang, Li Zhang, Yuxiang Xing	*A CT Image Feature Space (CTIS) Loss for Reconstruction with Deep Learning Methods*	384
15:00 – 16:20	Hye Sun Yun, Tae Jun Jang, Sung Min Lee, Sang-Hwy Lee, Jin Keun Seo	*3D Landmark Annotation for Automatic 3D Cephalometry*	388
15:00 – 16:20	Florian Schiffers, Thomas Bochynek, André Aichert, Tobias Würfl, Michael Rubenstein, Oliver Cossairt	*Disassemblable Fieldwork CT Scanner Using a 3D-Printed Calibration Phantom*	392
15:00 – 16:20	Nele Blum, Maik Stille, Christian Hofmann, Thorsten M. Buzug	*Metal Artifact Reduction for 3D Cone-Beam CT by Non-Local Prior Image Integration*	396

Coffee Break

Oral Session : Artifact and Noise Reduction
Time : Wednesday, 16:40 – 18:20
Chairs : Emil Y. Sidky and Stanislav Žabić

Time	Author	Title	Page
16:40 – 17:00	Grace J. Gang, Tom Russ, Yiqun Q. Ma, Christian Toennes, Clifford R. Weiss, Jeffrey H. Siewerdsen, Lothar R. Schad, J. Webster Stayman	*Metal-Tolerant Noncircular Orbit Design and Implementation on Robotic C-Arm Systems*	400
17:00 – 17:20	Linxi Shi, N. Robert Bennett, Amy Shiroma, Mingshan Sun, Jin Zhang, Richard E. Colbeth, Josh Star-Lack, Minghui Lu, Adam Wang	*Single-Pass Metal Artifact Correction using a Dual Layer Flat Panel Detector*	404
17:20 – 17:40	Masakazu Matsuura, Jian Zhou, Hiroki Taguchi, Naruomi Akino, Zhou Yu	*Deep Learning Low Dose CT Image Denoising using Synthetic Training Data*	408
17:40 – 18:00	Julien Erath, Tim Vöth, Joscha Maier, Eric Fournié, Karl Stierstorfer, Martin Petersilka, Marc Kachelrieß	*Deep Learning-Based Cross-Scatter Correction for Clinical CT*	412
18:00 – 18:20	Philipp Roser, Annette Birkhold, Alexander Preuhs, Christopher Syben, Norbert Strobel, Markus Korwarschik, Rebecca Fahrig, Andreas K. Maier	*Deep Scatter Splines: Learning-Based Medical X-Ray Scatter Estimation Using B-splines*	416

Artificial intelligence is powerful...

Shared Intelligence is (empowering)

Because everyone in the cancer fight shares the same goal.

We're combining the insights and smarts from a wide range of sources, such as artificial intelligence, machine learning, and data analytics, to build a network of collective knowledge to help you deliver advanced care. We call it Shared Intelligence™. Because when all intelligence is shared, we'll get closer to what we're all fighting for—a world without fear of cancer.

Visit varian.com/victories to see how Shared Intelligence can help elevate your cancer care.

Safety information: Radiation may cause side effects and may not be appropriate for all cancers.

© 2020 Varian Medical Systems, Inc. Varian is a registered trademark, and Shared Intelligence is a trademark of Varian Medical Systems, Inc.

varian

Thursday, August 6

Oral Session : Photon Counting
Time : Thursday, 10:00 – 11:40
Chairs : Mats Danielsson and Taly Gilat-Schmidt

Time	Author	Title	Page
10:00 – 10:40	Thomas Flohr	*Basic short course: Photon Counting*	
10:40 – 11:00	Katsuyuki Taguchi	*The Number of Energy Windows for Photon Counting Detectors: Is More Actually More?*	422
11:00 – 11:20	Mengzhou Li, David S. Rundle, Ge Wang	*X-Ray Photon-Counting Data Correction via Deep Learning*	426
11:20 – 11:40	Stefan J. van der Sar, Stefan E. Brunner, Dennis R. Schaart	*First Steps towards SiPM-Based Scintillation Detectors for Photon-counting CT*	430

Coffee Break

Oral Session : Breast CT
Time : Thursday, 12:00 – 13:40
Chairs : John Boone and Ioannis Sechopoulos

Time	Author	Title	Page
12:00 – 12:20	John Boone, Ioannis Sechopoulos	*Introduction: Breast CT*	
12:20 – 12:40	Yunhuan Yang, Chengyijue Fang, Lei Zhu	*Sparse-View Cone-Beam Breast CT via Conditional Generative Adversarial Network Constrained by Image Edges*	434
12:40 – 13:00	Koen Michielsen, Nikita Moriakov, Jonas Teuwen, Ioannis Sechopoulos	*Deep Learning-Based Initialization of Iterative Reconstruction for Breast Tomosynthesis*	438
13:00 – 13:20	Antonio Sarno, Giovanni Mettivier, Francesca di Franco, Antonio Varallo, Andrew M. Hernandez, Kristina Bliznakova, John Boone, Paolo Russo	*A Dataset of Patient Derived Digital Breast Phantoms for in Silico Computed Tomography Dedicated to the Breast*	442
13:20 – 13:40	Stefano van Gogh, Michał Rawlik, Zhentian Wang, Marco Stampanoni	*Investigating Deep Learning-Based Grating Interferometry Breast Computed Tomography Image Denoising*	446

Lunch Break

Poster Session : Poster Session 3
Time : Thursday, 14:40 – 16:20

Time	Author	Title	Page
14:40 – 15:00	**Poster Fast Forward**: 30 s presentation for each poster		
Poster Session 3a (Chair: Wenli Wang and Jerome Liang)			
15:00 – 16:20	Daniil Kazantsev, Nicola Wadeson	*TOmographic MOdel-BAsed Reconstruction (ToMoBAR) Software for High Resolution Synchrotron X-Ray Tomography*	450
15:00 – 16:20	Marion Savanier, Cyril Riddell, Yves Trousset, Emilie Chouzenoux, Jean-Christophe Pesquet	*A Matched CBCT Projector-Backprojector Based on the Convolution of B-splines*	454
15:00 – 16:20	Eckhard Wehrse, Stefania Petra, Laura Klein, Lukas Rotkopf, Christian Herbert Ziener, Marc Kachelrieß, Heinz-Peter Schlemmer, Stefan Sawall	*Structured Regularization for Material Decomposition of Photon Counting CT Data using Collaborative Total Variation*	458
15:00 – 16:20	Obaidullah Rahman, Ken Sauer, Connor J. Evans, Ryan K. Roeder	*Direct Iterative Reconstruction of Multiple Basis Material Images in Photon-Counting Spectral CT*	462
15:00 – 16:20	Wenying Wang, Grace J. Gang, Matthew Tivnan, J. Webster Stayman	*Perturbation Response of Model-Based Material Decomposition with Edge-Preserving Penalties*	466
Poster Session 3b (Chair: Hiroyuki Kudo and Stefan Sawall)			
15:00 – 16:20	David Weller, Maik Stille, Philipp Koch, Mark Dreier, Alfred Mertins, Thorsten M. Buzug	*Convolutional Neural Network based Iterative Metal Artifact Reduction*	470
15:00 – 16:20	Alessandro Piol, Patricia M. Berdon, Manuel Desco, Monica Abella	*Prior Information for CT Reconstruction from Thermal Data*	474
15:00 – 16:20	David Tellenbach, Tobias Lasser	*elsa - An Elegant Framework for Precision Learning in Tomographic Reconstruction*	478
15:00 – 16:20	Maximilian Wattenberg, Lasse Hansen, Philipp Klein, Mattias P. Heinrich, Maik Stille, Thorsten M. Buzug	*Reconstruction of Blood Vessel Paths from Sparse Cone Beam Projection Images using Neural Networks and Ray-Tracing*	482
15:00 – 16:20	M. Nauwynck, Shabab Bazrafkan, A. H. van Heteren, Jan De Beenhouwer, Jan Sijbers	*Ring Artifact Reduction in Sinogram Space Using Deep Learning*	486

Poster Session 3c (Chair: Kirsten Boedeker and Zhye Yin)

15:00 – 16:20	Daouda Diakite, Maxime Martelli, Nicolas Gac	*An OpenCL Pipeline Implementation on Intel FPGA for 3D Backprojection*	490
15:00 – 16:20	Daniel Hadhazi, Gabor Horvath	*PGM Based TV L0 Regularized Linear Tomosynthesis Reconstruction*	494
15:00 – 16:20	Michael H. Schmitt	*Iterative Reconstruction-Based Increase of Operating Pitch in Aviation Security CT*	498
15:00 – 16:20	Lina Felsner, Tobias Würfl, Christopher Syben, Philipp Roser, Alexander Preuhs, Andreas K. Maier, Christian Riess	*Reconstruction of Voxels with Position- and Angle-Dependent Weightings*	502
15:00 – 16:20	Maria Magnusson, Gudrun Alm Carlsson, Michael Sandborg, Asa Carlsson Tedgren, Alexandr Malusek	*Investigation of Base Materials for Accurate Dual-Energy Computed Tomography. Comparison between the Alvarez-Macovski method and DIRA*	506

Poster Session 3d (Chair: Jingyan Xu and Peter B. Noël)

15:00 – 16:20	Sumin Baek, Okkyun Lee	*KNN-Based Nanoparticle Agent Identification in PCD-CT*	510
15:00 – 16:20	J. Graetz Dittmann, D. Müller, A. Balles, T. Tuohimaa, T. Donath, C. Fella	*ntCT: Submicrometer Laboratory X-Ray Microtomography down to 0.2µm Resolution*	514
15:00 – 16:20	Van Nguyen, Jan De Beenhouwer, Shabab Bazrafkan, A-T. Hoang, Sam Van Wassenbergh, Jan Sijbers	*BeadNet: a Network for Automated Spherical Marker Detection in Radiographs for geometry Calibration*	518
15:00 – 16:20	Emil Y. Sidky, Emily R. Paul, Taly Gilat-Schmidt, Xiaochuan Pan	*Spectral Calibration and Non-Linear Intensity Correction of Photon-Counting Detectors for Spectral CT*	522
15:00 – 16:20	Aurélien Coussat , Simon Rit , Rolf Clackdoyle, Michel Defrise, Laurent Desbat, Jean Michel Létang	*ROI CT Reconstruction Combining Analytic Inversion of the Finite Hilbert Transform and SVD*	526

Poster Session 3e (Chair: Lifeng Yu and Günter Lauritsch)

15:00 – 16:20	Hung Nguyen, Laurent Desbat, Rolf Clackdoyle	*Automatic Geometric Calibration in 3D Cone-Beam Geometry with Sources on a Line*	530
15:00 – 16:20	Zheng Zhang, Buxin Chen, Dan Xia, Emil Y. Sidky, Xiaochuan Pan	*A Preliminary Study on the Inversion of DXT from Limited-Angular-Range Data*	534
15:00 – 16:20	Hyunuk Jung, Chenyang Shen, Xun Jia	*Simultaneous Image Reconstruction and Element Decomposition for Iodine Contrast Agent Visualization in Multi-energy Cone Beam CT*	538
15:00 – 16:20	Buxin Chen, Emil Y. Sidky, Zheng Zhang, Dan Xia, Xiaochuan Pan	*A Preconditioning Scheme for Accelerating A Non-Convex Primal-Dual Reconstruction Algorithm in Multispectral CT*	542
15:00 – 16:20	Priyanka Ray, Parikshit Moitra, Yicheng Zhang, Tor Jensen, Scott Santeler, Dipanjan Pan	*Radiopaque Temporary Embolic Agent for Direct Visualization During Surgical Procedure and Follow up Treatments*	546

Poster Session 3f (Chair: Yuxiang Xing and Elias Eulig)

15:00 – 16:20	Xiaokun Huang, You Zhang, Jing Wang	*U-net Based Automatic CBCT-Based Liver Tumor Localization using Biomechanical Modeling*	550
15:00 – 16:20	Fatemeh Ostadhossein, Indu Tripathi, Lily Benig, Denae LoBato, Mahdieh Moghiseh, Chiara Lowe, Aamir Raja, Anthony Butler, Raj Panta, Marzieh Anjomrouz, Alex Chernoglazov, Dipanjan Pan	*Gram Scale Synthesis of Hafnia Nanodots for Detection of Bone Microcracks using Photon Counting CT Imaging*	554
15:00 – 16:20	Stephen Z. Liu, Qian Cao, Jeffrey H. Siewerdsen, J. Webster Stayman, Wojciech Zbijewski	*Three-Material Dual Energy Decomposition Using a Constrained Model-Based Algorithm*	558
15:00 – 16:20	Andriy Andreyev, Arkadiusz Sitek, Christoph Graf vom Hagen	*Cone Beam CT Image Reconstruction by Origin Ensemble Algorithm*	562
15:00 – 16:20	Zhicong Yu, Amit Jain, Chuanyong Bai, Daniel Gagnon	*A Constrained Curve-Fitting Method for Scatter Estimation for Helical Cone-Beam CT*	566

Poster Session 3g (Chair: Johan Nuyts and Carlo Amato)

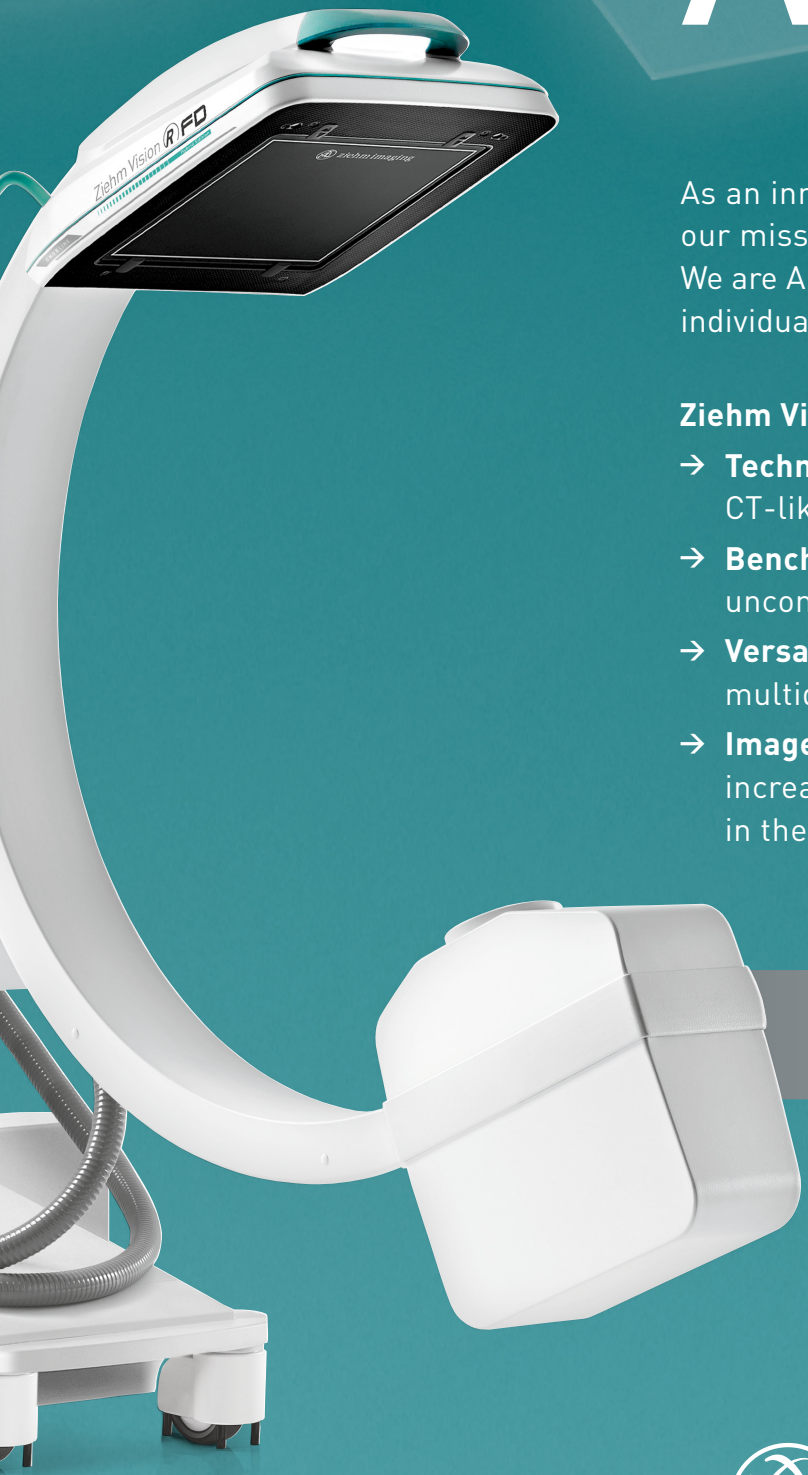
15:00 – 16:20	Alexander Meaney, Jussi Toivanen, Samuli Siltanen, Ville Kolehmainen	*Joint Reconstruction and Image Space Material Decomposition in Sparse View Spectral CT*	570
15:00 – 16:20	Hewei Gao, Tao Zhang, N. Robert Bennett, Adam Wang	*Scatter Correction for X-Ray CT Using Spectral Modulator with Flying Focal Spot*	574
15:00 – 16:20	Jingwen Zhuang, Yaguang Lu, Mei Bai	*A phantom Study on Exploring the Accuracy of the Weighted CT Dose Index for Multidetector CT*	578
15:00 – 16:20	Ludwig Ritschl, Marcel Beister, Magdalena Herbst, Steffen Kappler	*Isotropic Ultra-High Resolution CBCT Extremity Imaging Using a Robotic Radiography System*	582
15:00 – 16:20	Robert Cierniak, Piotr Pluta	*Fast Statistical Reconstruction Algorithm for a CT Scanner with Flying Focal Spot*	586
15:00 – 16:20	Paolo Gnudi, Bernd Schweizer, Marc Kachelrieß, Yannick Berker	*Denoising of X-Ray Projections and Computed Tomography Images using Convolutional Neural Networks without Clean Data*	590

Coffee Break

Oral Session : Security and Phase Contrast CT
Time : Thursday, 16:40 – 18:20
Chairs : Guang-Hong Chen and Peter B. Noël

Time	Author	Title	Page
16:40 – 17:00	Gengsheng L. Zeng	*Projection-Domain Iteration to Estimate Unreliable Measurements*	594
17:00 – 17:20	Wei Fang, Liang Li, Zhiqiang Chen, Chunguang Zong	*First Commercial Dual Energy MeV CT for Container Imaging: Experimental Results*	598
17:20 – 17:40	Sunho Lim, Donghyeon Lee, Hwanhui Jo, Seungryong Cho	*A New Empirical Dual-Energy Calibration (EDEC) Method for Material Decomposition in X-Ray Cargo Inspection*	602
17:40 – 18:00	Marius Reichardt, Jasper Frohn, Amara Khan, Frauke Alves, Tim Salditt	*X-Ray Phase-Contrast Tomography of Heart Tissue*	608
18:00 – 18:20	Grant T. Gullberg, Weijie Tao, Sally J. Kim, Yongjin Sung, Qiu Huang, Youngho Seo, Michael Fuller	*Tensor Tomography of Dark Field Scatter Using Single-Exposure Moiré Fringe Analysis of X-Ray Bi-Prism Interferometry*	612

ALWAYS AHEAD



As an innovation leader, we are committed to our mission of setting new technology standards. We are ALWAYS AHEAD in helping you meet your individual goals.

Ziehm Vision RFD 3D

- → **Technological leadership** with CT-like image quality
- → **Benchmark dose settings** and uncompromised image quality
- → **Versatile use** for 2D, 3D and multidisciplinary applications
- → **Image-guided navigation** for increased control and confidence in the OR

Visit us: **www.ziehm.com**

ziehm imaging

Friday, August 7

Oral Session: Observer
Time : Friday, 12:00 – 13:40
Chairs : Günter Lauritsch and Ge Wang

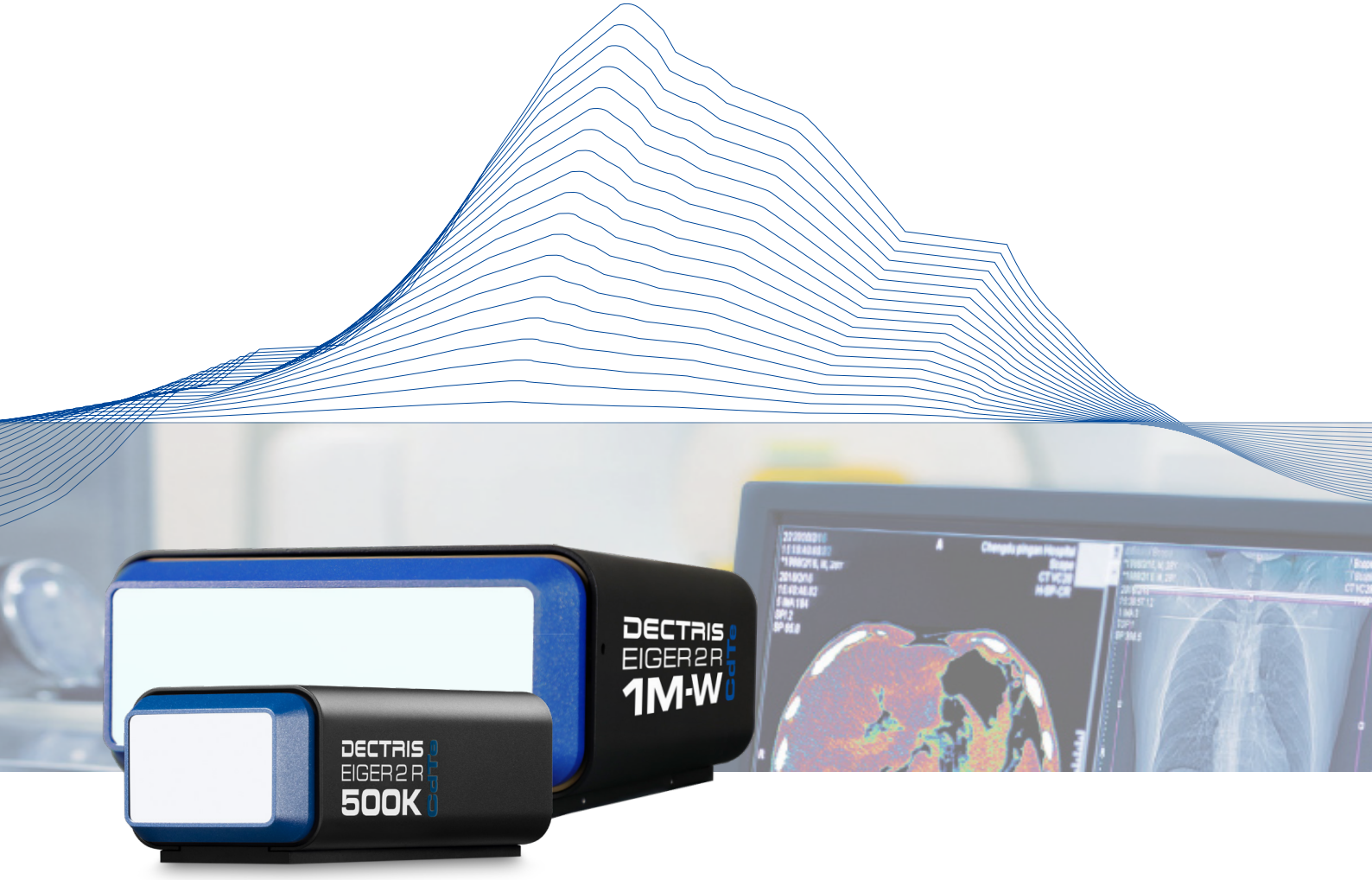
Time	Author	Title	Page
12:00 – 12:40	Frederic Noo	*Basic short course: Observer*	
12:40 – 13:00	Emil Y. Sidky, Weimin Zhou, Greg Ongie, Juan P. Cruz-Bastida, Ingrid S. Reiser, Mark A. Anastasio, Xiaochuan Pan	*Developing Signal Detectability as an Image Quality Metric for use with Non-Linear Image Reconstruction*	618
13:00 – 13:20	Irene Hernandez-Giron, Johan Michiel den Harder, Geert J. Streekstra, Wouter J. H. Veldkamp	*Objective Assessment of Clinical Image Quality in Thorax CT: Nodule Detectability in a 3D Printed Anthropomorphic Phantom by a 3D Model Observer*	622
13:20 – 13:40	Byeongjoon Kim, Minah Han, Jongduk Baek	*Training Strategies of a Convolutional Neural Network-Based Anthropomorphic Model Observer for Signal-Known-Exactly and Background-Known-Statistically Detection Tasks*	626

Lunch Break

Oral Session: Novel CT Technologies
Time : Friday, 14:40 – 16:20
Chairs : Thomas Flohr and Cyril Riddell

14:40 – 15:00	J. Webster Stayman, Matthew Tivnan, Grace J. Gang, Wenying Wang, Nadav Shapira, Peter B. Noël	*Grating-Based Spectral CT using Small Angle X-Ray Beam Deflections*	630
15:00 – 15:20	P. Wu, N. Sheth, T. Wang, A. Sisniega, A. Uneri, R. Han, R. Vijayan, P. Vagdargi, Björn W. Kreher, H. Kunze, G. Kleinszig, Sebastian Vogt, Jeffrey H. Siewerdsen	*C-Arm Non-Circular Orbits: Geometric Calibration, Image Quality, Avoidance of Metal Artifacts*	634
15:20 – 15:40	Matthew Tivnan, Wenying Wang, J. Webster Stayman	*Multi-Contrast CT Imaging with a Prototype Spatial-Spectral Filter*	638
15:40 – 16:00	Jannis Dickmann, Christina Sarosiek, George Coutrakon, Simon Rit, Nick Detrich, Victor Rykalin, Mark Pankuch, Robert P. Johnson, Reinhard W. Schulte, Katia Parodi, Guillaume Landry, George Dedes	*Dynamic Fluence Modulation using Proton CT for Low–dose Imaging in Particle Therapy*	642
16:00 – 16:20	Johan Nuyts, J. Webster Stayman	***Preview:** Fully3D 2021 + CT Meeting 2022*	

DECTRIS
detecting the future



EIGER2 R CdTe

Detector Series

Join us for the new era of spectral computed tomography

- Hybrid Photon Counting Technology
- Spectral imaging with multiple thresholds
- High resolution with 75 µm pixel size
- High absorption efficiency up to 120 keV with CdTe sensor
- High count rate up to 10^9 photons/sec/mm^2
- Excellent count rate stability

sales@dectris.com | www.dectris.com

Tuesday, August 4

Image Reconstruction

Dual and Multi-Energy Imaging

Poster Session 1

Diagnostic and Interventional

Benefits of distributed multidimensional undersampling for compressed sensing CT

Baiyu Chen, Erich Kobler, Aaron Sodickson, Daniel Sodickson, Ricardo Otazo[1]

***Abstract*— Compressed sensing exploits the natural sparsity of medical images and uses undersampled projection data for CT reconstruction. This study investigated the impact of undersampling scheme on reconstructed image quality. Two schemes were compared: angular undersampling and SparseCT undersampling. The former pulsed the x-ray beam during gantry rotation to acquire projections at discontinuous angles, while the latter interrupted the x-ray beam to acquire projections at discontinuous detector rows. This study also investigated undersampling imperfection, caused by the fact that x-ray beam cannot be pulsed on the timescale of millisecond and has penumbra.**

***The undersampled projection data were simulated by applying binary mask to fully-sampled patient projection data. They were then reconstructed iteratively using a sparsity-enforcing penalized weighted least squares cost function. Quantitative comparisons of image quality showed that SparseCT undersampling yielded higher SSIM and PSNR than angular undersampling. Visual comparisons also showed that SparseCT undersampling yielded better low-contrast detectability. With undersampling imperfection modeled into the simulation, both angular and SparseCT undersampling showed streak artifacts that impede the visibility of the pathology.**

In conclusion, SparseCT undersampling outperformed angular undersampling in terms of reconstruction image quality, possibly because it sampled the projection space more uniformly with less correlations. Undersampling scheme has significant impact on CT image quality, and must be carefully designed and optimized for compressed sensing.

***Index Terms*— CT, compressed sensing, SparseCT, undersampling**

I. Introduction

Compressed sensing is a promising approach for reducing CT radiation dose. It exploits the fact that CT images are sparse in a certain transform domain, and reconstructs images using undersampled projection data [1, 2]. The choice of undersampling scheme has a direct impact on reconstructed image quality, and therefore warrants investigation.

In this study we compared two undersampling schemes, angular undersampling and SparseCT undersampling (Figure 1(a)). Angular undersampling pulses the x-ray during gantry rotation to acquire projections at discontinuous angles [2]. SparseCT undersampling, on the other hand, acquires projections at every angle, but interrupts the beam with a multi-slit collimator (MSC) to irradiate discontinuous detector rows (Figure 1(b)) [3]. Furthermore, the beam blocker shifts from view to view to vary the undersampling pattern and thus increase incoherence for compressed sensing reconstruction.

We also investigated the impact of undersampling imperfection on image quality. For angular undersampling, the imperfection rises from the fact that the x-ray beam cannot be pulsed on the timescale of millisecond. As a result, each pulse irradiates several continuous projections together. For SparseCT, the imperfection rises from the fact that focal spot size has finite size and creates penumbra on both sides of each undersampled beam. As a result, each undersampled beam (the beam through one slit of the multi-slit collimator) irradiates several continuous detector rows.

We simulated the undersampled projection data from fully-sampled patient projection data acquired with a standard dose. All patients have known, low-contrast pathologies. The reconstructed images were compared in terms of structure similarity (SSIM), peak signal-to-noise ratio (PSNR), and the detectability of the low contrast pathologies.

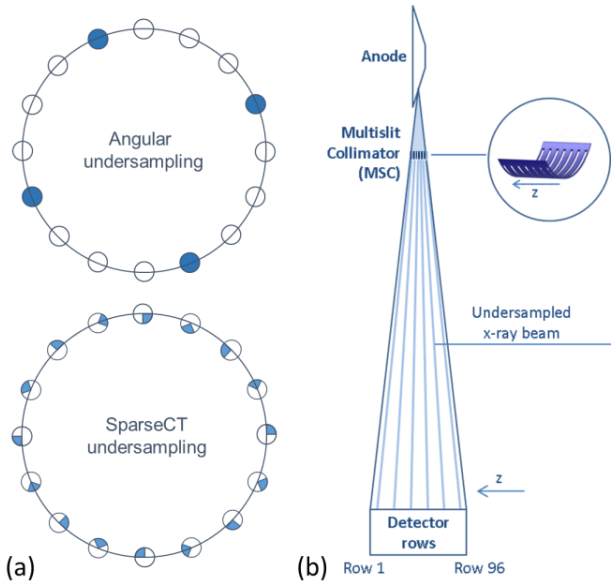


Figure 1: (a) Two undersampling schemes for achieving 4-fold dose reduction: angular and SparseCT. (b) The SparseCT undersampling is achieved by interrupting the beam with a multislit collimator (MSC). The MSC moves along the detector row direction during gantry rotation to change the undersampling pattern for each view.

II. Methods

A. Simulation of Undersampled Projection Data

Undersampled projection data were simulated by multiplying fully-sampled projection data with binary masks. The binary mask mathematically represented the undersampling schemes, where "1" means sampled and "0" means skipped. For angular

This work was supported by the National Institutes of Biomedical Imaging and Bioengineering under Grant U01 EB018760.

Baiyu Chen (e-mail: baiyu.chen@nyumc.org) and Daniel Sodickson are with NYU School of Medicine, New York, NY USA.

Erich Kobler is with Graz University of Technology, Austria.

Aaron Sodickson is with Brigham and Women's Hospital, Boston, MA USA.

Ricardo Otazo (e-mail: otazotoj@mskcc.org) is with Memorial Sloan Kettering Cancer Center, New York, NY, USA

undersampling, the naming convention of the mask is "Angular_x_y", where x is the number of views irradiated in one pulse and y is the number of views between pulses. For SparseCT undersampling, the naming convention is "SparseCT_x_y_z", where x is the number of detector rows irradiated by one undersampled beam, y is the distance between two neighboring undersampled beams in the unit of detector rows, and z is the shift of the undersampled beams from one view to the next view in the unit of detector rows.

Four types of binary masks were tested in this study, all enabling a 4-fold dose reduction: Angular_1_4, Angular_4_16, SparseCT_1_4_1, and SparseCT_4_16_1. Figure 2 and Figure 3 illustrate undersampled projection data using angular undersampling and SparseCT undersampling, respectively. Angular_1_4 represented the scenario where the x-ray beam can be pulsed quickly. SparseCT_1_4_1 represented the scenario where the focal spot is small (no penumbra). Angular_4_16 and SparseCT_4_16_1, on the other hand, represented more realistic undersampling scenarios.

The fully-sampled projection data to be undersampled were chosen from a standardized patient database, where each patient has known pathologies [4, 5]. The database also provided fully-sampled but reduced-tube-current projection data (simulated via noise-insertion), which we used to compare with compressed sensing results.

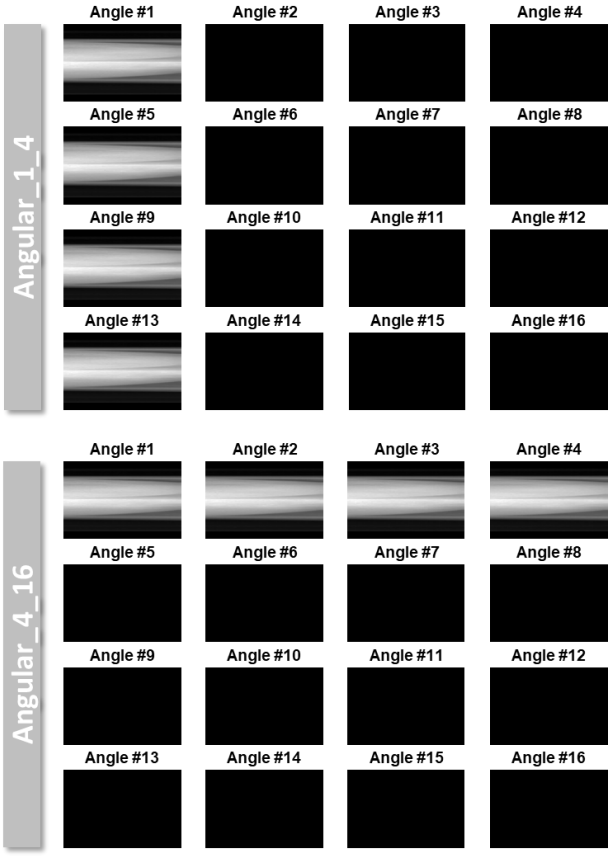


Figure 2: Illustration of angular undersampling using Angular_1_4 (ideal) and Angular_4_16 (more realistic). Only the first 16 projections of a scan are shown.

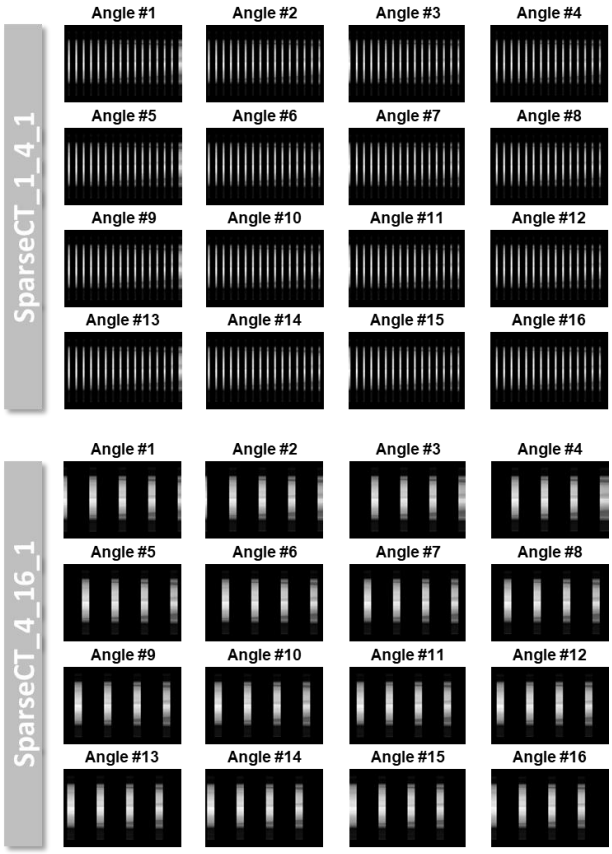


Figure 3: Illustration of SparseCT undersampling using SparseCT_1_4_1 (ideal) and SparseCT_4_16_1 (more realistic). Only the first 16 projections of a scan are shown.

B. Reconstruction of Projection Data

Image reconstruction was performed by solving the following optimization problem:

$$\underset{u}{\operatorname{argmin}} \frac{1}{2}\|Au - f\|_W^2 + \lambda\|Du\|_{2,1},$$

where A is the data acquisition operator, u is the image to be reconstructed, f is the projection data, $\|Du\|_{2,1}$ is the three-dimensional isotropic total variance regularizer, which favors sparse image gradients, and $\lambda > 0$ is a scalar weight that balances the tradeoff between regularization and data fidelity. The subscript W indicates the inclusion of statistical weights, which accounts for the varying quantum and electronic noise level in projection data.

The optimization problem was solved using a primal dual hybrid gradient [6] with a preconditioning described in [7]. All datasets were reconstructed using the same algorithm for a fair comparison. All reconstructions use a λ of 20, slice thickness of 3 mm, and 80 iterations.

C. Comparison Metrics

The images reconstructed from undersampled projections were compared to the images reconstructed from fully-sampled projections in terms of SSIM and PSNR. The detectability of the low contrast pathologies was also visually compared.

III. Results

A. Reconstruction

Figure 4 -Figure **6** show the reconstructed images of 3 patients under 6 settings: fully-sampled, fully-sampled but using a quarter of original tube current, Angular_1_4, Angular_4_16, SparseCT_1_4_1, and SparseCT_4_16_1.

The SSIM and PSNR of the reconstructions, computed with respect to the fully-sampled images, are marked in each figure. In general, SparseCT undersampling presents higher SSIM and PSNR than angular undersampling, which is likely due to the fact that SparseCT samples the projection space in a more uniform and less correlated fashion. For example, angular undersampling presents a few large gaps in projection space (for 4-fold undersampling, 3 full views are omitted), while SparseCT presents many small gaps distributed in projection space (for 4-fold undersampling, all views are sampled with a different slice undersampling pattern for each view). SparseCT_1_4_1 undersampling also outperforms reduce-tube-current iterative reconstruction in 2 out of 3 patient cases. With the undersampling imperfection taken into consideration (Angular_4_16 and SparseCT_4_16_1), both SSIM and PSNR decrease, which is expected because the projections are sampled less uniformly.

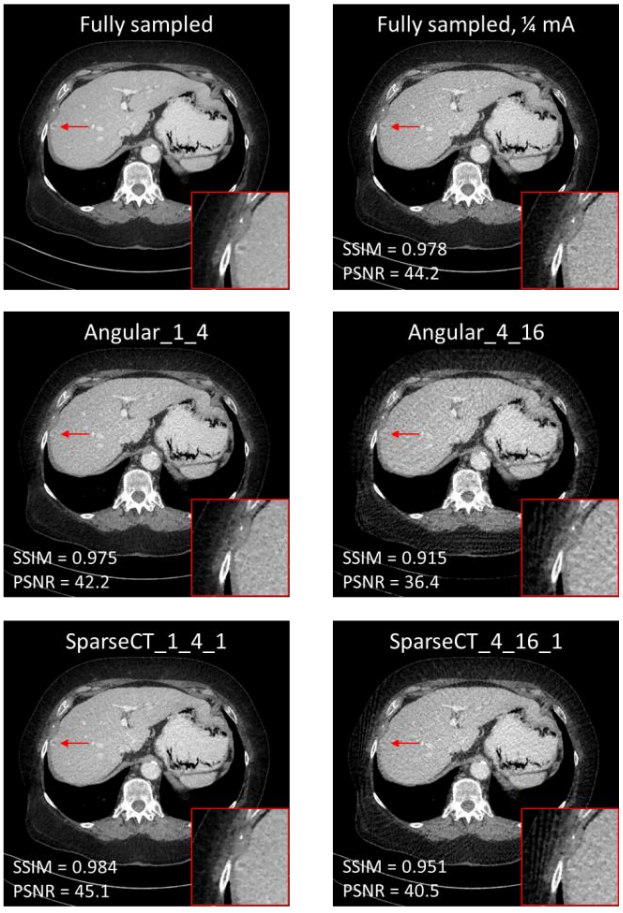


Figure 4: Reconstructed images in a representative patient using various undersampling settings. Full dose image and reduced-tube-current image are also provided as reference. The arrow points to a hemangioma, which is magnified in the red box.

The different pathologies in the patients (hemangioma, metastasis, and benign cyst) are indicated by arrows in the figures and magnified in the red boxes. Visually, SparseCT_1_4_1 has detectability comparable to fully-sampled full dose and quarter dose, and better than Angular_1_4. Angular_4_16 and SparseCT_4_16_1, due to the undersampling imperfection, have streak artifacts that impede visibility of the pathology.

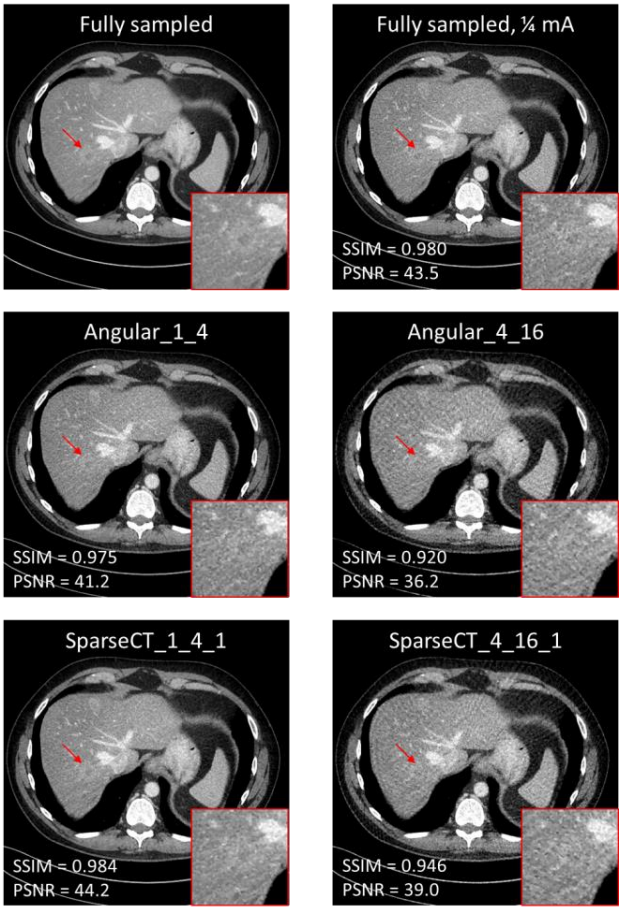


Figure 5: Reconstructed images in a representative second patient using various undersampling settings. Full dose image and reduced-tube-current image also provided as reference. The arrow points to a metastasis, which is magnified in the red box.

IV. Conclusion

This study aimed at analyzing an old question in the application of compressed sensing to CT: what is the best way to undersample a CT acquisition? We compared two different undersampling schemes for compressed sensing CT: angular undersampling and SparseCT undersampling. Quantitative comparisons of image quality showed that SparseCT undersampling yielded higher SSIM and PSNR than angular undersampling. Visual comparisons also showed that SparseCT undersampling yielded better low-contrast detectability in patient datasets. The higher performance of SparseCT undersampling is mainly due to the more uniform distribution of samples over projection space, which removes the challenge of dealing with large gaps in projection space, as in angular undersampling.

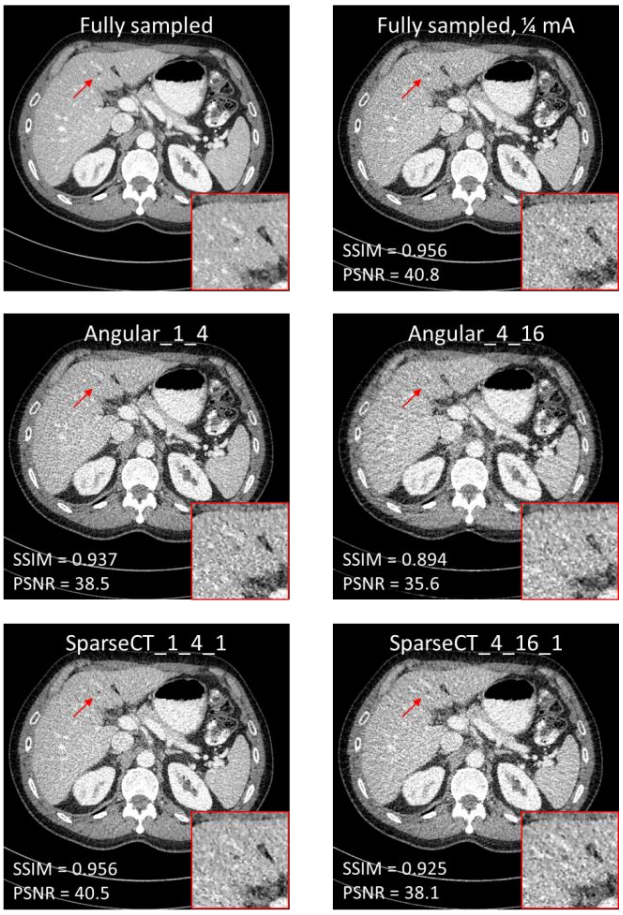


Figure 6: Reconstructed images in a representative third patient using various undersampling settings. Full dose image and reduced-tube-current image also provided as reference. The arrow points to a benign cyst, which is magnified in the red box.

This study also studied the impact of undersampling imperfection on reconstruction image quality and showed decrease in image quality when the undersampling pattern become less uniform. This point would be particularly important when taking these concepts to practice.

In conclusion, this study demonstrates that spreading the samples over multiple dimensions in the projection space increases the performance of compressed sensing for CT and can lead to higher dose reduction factors.

ACKNOWLEDGMENTS

This work was funded by a National Institute of Biomedical Imaging and Bioengineering (NIBIB) grant (U01 EB018760).

REFERENCES

[1] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207-1223, 2006.

[2] G. H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med Phys*, vol. 35, no. 2, pp. 660-3, Feb 2008.

[3] B. Chen *et al.*, "SparseCT: System concept and design of multislit collimators," *Medical physics*, vol. 46, no. 6, pp. 2589-2599, June 2019 2019.

[4] B. Chen, X. Duan, Z. Yu, S. Leng, L. Yu, and C. McCollough, "Technical Note: Development and validation of an open data format for CT projection data," *Med Phys*, vol. 42, no. 12, pp. 6964-72, Dec 2015.

[5] B. Chen, S. Leng, L. Yu, D. Holmes, 3rd, J. Fletcher, and C. McCollough, "An Open Library of CT Patient Projection Data," *Proc SPIE Int Soc Opt Eng*, vol. 9783, Feb 27 2016.

[6] A. Chambolle and T. Pock, "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging," (in English), *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120-145, Mar 2011.

[7] D. Kim, S. Ramani, and J. A. Fessler, "Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction," *IEEE transactions on medical imaging*, vol. 34, no. 1, pp. 167-178, 2015.

A k-NEAREST NEIGHBOR REGULARIZER FOR MODEL BASED CT RECONSTRUCTION

Jingyan Xu and Frédéric Noo

***Abstract*—Commonly used regularizers in model based image reconstruction (MBIR) penalize neighboring differences regardless of the image content, by which we mean that the neighboring differences may arise from the same organ or from different organs and penalizing the latter leads to reduced image contrast. In this work, we aim to improve fine feature preservation by designing regularizers that can potentially differentiate the different kinds of neighboring differences. The proposed regularizer relies on the ranking order of neighboring differences. That is, imagine all neighboring differences in absolute values are ordered from the smallest to the largest, and only the minimal k among those are penalized. In other words, the regularizer only penalize the k-nearest neighbors (k-NN) of each voxel. As the minimal differences likely arise from voxels all from the same organ, the k-NN regularizer applies smoothing to where it is needed the most. The k-NN regularizer is nonconvex and nonsmooth. For image reconstruction, we developed a smoothing technique to approximate the regularizer by a smooth nonconvex function, and subsequently apply gradient descent to the smooth objective function. We performed numerical studies using a hybrid data set consisting of physical phantom data and digitally inserted fine features. We compared the proposed k-NN regularizers with the Huber regularizer. Our quantitative evaluation demonstrates that the k-NN regularizer improved the conspicuity of low contrast features.**

I. INTRODUCTION

Model based image reconstruction (MBIR) is often formulated as the problem of minimizing

$$\frac{1}{2}\|Ax-y\|^2+\beta R(x) \tag{1}$$

where $x \in R^n$ is the unknown image, y is the acquired projection data, and A is the forward projection matrix. The regularizer $R(x)$ can incorporate a priori knowledge of the unknown x, and the hyperparameter $\beta > 0$ balances between the data fitting term and the prior information.

A large class of regularizers can be represented by

$$R(x)=\sum_i \sum_{j\in N_i} \phi(x_i - x_j) \tag{2}$$

where $\phi(s) : R \to R$ can be (non)convex and (non)smooth, and $1 \le i \le n$ is the voxel index, and N_i is the neighborhood of i that typically involves the 4 or 8 nearest neighbors (in 2D). For example, when $\phi(s) = |s|$, (2) is the (anisotropic) TV regularizer that is widely used in MBIR.

A weakness of (2) is that it is not "content-aware." That is, it penalizes all neighboring differences regardless of the source of the differences. For example, if the voxel i is in the interior of a piecewise-constant organ, then penalizing all neighboring differences is effective in reducing noise; if the voxel i is near the organ boundary, then it would be natural if we only penalize those differences that are calculated from the same organ, and not penalize those from across the organ boundaries. This natural strategy, however, is not what (2) does; and this indifference to the underlying image content leads to contrast reduction. Many clinical tasks rely on detecting and differentiating fine features for lesion classification. Regularizers of the form (2), whether $\phi(s)$ is convex or not, may be suboptimal for such tasks.

The proposed regularizers in this work can potentially differentiate the neighboring differences that are calculated within the same organ from those that are calculated across different organs, and penalize only the first kind. In this sense, the proposed regularizers are adaptive to the image content and can better preserve fine feature contrast and may improve lesion classification performance.

A. Notation

The proposed regularizer will be described using sorted sequences.[1] Let $\boldsymbol{s} = (s_1, \cdots, s_m) \in \mathrm{R}^m$. We denote by $s_{(i)}$, $i = 1, \cdots, m$, the i-th smallest component of $\boldsymbol{s}$, *i.e.*,

$$s_{(1)} \le s_{(2)} \le \cdots \le s_{(m)}$$

It may also be convenient to sort in the descending order. We denote by $s_{[i]}$ be the i-th largest component of $\boldsymbol{s}$, *i.e.*,

$$s_{[1]} \ge s_{[2]} \ge \cdots \ge s_{[m]},$$

then obviously

$$\sum_{i=1}^{m} s_i = \sum_{i=1}^{\underline{k}} s_{(i)} + \sum_{i=1}^{\bar{k}} s_{[i]} \tag{3}$$

for any nonnegative integers $\underline{k}$, $\bar{k}$, such that $\underline{k} + \bar{k} = m$. We use the convention that if $\underline{k} = 0$, then $\sum_i^{\underline{k}} = 0$, and similarly for $\bar{k}$. The sorting operation depends on the vector dimension m, which may not be explicitly mentioned if the context is clear.

In MBIR, the neighboring differences are often measured in absolute values. In other words, the potential function ϕ in (2) is often even. We denote by $|\boldsymbol{s}|$ the element-wise absolute value of $\boldsymbol{s} \in R^m$, and use $|s|_{(i)}$ and $|s|_{[i]}$ to denote the i-th minimum, the i-th maximum value of $|\boldsymbol{s}|$. Similar to (3), we have

$$\sum_{i=1}^{m} |s|_i = \sum_{i=1}^{\underline{k}} |s|_{(i)} + \sum_{i=1}^{\bar{k}} |s|_{[i]} \tag{4}$$

for any nonnegative integer $\underline{k}$, $\bar{k}$, such that $\underline{k} + \bar{k} = m$.

II. METHOD

The proposed k-NN regularizer penalizes only the smaller neighboring differences, and skip the larger ones that more likely arise from across organ boundaries. This regularizer is nonsmooth and nonconvex, for which we develop a smooth (nonconvex) approximation and discuss its properties. The smoothed k-NN regularizer is then combined with the data fitting term as the objective function for MBIR.

J. Xu (jxu@jhmi.edu) is with Department of Radiology, Johns Hopkins University. F. Noo is with Department of Radiology and Imaging Sciences, University of Utah.

[1] Note that we use sorting to explain the concept, but the sorting operation itself can be avoided in numerical algorithms for efficiency reasons. We will come back to this later.

A. The k-nearest neighbor (k-NN) regularizer

Let $\boldsymbol{s} = (s_1, \cdots, s_m) \in R^m$, and define

$$\underline{\phi}(\boldsymbol{s}) = \sum_{k=1}^{\underline{k}} |s|_{(k)} = |s|_{(1)} + \cdots + |s|_{(\underline{k})} \tag{5}$$

as the min-$\underline{k}$-sum function, then the proposed regularizer can be written as a composition of $\underline{\phi}$ with a finite difference operator.

$$R(x) = \sum_i \underline{\phi}(\boldsymbol{s}_i), \quad \boldsymbol{s}_i = Dx_i = \{x_i - x_j | j \in N_i\} \tag{6}$$

If $\underline{k} = 1$, then (6) penalizes the least one among all neighboring differences for each voxel i; if $\underline{k} = 2$, then (6) penalizes the sum of the least two among all neighboring differences. If $\underline{k} = |N_i|$, then (6) penalizes all neighboring differences, *i.e.*, the same as the anisotropic TV. By choosing $\underline{k} < |N_i|$ properly, we can use (6) to avoid penalizing larger differences and preserve fine feature contrast.

Properties of the regularizer (6) are inherited from properties of the min-$\underline{k}$-sum function (5). It is well known that $\underline{\phi}(\boldsymbol{s})$ is a non-convex, non-smooth function of $\boldsymbol{s}$ (unless when $\underline{k} = m$ then $\underline{\phi}(\boldsymbol{s})$ is convex and possibly nonsmooth). Fig. 1(a) is the plot of a special case of (5) where $m = 2$ and $\underline{k} = 1$. The nonconvex and nonsmooth properties are obvious.

Image reconstruction using the k-NN can be formulated by replacing $R(x)$ in (1) by (6). Since the regularizer is nonconvex and nonsmooth, so is the MBIR objective function. Existing algorithms for nonconvex minimization are often developed for specific models. For example, they may require the nonconvex component to be smooth or simple [1], [2], or a simple function in composition with an surjective linear operator [3], [4], but none of these conditions hold for the k-NN regularizer (6).

In this work, we develop a smoothing technique to approximate the nonsmooth, nonconvex function (5) by a smooth (nonconvex) function. This smooth approximation induces a smooth regularizer which is then joined by the data fidelity term of (1) to form a nonconvex but smooth objective function. The smoothing technique allows us to apply a gradient descent algorithm to obtain the solution.

B. A smooth approximation of (5)

Let $\boldsymbol{s} = (s_1, \cdots, s_m) \in R^m$. To complement (5), we define

$$\bar{\phi}(\boldsymbol{s}) = \sum_{k=1}^{\bar{k}} |s|_{[k]} = |s|_{[1]} + \cdots + |s|_{[\bar{k}]} \tag{7}$$

which is convex and nonsmooth. Rewriting (4), we have

$$\phi(\boldsymbol{s}) \triangleq \sum_{i=1}^{m} |s|_i = \underline{\phi}(\boldsymbol{s}; \underline{k}) + \bar{\phi}(\boldsymbol{s}; \bar{k}) \tag{8}$$

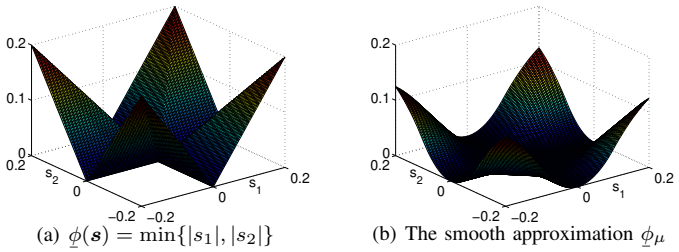
for any nonnegative integer $\underline{k}$, $\bar{k}$, such that $\underline{k} + \bar{k} = m$. From (8), we see that the min-$\underline{k}$-sum (5) can be written as a difference-of-convex (DC) function, *i.e.*,

$$\underline{\phi}(\boldsymbol{s}; \underline{k}) = \phi(\boldsymbol{s}) - \bar{\phi}(\boldsymbol{s}; \bar{k}). \tag{9}$$

Since the two components on the right hand side of (9) are both convex, they each can be approximated using their Moreau-Yosida regularization [5] to obtain, say $\phi_{\mu_1}(\boldsymbol{s})$ and $\bar{\phi}_{\mu_2}(\boldsymbol{s})$. Our smooth approximation of $\underline{\phi}(\boldsymbol{s})$ is indeed based on such individual smoothing with the additional condition that $\mu_1 = \mu_2$. More precisely, we define

$$\underline{\phi}_\mu(\boldsymbol{s}; \underline{k}) = \phi_\mu(\boldsymbol{s}) - \bar{\phi}_\mu(\boldsymbol{s}; \bar{k}). \tag{10}$$

The properties of $\underline{\phi}_\mu(\boldsymbol{s}; \underline{k})$ come from the two components functions in (10) , which we discuss next.



(a) $\underline{\phi}(\boldsymbol{s}) = \min\{|s_1|, |s_2|\}$ (b) The smooth approximation $\underline{\phi}_\mu$

Fig. 1: Surface plot of $\underline{\phi}(s_1, s_2)$ and its approximation $\underline{\phi}_\mu(s_1, s_2)$.

C. Properties of the smooth approximation (10)

Since $\phi(\boldsymbol{s})$ is simply the ℓ_1 norm, its Moreau-Yosida regularization $\phi_\mu(\boldsymbol{s})$ is the Huber function with the parameter μ applied to the individual components of $\boldsymbol{s}$. Similarly, the Moreau regularization $\bar{\phi}_\mu(\boldsymbol{s}; \bar{k})$ of $\bar{\phi}(\boldsymbol{s}; \bar{k})$ is obtained as the following[2]

$$\begin{aligned} \bar{\phi}_\mu(\boldsymbol{s}) &= \inf_{\boldsymbol{t}} \{\bar{\phi}(\boldsymbol{t}) + \frac{1}{2\mu}\|\boldsymbol{s} - \boldsymbol{t}\|^2\} \\ &\triangleq \inf_{\boldsymbol{t}} \{\sum_{k=1}^{\bar{k}} |t|_{[k]} + \frac{1}{2\mu}\|\boldsymbol{s} - \boldsymbol{t}\|^2\} \end{aligned} \tag{11}$$

Since $\bar{\phi}(\boldsymbol{s})$ is convex, $\bar{\phi}_\mu(\boldsymbol{s})$ is smooth with Lipschitz-continuous gradient (with Lipschitz constant $1/\mu$). Let the solution of (11) be $\boldsymbol{t}^*(\boldsymbol{s}) = \{t_1^*, \cdots, t_m^*\}$. Then the gradient of $\bar{\phi}_\mu(\boldsymbol{s})$ can be evaluated as

$$\bar{\phi}'_\mu(\boldsymbol{s}) = \frac{\boldsymbol{s} - \boldsymbol{t}^*(\boldsymbol{s})}{\mu} \tag{12}$$

Some observations of the solution $\boldsymbol{t}^*(\boldsymbol{s})$ are immediate.

- $|t_i^*| \le |s_i|$. That is, $t^*(\boldsymbol{s})$ is a shrinkage operator.
- If $|s_i| \ge |s_j|$, then $|t_i^*| \ge |t_j^*|$. Otherwise, we could make a switch $|t_i^*| \leftrightarrow |t_j^*|$ and obtain a lower objective value on the left-hand-side of (11). In other words, the mapping $\boldsymbol{t}^*(\boldsymbol{s}) : R^m \to R^m$ is order-preserving.

The approximation (10) is based on separate approximation of the components of the DC function. The following properties can be obtained for $\underline{\phi}(\boldsymbol{s}; \underline{k})$ (details omitted due to space).

- $\underline{\phi}_\mu(\boldsymbol{s}) \ge 0$, and $\min_{\boldsymbol{s}} \underline{\phi}_\mu(\boldsymbol{s}) = 0$.
- $\underline{\phi}_\mu(\boldsymbol{s})$ is a lower bound of the original, *i.e.*, $\underline{\phi}_\mu(\boldsymbol{s}) \le \underline{\phi}(\boldsymbol{s})$.

Combining the above two, we see that both $\underline{\phi}(\boldsymbol{s})$ and its approximation $\underline{\phi}_\mu(\boldsymbol{s})$ are nonnegative and share the same global minimal. The plot of an example of $\underline{\phi}_\mu(\boldsymbol{s}; \underline{k})$ for $\underline{k} = 1, m = 2$ is shown in Fig. 1(b).

[2]We omit the notation $\bar{k}$ when there is no confusion.

D. Image reconstruction algorithm

Using (10), the smoothed k-NN regularizer is given by

$$R_\mu(\boldsymbol{x}) = \sum_i \underline{\phi}_\mu(\boldsymbol{s}_i) = \sum_i \phi_\mu(\boldsymbol{s}_i) - \bar{\phi}_\mu(\boldsymbol{s}_i), \qquad \boldsymbol{s}_i = D\boldsymbol{x}_i = \{x_i - x_j | j \in N_i\}. \tag{13}$$

Since $\underline{\phi}_\mu(\boldsymbol{s}) \geq 0$, so is $R_\mu(\boldsymbol{x})$. In this sense, (13) is well-defined.

Combining (13) and the data fidelity term in (1), the objective function for our MBIR problem is

$$\frac{1}{2}\|y - Ax\|^2 + \beta R_\mu(\boldsymbol{x}). \tag{14}$$

By construction, both the data fitting and the (smoothed) regularizer terms in (14) have Lipschitz continuous gradients. We can apply gradient descent algorithms to (14) which is guaranteed to reach a local minimal.

Calculating the gradient of (14) requires evaluating (12), which in turn requires the solution to the minimization problem (11). This problem is extensively studied in the literature and its solution can be obtained by a finite-termination algorithm in $O(m)$. See [6] and the references within.[3]

III. Numerical studies

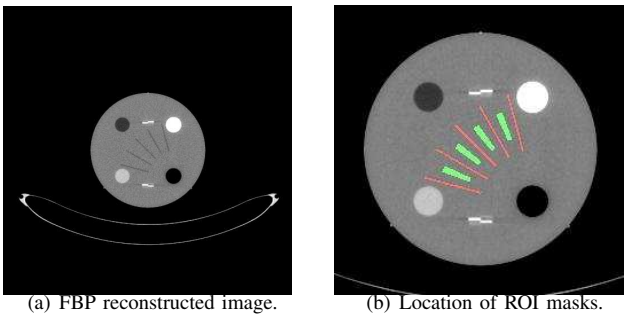
We performed numerical experiments to study the effectiveness of the k-NN regularizer in preserving fine features. In this preliminary study, we focus on the following two aspects: (1) effectiveness in feature preservation in comparison with other (non)convex regularizers; and (2) effect of the parameter $\underline{k}$ in the k-NN regularizer.

A. Data generation

Our image reconstruction used hybrid projection data by combining the projection of a physical phantom and digitally simulated signals. The projection data of an ACR phantom were acquired on a Siemens Somatom Flash scanner [7]. Using the same acquisition geometry, we simulated fanbeam projection data of thin cylinders that varied in their orientations, but all were 5 cm in length, 2 mm in diameter, and had the same uniform attenuation distribution (Fig. 2). We then combined the simulated data of the cylinders with the ACR data. The contrast between the cylinders and the background was tuned to be ~ 47.8 HU. The size of the hybrid projection data was 736×2304 (channels $\times$ views) with 2FFS in-plane flying focal spots.

We considered MBIR with two regularizers, the proposed k-NN regularizer and the Huber regularizer, both with an 8-pixel neighborhood, *i.e.*, $|N_i| = 8$. All image reconstruction was obtained by applying the gradient descent algorithm with the step size of $1/L$ where L is the Lipschitz constant of the gradient of the objective function. For reconstruction with the k-NN regularizer, the Lipschitz constant can be approximated by $L = L_d + \beta(L_1 + L_2)$, where L_d is the Lipschitz constant of the data-fitting term, and $L_1 = L_2 = \|D\|^2/\mu$ for the smoothed k-NN regularizer. For reconstruction with the Huber regularizer, we have $L = L_d + \beta L_1$. For both methods (regularizers) we performed a parameter sweep over the hyperparameter β; for the k-NN regularizer, we also varied the integer $\underline{k}$ to study its effects. Multicore-programming with the openMP library was used for the forward and backward projection to speed up reconstruction. All reconstructed images were obtained at 2000 iterations, after which point the successive updates were smaller than 0.1 HU.

[3]These linear time algorithms do not require sorting a size-m sequence, as sorting itself has a complexity of $O(m \log m)$.



(a) FBP reconstructed image. (b) Location of ROI masks.

Fig. 2: (a) The FBP reconstructed image from the hybrid projection data. The contrast of the five 2 mm cylinders was 47.8 HU. Display (C, W) = (1000, 400) HU. (b) The location of the ROI masks for the background (green) and the cylinders (red).

B. Figure of merit

The imaging task was to estimate the mean intensity of the 2 mm cylinders. We used the region-of-interest (ROI) masks (Fig. 2, red) over the cylinders and calculated the standard deviation (stdv_f) and mean of the pixel intensities within the masks. The background noise (stdv_b) was measured using the background ROIs located between the cylinders (Fig. 2, green). A composite measure calculated by $\sqrt{(\mathrm{stdv}_f)^2 + (\mathrm{stdv}_b)^2}$ was used as the noise in the contrast-noise calculation. The calculation was repeated for all methods and for all hyperparameter values.

C. Results

The noise and contrast curves for all compared methods are shown in Fig. 3. We notice that by skipping the larger neighboring differences, the k-NN regularizer can indeed improve the noise-contrast tradeoff over the Huber regularizer, which corresponds to setting $\underline{k} = 8$ in the k-NN regularizer.

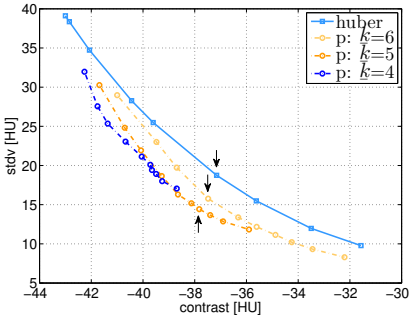


Fig. 3: Noise-contrast tradeoff for measuring the mean intensity of the cylinders. For the proposed regularizer, marked with 'p', we also varied the parameter k. The arrows point to the sampled images in Fig. 4.

The different parameters $\underline{k}$ also affect the noise-contrast tradeoff. A larger $\underline{k}$ will penalize more of the smaller neighboring differences. In the limit, when all neighboring differences are penalized, there should be no difference between the Huber regularizer and k-NN. A lower value $\underline{k}$ will penalize fewer smaller neighboring differences, increasing the likelihood of skipping the

larger differences that come from across region boundaries and may lead to a better noise-contrast tradeoff. The noise-contrast curves in Fig. 3 agree with this analysis.

The arrows in Fig. 3 point to images with similar contrast for the different methods. These images, zoomed in to the cylinders, are shown in Fig. 4 with corresponding labels. As a reference for comparison, we also included an image reconstructed by the FBP algorithm with a filter selected to match the same contrast level as the iterative methods. Among the four images, the ones obtained by the proposed regularizers yield better conspicuity of the thin cylinders.

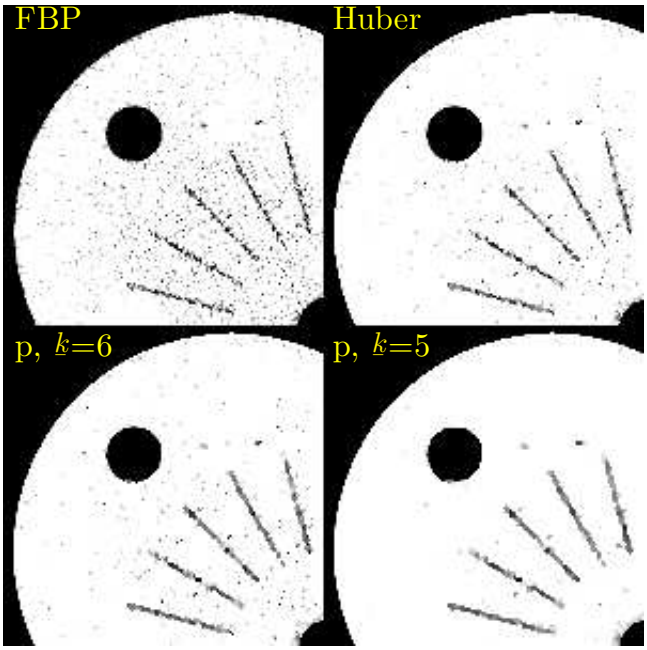


Fig. 4: Sample images marked by arrows in Fig. 3. The cylinders appear wider using the proposed regularizers, indicating better contrast preservation. The display window, (C, W) = (962, 40) HU, was chosen to visualize the cylinders better.

The voxel intensity histograms within the cylinder masks (Fig. 2 red) of the different methods are shown in Fig. 5. We observe that the intensity histogram of the FBP and the Huber regularizer are more spread out than that of the proposed regularizers, which agrees with the noise-contrast curves in Fig. 3.

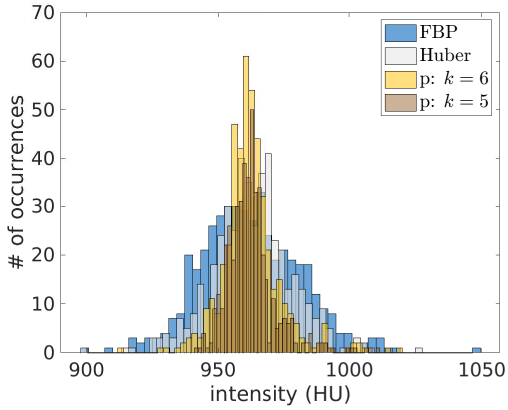


Fig. 5: Histogram of the intensity values within the thin cylinders.

IV. Discussion

The k-NN regularizer can be extended by using a more general definition for voxel differences. For example, we may replace $\boldsymbol{s}_i$ in (6) by a generic distance function, *e.g.*, the patch difference

$$\boldsymbol{s}_i = \{\|x_{\{i\}} - x_{\{j\}}\| | j \in N_i\}, \tag{15}$$

where $x_{\{i\}}$ represents a small patch around the voxel i. There is also no restriction on the neighborhood N_i being local. The voxel neighborhood N_i can be as large as the entire image. Computational concerns may limit the neighborhood to a smaller, but nonetheless nonlocal, region. Unlike other nonlocal regularizers that require a prescription of the neighborhood weights [8], the k-NN regularizer automatically determines these weighting factors: the $\underline{k}$ smallest differences have a weight of 1, the larger differences have a weight of 0.

V. Conclusions

We propose a k-NN regularizer for MBIR that penalizes the sum of $\underline{k}$ smallest (in absolute value) differences among all neighboring differences, and skip the larger differences that likely come from across organ boundaries. Doing so can potentially improve edge preservation and fine-feature detection that are important for some clinical tasks. Alternatively, such region-aware penalty could be implemented with the aid of separately obtained organ masks, *e.g.*, from a preliminary reconstruction. An advantage of the k-NN regularizer is that we can achieve the same goal without an explicit segmentation or explicit knowledge of the boundary locations.

The k-NN regularizer is nonconvex and nonsmooth. We developed a smooth approximation to the k-NN regularizer so that a gradient descent algorithm can be applied to find the MBIR solution. Our numerical studies showed that the (smoothed) k-NN regularizer can improve the noise-contrast tradeoff over the Huber regularizer in an estimation task. We also performed preliminary study of the effect of the parameter $\underline{k}$ on image reconstruction. Our quantitative results demonstrate that the k-NN regularizer is promising in improving the detectability of fine features.

References

[1] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, pp. 459–494, Aug 2014.

[2] P. Ochs, Y. Chen, T. Brox, and T. Pock, "iPiano: Inertial proximal algorithm for nonconvex optimization," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 1388–1419, 2014.

[3] R. I. Bot, E. R. Csetnek, and D.-K. Nguyen, "A proximal minimization algorithm for structured nonconvex and nonsmooth problems," *SIAM Journal on Optimization*, vol. 29, no. 2, pp. 1300–1328, 2019.

[4] T. Sun, R. Barrio, M. Rodríguez, and H. Jiang, "Inertial nonconvex alternating minimizations for the image deblurring," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 6211–6224, 2019.

[5] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, pp. 932–946, July 1995.

[6] K. C. Kiwiel, "Breakpoint searching algorithms for the continuous quadratic knapsack problem," *Mathematical Programming*, vol. 112, no. 2, pp. 473–491, 2008.

[7] V. Haase, K. Hahn, H. Schöndube, K. Stierstorfer, A. Maier, and F. Noo, "Impact of the non-negativity constraint in model-based iterative reconstruction from CT data," *Medical Physics*, vol. 46, no. 12, pp. e835–e854, 2019.

[8] H. Zhang, D. Zeng, H. Zhang, J. Wang, Z. Liang, and J. Ma, "Applications of nonlocal means algorithm in low-dose X-ray CT image processing and reconstruction: A review," *Medical physics*, vol. 44, no. 3, pp. 1168–1185, 2017.

Learned Texture Prior for Bayesian Reconstruction of Super-resolution CT

Yongyi Shi, Yongfeng Gao, Jiaxing Tan, Siming Lu, Xuanqin Mou and Zhengrong Liang

Abstract—The spatial resolution of currently reconstructed images is limited by blurring effects due to finite detector elemental, source focal size, reconstruction algorithms or other factors. To enhance the spatial resolution and maintain the useful texture information, in this paper we propose a novel single image super resolution CT (SRCT) reconstruction method, which incorporates a learned tissue-specific texture prior into a Bayesian SRCT image reconstruction framework. Specifically, the SRCT problem can be formulated as a down-sampling kernel and a Gaussian blur kernel on the desired high-resolution CT images. Using a convolutional neural network (CNN) model, we learnt the texture priors from four main tissue types of the previous high-resolution CT chest images, i.e. the lung, bone, fat and muscle, and then incorporate the learned texture prior into the Bayesian framework. Hence, a higher resolution image can be directly reconstructed from lower resolution projection data with the measurement fidelity and learned texture prior information. We refer our proposed method as SRCT-CNN. Evaluation of SRCT-CNN was performed on traditional CT simulated data with comparison to the well-known low rank and total variation (LRTV) image post-processing. SRCT-GMRF (similar with SRCT-CNN, but replace the learned texture prior with Gaussian Markov random field prior) also employed as a comparison method. The proposed SRCT-CNN showed superior performance, in terms of resolution recovery and texture enhancement.

Index Terms—Learned Texture Prior, Bayesian Reconstruction, Super Resolution Computed Tomography.

This work was partially supported by the NIH/NCI Grant CA206171. Y. Shi is also supported partially by China Scholarship Council (No. 201806280297). X. Mou is supported by the National Key Research and Development Program of China (No. 2016YFA0202003) and the National Natural Science Foundation of China (NSFC) (No. 61571359).

Y. Shi is with the Department of Radiology, State University of New York at Stony Brook, Stony Brook, NY 11794, USA, and is also with the Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China.

Y. Gao and Jiaxing Tan with the Department of Radiology, State University of New York at Stony Brook, Stony Brook, NY 11794, USA.

S. Lu is with the Department of Radiology, Computer Science and Biomedical Engineering, State University of New York at Stony Brook, Stony Brook, NY 11794, USA.

X. Mou (corresponding author) are with the Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China (e-mail: xqmou@mail.xjtu.edu.cn).

Z. Liang (corresponding author) is with the Departments of Radiology, Electrical and Computer Engineering, Computer Science and Biomedical Engineering, State University of New York at Stony Brook, Stony Brook, NY 11794, USA (e-mail: Jerome.Liang@sunysb.edu).

I. Introduction

THE spatial resolution of a reconstructed computed tomography (CT) image is limited by blurring effects due to finite detector elemental, source focal size, the reconstruction algorithms and other factors [1]. The resolution of 0.5-0.625 *mm* measured at isocenter of the scanner [2], is not sufficient where some small anatomy and pathologic structure is the target, such as lung parenchyma. Super-resolution (SR) method has achieved a promising performance to recover high-resolution images from their low-resolution counterparts [3]. In 2015, Shi et al. proposed a low rank TV regularizations (LRTV) method for SR magnetic resonance image reconstruction [4]. This method formulates the SR problem as an inverse problem, which utilizes prior information to regularize the reconstruction. Although the results of these methods are encouraging, they are the post-processing methods, and inherently ignore the data consistence. Additionally, they are limited in texture features preservation while the texture characteristics are desired in many clinical tasks.

In recent years, we proposed a texture-preserving Markov random field (MRF) prior to enhance the low dose CT (LDCT) image via learning tissue-specific textures, such as muscle, fat, bone and lung, from the same patient's previous full dose CT (FDCT) scans [5-7]. Compared to the homogeneous assumption in the commonly used MRF model such as Gaussian MRF (GMRF), texture-preserving MRF prior is already proven to be an effective way to keep the image details in low-dose CT reconstruction. Since the GMRF was perform well in SR reconstruction [8], we also proven that the texture-preserving MRF prior also can be used for SRCT reconstruction [9]. In this paper, we propose a novel single-image SRCT reconstruction method which incorporates the learned tissue-specific textures into a Bayesian reconstruction framework. Our method consists of two components. The first component is the measured fidelity knowledge, which guarantees the reconstructed image consistent with the sinogram. The second component is the learned tissue-specific texture prior, which learnt by convolutional neural network (CNN) in specific tissue region, such as muscle, fat, lung and bone. We refer our proposed method as SRCT-CNN.

II. Mathod

In this paper, we propose a learned texture prior under the Bayesian reconstruction framework for super-resolution CT

imaging. We first introduce the Bayesian framework, which requires that the to-be-recovered CT image should be consistent with the input sinogram. Then we will describe the learned texture prior which learnt tissue-specific texture by CNN architecture. The overall reconstruction algorithm will be presented following.

A. *Bayesian framework for Super Resolution CT*

The measured low-resolution sinogram can be modeled as a forward projection of a blurred and down-sampled version of the desired high-resolution image. Hence, the measurement of projection can be expressed as following:

$$y = ADSx + n \quad (1),$$

where y represents the measured projection, A is the system matrix, D is a down-sampling operator, S is a blurring operator to model the point spread function (PSF), x denotes the high-resolution image that we want to recover, and n is the observation noise. To solve x, we add a regularization term to stabilize the solution. The cost function can be written as following:

$$x^* = \min_{x\geq 0}(ADSx - y)^T\Sigma^{-1}(ADSx - y) + \beta U(x) \quad (2),$$

where Σ is a weighting matrix, β is the regularization parameter, $U(x)$ is the regularization term which represents prior information on the corresponding high-resolution images. In this paper, the prior information can be learnt by a CNN model. Therefore, we introduce an auxiliary variable, which satisfies condition of $z = x$, to Eq. (2):

$$x^* = arg\min_{x\geq 0}(ADSx - y)^T\Sigma^{-1}(ADSx - y) + \beta U(z), \quad s.t. z = x \quad (3).$$

By Half Quadratic Splitting method, we have:

$$x^* = arg\min_{x\geq 0}(ADSx - y)^T\Sigma^{-1}(ADSx - y) + \beta U(z) + \frac{\lambda}{2}\|z - x\|^2 \quad (4),$$

where λ is the penalty parameter. The two variables z and x in Eq. (4) can be alternating optimized by the following iterative scheme:

$$\begin{cases} x_{t+1} = arg\,\min_x (ADSx - y)^T\Sigma^{-1}(ADSx - y) + \frac{\lambda}{2}\|x - z_t\|^2 \\ z_{t+1} = arg\,\min_z \frac{\lambda}{2}\|z - x_{t+1}\|^2 + \beta U(z) \end{cases} \quad (5),$$

where t represents the number of iterations. The two variables z and x can be solved individually, where x could be updated by the well-known penalized reweighted least squares (PWLS) algorithm. In Eq. (5), $U(z)$ could be substituted by $\|z - G(x)\|^2$, the optimization of z can be rewritten as following:

$$z_{t+1} = arg\,\min_z \frac{\lambda}{2}\|z - x_{t+1}\|^2 + \beta\|z - G(x_{t+1})\|^2 \quad (6),$$

where $G(x)$ was produced by CNN texture prior, which will discuss in Section IIB. Eq. (6) could be simplified as follows:

$$z_{t+1} = arg\,\min_z \|z - x_{t+1}\|^2 + \|z - G(x_{t+1})\|^2, \alpha = \frac{2\beta}{\lambda} \quad (7),$$

The solution of Eq. (7) is:

$$z_{t+1} = \frac{x_{t+1} + \alpha G(x_{t+1})}{1 + \alpha} \quad (8),$$

The flowchart of the proposed method was shown in Fig. 1.

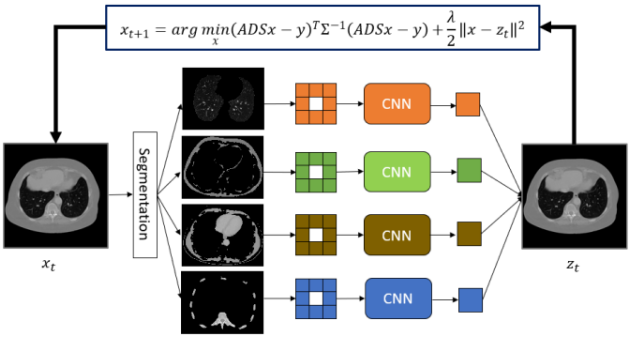


Fig. 1 The flowchart of the proposed method.

B. *Learning Tissue-specific Texture Prior*

We define the tissue-specific texture prior to be the relationship between one pixel and its neighboring pixels, where the center pixel value could be inferred by its neighbors based on the pattern specified by its texture. The analytical form of the tissue-specific texture prior could be written as:

$$w_r = arg\,\min_{w_r} \sum_k \sum_{m\in\Omega_k} \left(x_m - w_r^T x_{\Omega_k}\right) \quad (10),$$

where w_r is the texture prior weights, k is pixels index within one tissue type, m is pixel index within kth pixel surrounding region. It has demonstrated that the aforementioned tissue-specific texture prior is superior to GMRF prior in LDCT reconstruction. In this paper, we use the CNN model to learn a relationship between the center pixel and its surrounding pixels.

To train the CNN model, we first segment the corresponding high-resolution image into different tissue regions, such as lung, fat, muscle and bone, using our previously reported vector quantization algorithm [10]. The network architecture of CNN model has shown in Fig. 2. Our CNN model has one convolution layer with 16 kernels of size 3×3. After the convolution, we use rectified linear unit (ReLu) as the activation function. In the last layer, we apply one fully-connected layer with 784 and 1 neurons. The input of the model is a patch with size of 7×7 from high-resolution image with center pixel removed and to be estimated. The output is the estimated value of the center pixel.

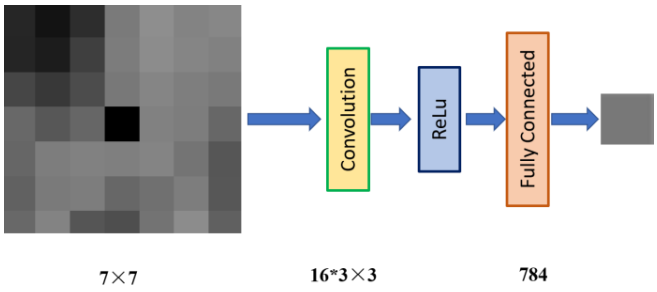


Fig. 2 The network design of the proposed CNN model.

C. *Optimization*

We can solve our proposed algorithm by alternating optimized scheme. The input image could be initialized by simultaneous algebraic reconstruction technique (SART). The workflow of SRCT-CNN can be summarized as following:

TABLE I WORKFLOW OF THE PROPOSED SRCT-CNN ALGORITHM.

Algorithm: SRCT-CNN

Learn the tissue-specific texture prior:

Segment the high-resolution image into four regions.

Learn the texture prior using the CNN model.

Bayesian reconstruction for SRCT-CNN:

Initialization x by SART.

Segment x to lung, fat, muscle and bone.

Set parameters λ, β, Σ.

While stop criterion is not met:

Update x.

for pixel j inside lung, fat, muscle and bone:

determine tissue type for pixel j ;

update z.

end

End until the stop criterion is satisfied.

III. Results

A. Experimental Data

To evaluate the performance of the proposed algorithm, one numerical clinical image was downloaded from the dataset "the NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge". A represented high-resolution chest image was selected for our simulation study, which includes 512×512 pixels with area 0.664 *mm* $\times$ 0.664 *mm* per pixel. And a fan-beam geometry with equiangularly distributed detectors was assumed. The distance from the source to the system center was 595 *mm*, and from the source to the detector was 1085.6 *mm*. Totally, 2304 projection views were uniformly collected over a full scan range. For each projection view, the original number of detectors is 736 of width 1.2858 *mm* per element, which was used to simulate a high-resolution sinogram. In this study, 368 detector elements of width 2.5716 *mm* per element was used to simulate the low-resolution sinogram. The corresponding reconstructed low-resolution image includes 256 $\times$ 256 pixels with area 1.328 *mm* $\times$ 1.328 *mm* per pixel.

B. Network Training

In our experiments, the input patches are with size of 7×7 from high-resolution image with center pixel removed and used as the training samples. We first extract the training samples from the segmented regions. There are 61313, 12942, 29801 and 6380 training samples for lung, fat, muscle and bone, respectively. To train the CNN models, we select Adam optimizer, with learning rate of 0.01, weight decay of 0.005, a momentum of 0.9 and a mini-batch size of 100. The maximum number of training epochs is set to 15. The loss curve of CNN model for lung was shown in Fig. 3. It can be observed that the loss of our CNN model decreased initially and converged rapidly. The networks were implemented in Python with the Pytorch library. A NVIDIA Geforce GTX 1080 GPU was used in this study. Since we just learn a relationship between the center pixel and its surrounding pixels, the computer time is more economical than traditional deep learning method. It only takes 197 seconds for training the CNN model in lung region.

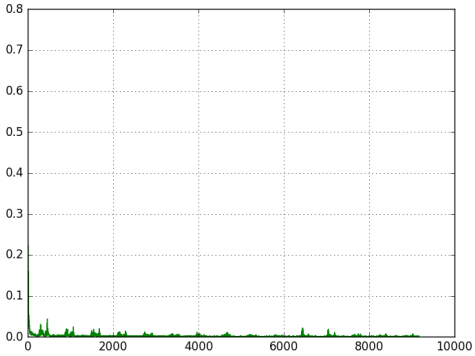


Fig. 3 The loss curve of CNN model for lung.

C. Visualization and Quantitative Evaluation

To evaluate the SRCT-CNN method, we use a high-resolution projection to reconstruct a high-resolution image as the reference. The low-resolution image is reconstructed by the measured low-resolution projection. The nearest neighbor (NN) interpolation method, LRTV method and SRCT-GMRF method were selected as the comparison methods. The results reconstructed by the aforementioned methods were shown in Fig. 3. The magnified images of selected regions of interest (ROIs) were also shown in Fig. 4. It can be seen that the jagged-edged blood vessels in NN recovered image were seriously reduced the image quality. LRTV and SRCT-GMRF methods can effectively eliminate the jagged-edged of blood vessels. However, the texture details were blurred in some extent. The proposed SRCT-CNN method best preserves the details. In Fig. 5, we showed the absolute difference images in the corresponding selected ROIs by subtracting the generated images from the reference image. It can be seen that the proposed SRCT-CNN method has the smallest error while other methods failed to keep fine detail in structure edges.

Quantitative results using the RMSE, PSNR, SSIM and FSIM indexes are shown in TABLE I for the ROI evaluation in Fig. 4. Both the LRTV and SRCT-GMRF method can remarkably improve the image quality, while the proposed SRCT-CNN method achieves the best performance in all metrics.

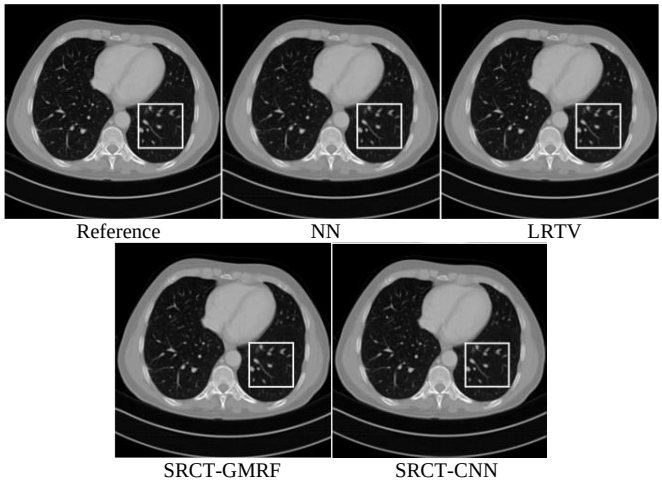


Fig. 3: The results of human chest simulation study. The display window is [0 0.035] mm^{-1}.

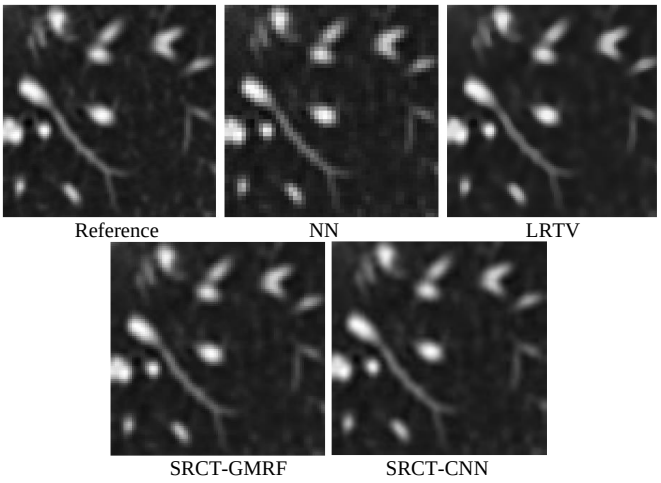


Fig. 4: The close-up views of corresponding selected regions in the white boxes in Fig. 3. The display window is [0 0.035] mm^{-1}.

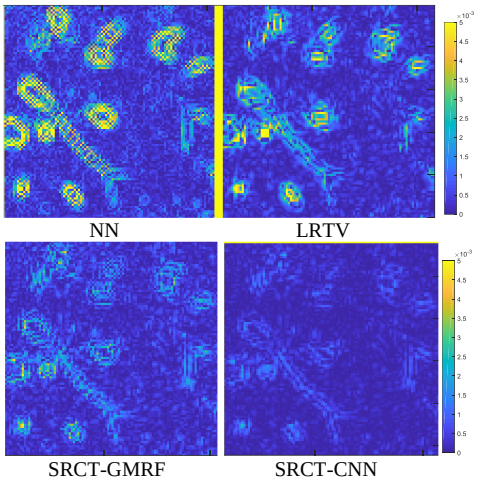


Fig. 5: The absolute difference images relative to the reference image in Fig. 4.

TABLE I
RMSE(1E-4), PSNR, SSIM AND FSIM INDEX FOR ROI IMAGES

	RMSE	PSNR	SSIM	FSIM
NN	7.9174	30.1573	0.9277	0.9633
LRTV	5.4782	33.3562	0.9694	0.9735
SRCT-GMRF	3.6501	36.8828	0.9718	0.9795
SRCT-CNN	**3.2443**	**37.9065**	**0.9746**	**0.9805**

IV. CONCLUSION

We have proposed a novel SRCT-CNN method for Bayesian reconstruction of super-resolution CT, where the learned tissue-specific texture prior is incorporated into a Bayesian reconstruction framework for the purposes of (1) the measured fidelity knowledge shall guarantee the reconstructed image is consistent with the sinogram data and (2) the learned texture prior shall enrich the tissue texture information inside the SRCT reconstruction. Experimental results from traditional CT simulated data indicate that the proposed method can improve the image quality in visualization and quantitative evaluation. More evaluation in the preclinical and clinical data is underway.

REFERENCES

[1] H. Yu and G. Wang, "Finite detector-based projection model for high spatial resolution," *Journal of X-ray Science and Technology*, vol. 20, no. 2, pp. 229-238, 2012.

[2] J. Wang and D. Fleischmann, "Improving Spatial Resolution at CT: Development, Benefits, and Pitfalls," *Radiology*, vol. 289, no. 1, pp. 261-262, 2018.

[3] S. C. Park, M. K. Park and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21-36, 2003.

[4] F. Shi, J. Cheng, L. Wang, P. T. Yap and D. Shen, "LRTV: MR image super-resolution with low-rank and total variation regularizations," *IEEE Transactions on Medical Imaging*, vol. 34, no. 12, pp. 2459-2466, 2015.

[5] H. Zhang, H. Han, J. Wang, J. Ma, Y. Liu, W. Moore et. al., "Deriving Adaptive MRF Coefficients from Previous Normal-dose CT Scan for Low-dose Image Reconstruction via Penalized Weighted Least-squares Minimization," *Medical Physics*, vol. 41, no. 4, p. 041916, 2014.

[6] H. Zhang, H. Han, Z. Liang, Y. Hu, Y. Liu, W. Moore et. al., "Extracting information from previous full-dose CT scan for knowledge-based Bayesian reconstruction of current low-dose CT images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 3, pp. 860-870, 2015.

[7] Y. Gao, Z. Liang, W. H. Moore, H. Zhang, M. J. Pomeroy, J. A. Ferretti et. al., "A feasibility study of extracting tissue textures from a previous full-dose CT database as prior knowledge for Bayesian reconstruction of current low-dose CT images," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1981-1992, 2019.

[8] R. Nayak and D. Patra, "New single-image super-resolution reconstruction using MRF model," *Neurocomputing*, vol. 293, pp. 108-129, 2018.

[9] Y. Shi, Y. Gao, X. Mou and Z. Liang, "Bayesian Approach to Super-resolution CT Reconstruction by Texture Prior Learning from Previous High-resolution CT Images," *IEEE Nuclear Science Symposium and Medical Imaging Conference*, Manchester, November, 2019.

[10] H. Han, L. Li, F. Han, B. Song, W. Moore and Z. Liang, "Fast and Adaptive Detection of Pulmonary Nodules in Thoracic CT Images Using a Hierarchical Vector Quantization Scheme," *IEEE Transactions on Information Technology in BioMedicine*, vol.19, no.2, pp. 648-659, 2015.

Limited-Angle CT Reconstruction with Random Walk Regularization

Xiaojuan Deng, Xing Zhao and Hongwei Li*

***Abstract*—Limited-angle computed tomography (CT) is a prospective nondestructive testing technique for many applications. However, it is a challenging task, because incomplete projection data can cause artifacts when a conventional CT reconstruction algorithm is applied. Therefore, many endeavors have been made to tackle this problem, and numerous algorithms have been proposed. But there are still some difficulties to be solved, such as the inconsistency between polyenergetic spectrum of x-ray in actual CT systems and the linear reconstruction model which often leads to the well-known beam hardening artifacts impairing the effectiveness of many reconstruction methods. In this work, we propose an iterative algorithm for reconstruction from incomplete polyenergetic data, using random walk penalty, which utilizes visible edges as priors to recover invisible edges and promotes piecewise-smooth reconstruction. The effectiveness of the proposed method is validated by both simulation and actual CT data sets.**

***Index Terms*—limited-angle CT, iterative reconstruction, random walk regularizer.**

I. Introduction

X-ray computed tomography (CT) plays a very important role in many fields, such as medical imaging [1] and industrial nondestructive testing [2], as it can provide a multi-dimensional view of the internal structure information of an object. The traditional reconstruction algorithms require that the projection data is complete, i.e the scanning range is at least a π plus fan angle range for fan-beam CT scan or a π range for parallel-beam CT scan [3]. However, this requirement can not be satisfied in some applications, e.g. the digital breast tomosynthesis [4] and the inspection of a plate-shaped object, which lead to the limited-angle projection data.

Limited-angle CT reconstruction is very challenging due to its high ill-posedness. The application of conventional reconstruction algorithms such as filtered back-projection (FBP) and (simultaneous) algebraic reconstruction technique (ART or SART) for limited-angle projection data can cause heavy artifacts in the reconstructed images. To overcome this problem, both theories and reconstruction methods for limited-angle CT are studied [5]–[7] in the past decades. These methods can be roughly divided into two categories: (A) projection data preprocessing approaches; (B) iterative reconstruction with proper regularizations.

Thanks for the support of the National Natural Science Foundation of China (NSFC) (61971292, 61827809 and 61871275). (Corresponding author: Hongwei Li.)

Xiaojuan Deng, Xing Zhao and Hongwei Li are with the school of Mathematical Sciences, Capital Normal University, Beijing, 100048, China, and are also with Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, 100048, China.

E-mail: hongwei.li91@cnu.edu.cn.

An obvious advantage of the methods in category (A) is that the inpainting and reconstruction procedures could be performed quite efficiently. However, a small error caused by the interpolation procedure on projection data should introduce extra artifacts in the reconstruction images.

Iterative reconstruction methods in category (B) are more popular thanks to their high flexibility in model designing and better artifacts suppression property. For example, the CT images can be approximated well by piecewise constant functions in certain applications, and its sparsity in the gradient transform domain can be integrated into the reconstruction procedure as regularized terms, such as total variation (TV) l_1 (l_0) norm [8] and wavelet tight frame [9]. There are some other methods introducing more prior knowledge to further improve the quality of the reconstructed images, e.g. prior image constrained compressed sensing (PICCS) [10] and multi-domain regularization [11] based methods. These methods can improve the quality of reconstructed images from limited-angle projection data significantly.

However, these popular regularizers are primarily designed based on the priors of the reconstructed images or projection data, and little consideration has been given to the theory of visible and invisible singularities of limited-angle data, developed by Quinto et al. [6]. Besides, most work on the limited-angle CT reconstruction assumes the linear model, which requires the monoenergetic projection data. However, the measured data are obtained from the polychromatic x-ray beam. This inconsistency with the linear model produces the well-known beam hardening artifacts, which impair the effectiveness of many reconstruction methods.

In fact, in many practical applications, CT images with beam hardening artifacts can be approximated by piecewise smooth functions. Based on this key observation and the theory of visible and invisible boundaries, a reconstruction model incorporating random walk is proposed in this work for limited-angle reconstruction. These random walk penalty functions are designed to achieve edge-preserving smoothing such that the visible edges are used as anchors to recover piecewise smooth reconstructions.

The remainder of this paper is organized as follows. The reconstruction model incorporating random walk as regularizers will be introduced in section II. Section III dedicates to experiments with simulated and real data. Conclusions will be given in Section IV.

II. Methodology

In this section, we first introduce the motivation of our proposed method and then present the random walk regularized

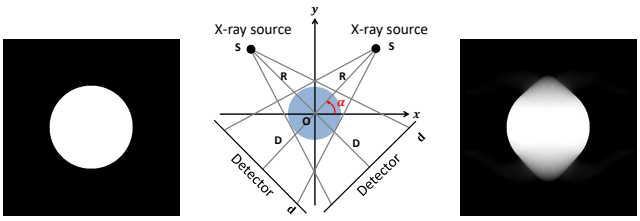
TABLE I
GEOMETRICAL PARAMETERS OF THE DISK PHANTOM

Parameter	Value
Distance of X-ray source to rotation center	500 mm
Distance of rotation center to detector	500 mm
Width of detector unit	0.25 mm
Number of detector units	1024
Scanning Angular Interval	0.5 degree

reconstruction method for limited-angle CT in detail.

A. Motivation

We first demonstrate the limited-angle artifacts with a numerical experiment on a disk phantom size of 512×512, shown in Fig. 1(a). Without loss of generality, a fan-beam CT scanning configuration for the image projection is considered in this work, and it should be easy to extend our idea to the parallel-beam CT or even 3D cone-beam CT. With the given scanning configuration shown in Fig. 1(b) and the parameters listed in Table I, the reconstructed image from limited-angle data (i.e. $\alpha \in [\pi/4, 3\pi/4]$) by SART is shown in Fig. 1(c). One can see that the serious blurring and distortion are approximately along the $y-$direction, i.e. the edges tangent to the missing x-rays are invisible. On the other hand, the edges approximately perpendicular to the $y-$direction, i.e. tangent to the x-rays are visible and more faithfully than that of other points.



(a) Phantom (b) Configuration (c) SART

Fig. 1. The images of a disk phantom reconstructed with the SART method from limited-angle scan data. (a) The disk phantom; (b) the scanning configuration; (c) the limited-angle reconstruction from 90-degree data ($[\pi/4, 3\pi/4]$) with SART (10 iterations). The display window is set to [0 1].

Indeed, it is a widely adopted assumption that the ideal CT image can be modeled as a piecewise constant function in many applications. So the key observation of visible and invisible edges motivates a promising reconstruction model for limited-angle CT, which is named as alternating edge-preserving diffusion and smoothing (AEDS) [7]. Let A denote as the projection matrix and b as the acquired data, the AEDS model can be formulated as

$$\min_u \frac{1}{2}\|Au - b)\|^2 + \lambda_1\phi(u) + \lambda_2\psi(u), \qquad (1)$$

where the u is the linear attenuation coefficient. A typical choice for $\phi(\cdot)$ is $\|\nabla_x u\|_0$ or $\|\nabla_x u\|_1$, where $\nabla_x u$ denotes the gradient operator along the $x-$direction. This regularization is designed to fix the trustworthy visible edges and diffuse their values to other areas to obtain a better approximation to the ideal image. But it might introduce horizontal striping artifacts (lines). So the other regularizer $\psi(\cdot)$ is designed to suppress such possible artifacts and a similar choice for $\psi(\cdot)$ is $\|\nabla_y u\|_0$ or $\|\nabla_y u\|_1$ or even one-dimensional median filter along the $y-$direction.

Even the AEDS model achieves very promising results compared to existing algorithms in the literature, it still suffers from some limitations: parameters tuning is tricky, too many iterations are needed, and it also relies heavily on the piecewise-constant assumption. In practical applications, the reconstructed image is always affected by beam hardening or metal artifacts, so, the image does not always satisfy the piecewise-constant assumption.

B. Random walk regularized reconstruction method

To address the above-mentioned limitations, we replace the l_0 (l_1) norm regularization with random walk. Motivated by the continuous random walk model [12], the one-dimensional random walk smoothing can be achieved by minimizing the following problem

$$\min_u \frac{\lambda}{2}\int_\Omega q(u_x)|u - f|^2 d\Omega + \frac{1}{2}\int_\Omega g(f_x)(u_x)^2 d\Omega, \qquad (2)$$

where f denotes the acquired image, u is the image to be recovered, f_x, u_x denote the $x-$direction derivative. The functions $q(\cdot)$ and $g(\cdot)$ are defined as follow

$$q(\cdot) = |\cdot|^\alpha, \qquad g(\cdot) = exp(-\beta|\cdot|^c),$$

where α, β and c are the scalar parameters greater than 1. For the edge points, big values would be calculated by $q(\cdot)$ and small values by $g(\cdot)$. The first term of (2) requires that the recovered image u should be similar to the obtained image f, especially at the edges, while the second term of (2) asks for a smooth field. So, the solution of the problem (2) would balance these two terms and try to fix the edge points of f while propagating their values to non-edge points. This is just what we need.

So, we would like to combine the random walk smoothing and the SART method into an iterative algorithm to recover the true image. Let denote $u^{(n)}$ as the result after the nth iteration, we first update the image $u^{(n)}$ to $u^{(n+1/3)}$ by the SART reconstruction algorithm from the obtained projection data, with $u^{(n)}$ as its initial guess, and the second step is seeking $u^{(n+2/3)}$ from $u^{(n+1/3)}$ by solving $\min_u E_x(u)$, where $E_x(u)$ is defined as following

$$E_x(u) = \frac{\lambda}{2}\int_\Omega q(u_x)|u - u^{(n+1/3)}|^2 d\Omega + \frac{1}{2}\int_\Omega g(u_x^{(n+1/3)})(u_x)^2 d\Omega. \qquad (3)$$

The third step updates $u^{(n+2/3)}$ to $u^{(n+1)}$ by solving $\min_u E_y(u)$, where $E_y(u)$ is defined as following

$$E_y(u) = \frac{\lambda}{2}\int_\Omega q(u_y)|u - u^{(n+2/3)}|^2 d\Omega + \frac{1}{2}\int_\Omega g(u_y^{(n+2/3)})(u_y)^2 d\Omega. \qquad (4)$$

Denote SART($u^{(n)}$, b) as the procedure of SART algorithm with the projection b and initial value $u^{(n)}$, so, the proposed reconstruction method can be summarized as following **Algorithm 1**.

Algorithm 1 The proposed method

Given an initial $u^{(0)}$ and the projection data b
$n := 0$
while (stop criterion is not met) **do**
  Update image by
  $u^{(n+1/3)} = \text{SART}(u^{(n)}, b)$ (one iteration)
  Update image by solving the optimization problem
  $u^{(n+2/3)} = \arg\min_u E_x(u, u^{(n+1/3)})$
  Update image by solving the optimization problem
  $u^{(n+1)} = \arg\min_u E_y(u, u^{(n+2/3)})$
  $n := n + 1$
end while

C. Numerical algorithm

For the proposed **Algorithm 1**, the numerical algorithm for solving the second step is the same as the third one. The Euler-Lagrange equation of the problem (2) reads

$$\lambda q(u_x)(u - f) - g(f_x)u_{xx} = 0. \tag{5}$$

Given the second order approximation of u_{xx},

$$u_{xx} = \frac{u(x+h) - 2u(x) + u(x+h)}{h^2} + \mathcal{O}(h^2), \tag{6}$$

the discretion of Euler-Lagrange equation (5) reads

$$-g_{i-\frac{1}{2},j}u_{i-1,j} + \left(\lambda q_{i,j}h^2 + g_{i-\frac{1}{2},j} + g_{i+\frac{1}{2},j}\right)u_{i,j} - g_{i+\frac{1}{2},j}u_{i+1,j} = \lambda q_{i,j}h^2 f_{i,j}. \tag{7}$$

This is a tridiagonal linear system and can be solved efficiently by the Thomas algorithm. Note that boundary treatment is required when u_{xx} and f_x approach the boundary nodes. In this study, the Newman boundary condition is always adopted.

III. Experiments

In this section, we perform numerical experiments with simulated data as well as real data, against the SART and AEDS methods to verify the effectiveness of the proposed reconstruction method. The simulated projection is obtained by the open source CT simulator (see: http://www.ctsim.org) and the real data is obtained by an industrial CT system located in our lab. Both the analytic simulation and real data could remove the inverse crime and they would help to demonstrate the proposed method for its tolerance to imaging model perturbations in many applications.

Parameter selection is always a challenging problem for iterative reconstruction algorithms. In this study, the parameters have been tuned to achieve their best performance in terms of artifacts removal and structure-preserving by sampling the parameter space. Additionally, to start the proposed algorithm, SART with 10 iterations is applied on the limited data to obtain an initial guess. This is done in all the experiments.

A. Simulated data experiment

It is known that if the forward projection procedure in the image reconstruction algorithm is the same as that of the projection simulation procedure, one can usually get over-positive reconstruction results, which is known as inverse crime. To avoid such a problem, the analytic phantom is employed and the polychromatic property of x-ray is taken into consideration to demonstrate the behaviors of the random walk regularized reconstruction method. In the simulation, the x-ray spectrum used in the experiments is simulated with the free software SpectrumGUI (see: https://sourceforge.net/projects/spectrumgui/). The voltage is set to 80 kV. The phantom is treated as a composition of copper and titanium materials with different densities. The mass attenuation coefficients of the materials, copper and titanium, are obtained from NIST (see: https://physics.nist.gov/PhysRefData/XrayMassCoef/tab4.html).

Fig. 2(a) shows the rasterized image of the designed rhombus phantom by using the CTSim software. The scanning configuration illustrated in Fig. 1(b) is adopted in the simulation. Fig. 2(a) is a very challenging phantom for limited-angle problem, because the two triangles are stacked together with a designed tilt angle 5 degrees, such that this edge is completely invisible. The full-angle SART reconstruction from the polychromatic projection shows obvious beam-hardening artifacts, and additional blurring and distortions are introduced by limited-angle SART reconstruction, see Fig. 2(c) and (d). Fig. 2(e) and (f) show the reconstructions from limited-angle polychromatic projection by using AEDS(l_0, l_0) and the proposed method, respectively. One can find that, the performance of AEDS(l_0, l_0) method is seriously affected by the beam-hardening and new artifacts are produced in the triangle region. The proposed method recovered two triangles better, although the tilt edge is not completely accurate.

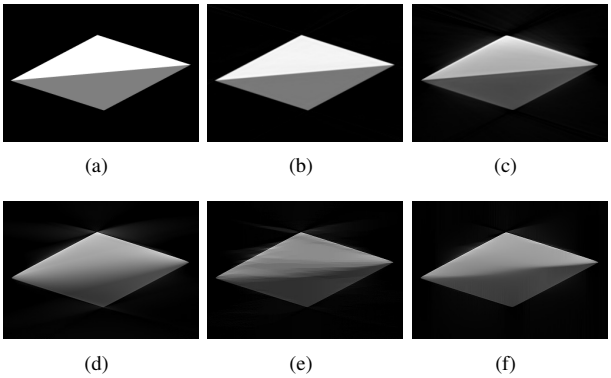
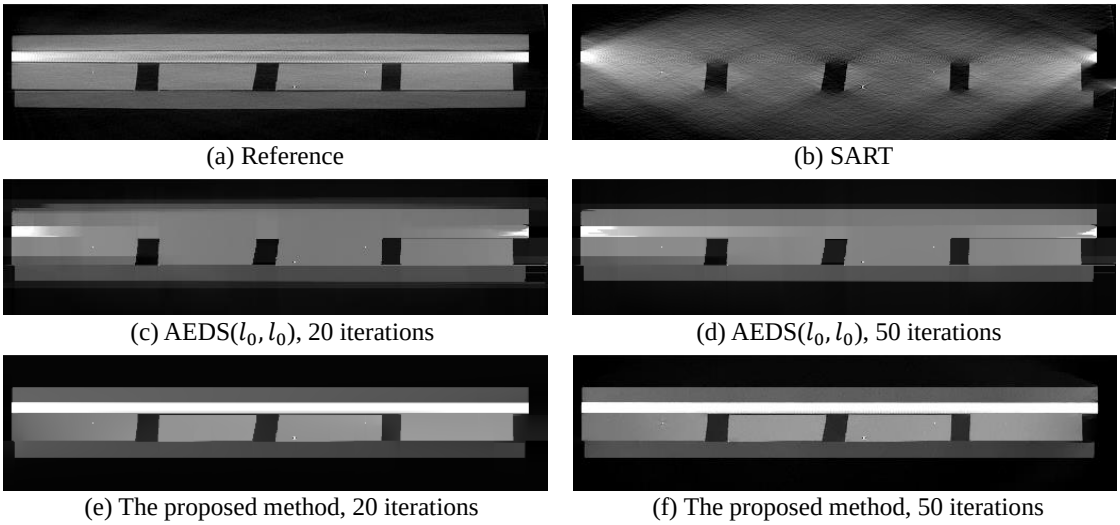


Fig. 2. (a) Rasterized image of the designed rhombus phantom by using CTSim; (b)-(c) full-angle SART reconstructions from the monochromatic and polychromatic projection, respectively; (d)-(f) limited-angle ($[\pi/6, 5\pi/6]$) reconstructions from the polychromatic projection by using SART, AEDS(l_0, l_0) and the proposed method, respectively. The display window is set to [0, 8].

B. Real data experiment

To further test the potential capability of the proposed method in practical applications, experiments are carried out



(a) Reference (b) SART

(c) AEDS(l_0, l_0), 20 iterations (d) AEDS(l_0, l_0), 50 iterations

(e) The proposed method, 20 iterations (f) The proposed method, 50 iterations

Fig. 3. (a) The reference image reconstructed from the full-angle data; (b) reconstructed image from 120-degree data ($[\pi/6, 5\pi/6]$) with SART (10 iterations); (c)-(d) show reconstructions with AEDS(ℓ_0, ℓ_0) at 20 and 50 iterations, respectively; (e)-(f) show reconstructions with the proposed method at 20 and 50 iterations, respectively. The display window is set to [0, 0.06]

on a real flat object. The x-ray source is YXLON-FXE-225.48, the detector is Varian PS2520V. The voltage and the current used are 140 kV and 160 mA, respectively. The number of detector cells is 1920, and the length of each detector cell is 0.127 mm. The total number of views is set to 720, which are evenly spanned on a circular track of 360 degrees. The distance of x-ray source to rotation center is set to 311.19mm and the distance of rotation center to detector is set to 697.82mm. Please note that the detector used is a panel detector, but in this study, we have extracted the data corresponding to the central plane to mimic a fan beam CT scan.

A full angular CT scan is carried out, from which we can reconstruct image as the reference, see Fig. 3(a). Limited-angle reconstruction with SART, AEDS(l_0, l_0) and the proposed method with different iterations are shown in Fig. 3(b)-(f), respectively. After 50 iterations, some blurring and distortions are still remained in the reconstruction of AEDS(l_0, l_0), please check Fig. 3(d). Fortunately, the proposed method can result in a satisfactory reconstruction in 20 iterations. After 30 iterations, the reconstruction does not make observable change, see Fig. 3(e) and (f).

It should be pointed out that, the proposed algorithm can still work well for biological scans, but low contrast of the structures will increase the difficulty of limited angle reconstruction.

IV. Conclusion

A reconstruction model regularized by the random walk penalty is proposed, which tries to consider the property of visible and invisible edges of limited-angle CT reconstructions. The random walk regularization is introduced to better handle the beam-hardening artifacts. Numerical experiments show that the proposed method is superior to state-of-the-art algorithms, which produce very promising reconstructions on polychromatic limited-angle data. Compared to the AEDS method, it's easy to obtain good reconstructions with fewer iterations.

References

[1] L. T. Niklason, B. T. Christian, L. E. Niklason, D. B. Kopans, D. E. Castleberry, B. Opsahl-Ong, C. E. Landberg, P. J. Slanetz, A. A. Giardino, R. Moore *et al.*, "Digital tomosynthesis in breast imaging." *Radiology*, vol. 205, no. 2, pp. 399–406, 1997.

[2] K. Tam, J. Eberhard, and K. Mitchell, "Incomplete-data CT image reconstructions in industrial applications," *IEEE Transactions on Nuclear Science*, vol. 37, no. 3, pp. 1490–1499, 1990.

[3] D. L. Parker, "Optimal short scan convolution reconstruction for fan beam CT," *Medical Physics*, vol. 9, no. 2, pp. 254–257, 1982.

[4] Y. Zhang, H.-P. Chan, B. Sahiner, J. Wei, M. M. Goodsitt, L. M. Hadjiiski, J. Ge, and C. Zhou, "A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis," *Medical Physics*, vol. 33, no. 10, pp. 3781–3795, 2006.

[5] E. T. Quinto, "Singularities of the X-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$," *SIAM Journal on Mathematical Analysis*, vol. 24, no. 5, pp. 1215–1225, 1993.

[6] ——, "Artifacts and visible singularities in limited data X-ray tomography," *Sensing and Imaging*, vol. 18, no. 1, p. 9, 2017.

[7] J. Xu, Y. Zhao, H. Li, and P. Zhang, "An Image Reconstruction Model Regularized by Edge-preserving Diffusion and Smoothing for Limited-angle Computed Tomography," *Inverse Problems*, vol. 35, no. 8, p. 085004, 2019.

[8] T. Wang, K. Nakamoto, H. Zhang, and H. Liu, "Reweighted anisotropic total variation minimization for limited-angle CT reconstruction," *IEEE Transactions on Nuclear Science*, vol. 64, no. 10, pp. 2742–2760, 2017.

[9] C. Wang, L. Zeng, Y. Guo, and L. Zhang, "Wavelet tight frame and prior image-based image reconstruction from limited-angle projection data," *Inverse Problems & Imaging*, vol. 11, no. 6, pp. 917–948, 2017.

[10] C. Gong, L. Zeng, and C. Wang, "Image reconstruction model for limited-angle ct based on prior image induced relative total variation," *Applied Mathematical Modelling*, vol. 74, pp. 586–605, 2019.

[11] Z. Wei, B. Liu, B. Dong, and L. Wei, "A joint reconstruction and segmentation method for limited-angle x-ray tomography," *IEEE Access*, vol. 6, pp. 7780–7791, 2018.

[12] M. Li, H. Gao, F. Zuo, and H. Li, "A Continuous Random Walk Model with Explicit Coherence Regularization for Image Segmentation," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1759–1772, 2018.

Cone-beam reconstruction from n-sin trajectories with transversely-truncated projections

Nicolas Gindrier, Rolf Clackdoyle, Simon Rit and Laurent Desbat

***Abstract*—In cone-beam tomography, we define the n-sin source trajectory as having n periods of a sinusoid traced on an imaginary cylinder enclosing the object. A 2-sin is commonly known as a saddle, and it is known that the convex hull of a saddle is the same as the union of all of its chords. The convex hull of a closed trajectory is the Tuy region, where cone-beam reconstruction is possible if there are no truncated projections. However, with truncated projections, the method of differentiated backprojection and Hilbert inversion can be applied along a chord if the chord is visible (not truncated) in the projections. Here, we consider a particular transaxial truncation which prevents chords from always being visible, but we establish that the more powerful method of M-lines can be applied to ensure reconstruction in the reduced field-of-view. The 3-sin, on the other hand, has a Tuy region which is not filled by its chords, and we do not have any cone-beam theory to determine if reconstruction is possible with transverse reconstruction. In our preliminary numerical experiment, the 3-sin seemed to perform equally well as the 2-sin trajectory even though there were no chords passing through the slice we examined. We tentatively suggest that there might be other, yet unknown theory that explains why 3-sin reconstruction is possible with the specified transaxial truncation. We believe that these results on transverse truncation and reconstruction from 2-sin and 3-sin trajectories are new.**

I. Introduction

A cornerstone of cone-beam (CB) reconstruction theory is Tuy's condition. For an object which is contained in some known convex region Ω, a point $x \in \Omega$ is said to satisfy Tuy's condition if every plane passing through the point x intersects the source trajectory non-tangentially [1]. Given a source trajectory (always assumed to not intersect Ω), it was proved that stable image reconstruction could be achieved for the whole Tuy region provided that none of the cone-beam projections was truncated [1] (not just the Tuy region, but none of Ω is allowed to be truncated). Conversely, Finch showed that all points of Ω outside the Tuy region cannot be stably (and accurately) reconstructed (unless strong a priori information is available) [2]. Now, aside from the cases amenable to differentiated backprojection (DBP) techniques, see below, there is no current theory to systematically indicate which part of the Tuy region can be reliably reconstructed if some projections are truncated. In a restricted context, we explore this problem here. In view of Finch's result, we are only interested in Ω_0 which is the intersection of Ω with the Tuy region. For simplicity, we assume this (convex) Ω_0 is a cylinder with elliptical base and finite height.

Our context is a specific set of source trajectories, and a specific pattern of (transverse) truncation. The source trajectories under consideration in this work are "n periods of a sinusoid on the surface of a cylinder" (n-sin) where $n \in \mathbb{N}, n > 1$. A 2-sin trajectory is often called a "saddle" although the term saddle can refer to more general curves, see [3]. Nearly all the theoretical results here will apply to more general closed trajectories. For example, the convex hull of any n-sin trajectory (and more generally, of any closed trajectory) is the same as the Tuy region [2]. For the specific pattern of truncation, we consider a cylindrical reduced-field-of-view (RFOV) whose axis is parallel to the axis of the n-sin trajectory and whose radius is small enough that the RFOV does not contain all of Ω_0. The RFOV is positioned to intersect the boundary of Ω_0. The measured rays for each projection are exactly those that intersect the RFOV. Fig. 1 provides an example illustration of a 2-sin trajectory with a cylindrical Ω_0 and a possible RFOV.

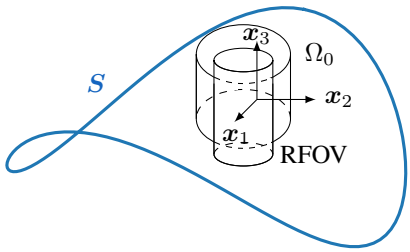


Fig. 1. A 2-sin trajectory with a cylindrical Ω_0 and a RFOV (taller than Ω_0).

We appeal to two methods of reconstruction when CB projections are truncated, and we refer to them here as the "DBP-chord method" and the "M-line method". Given a source trajectory S, parametrized by λ, a chord $C_{i,j}$ is a line joining $\boldsymbol{S}_{\lambda_i}$ to $\boldsymbol{S}_{\lambda_j}$ (Fig. 2 shows an example chord and an M-line). "DBP" refers to the operation or the 3D image obtained after CB backprojection of the differentiated projections. In the DBP-chord method, the DBP is performed along a sub-trajectory $\boldsymbol{S}_{[\lambda_1,\lambda_2]}$ and reconstruction takes place along the chord $C_{1,2}$ in the DBP image, followed by an inversion of the Hilbert transform. An M-line is a half-line starting at some source point $\boldsymbol{S}_{\lambda_M}$ and passing through Ω_0; see Fig. 2. In the M-line method, the Hilbert inversion takes place along a segment of the M-line. The M-line method is more powerful than the DBP-chord method, but is more complicated.

In this work, we will show that the DBP-chord method is

N. Gindrier, R. Clackdoyle and L. Desbat are with the TIMC-IMAG laboratory, CNRS UMR 5525 and Université Grenoble Alpes 38000 Grenoble, France. e-mail: nicolas.gindrier@univ-grenoble-alpes.fr

S. Rit is with Univ. Lyon, INSA-Lyon, UCB Lyon 1, UJM-Saint Etienne,CNRS, Inserm, CREATIS UMR5220, U1206, Centre Léon Bérard, F-69373, LYON, France.

This work was supported by the "Fonds unique interministériel" (FUI) and the European Union FEDER in Auvergne Rhône Alpes (3D4Carm project) and by the ANR (ROIdoré project ANR-17-CE19-0006-01).

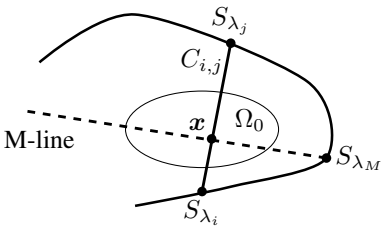


Fig. 2. A chord $C_{i,j}$ and an M-line from S_{λ_M}.

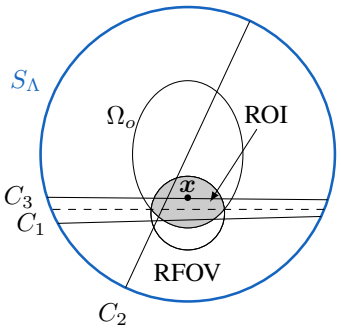


Fig. 3. Top-view for the configuration of Fig. 1. The dashed line delimits the 3 cases of reconstruction for chords containing a point in the ROI: both endpoints of C_1 are below the dashed line (C_1 intersects the support before the RFOV) therefore the inverse of the finite Hilbert transform can be used, C_2 crosses the dashed line then the one-endpoint Hilbert transform must be used, and both endpoints of C_3 are above the dashed line (C_3 intersects the RFOV before the support) line so there is no reconstruction from the chord method (from eq. (5)). To reconstruct $\boldsymbol{x}$, belonging to C_3, the M-line method must be used.

not strong enough to handle the cylindrical RFOV truncation problem, but that M-line method is suitable in the case of the 2-sin trajectory. We believe this result to be new. We also demonstrate that neither the DBP-chord method nor the M-line method is suitable for certain regions inside the 3-sin trajectory. We then report on numerical experiments performed to test whether accurate reconstruction is likely to be achieved anyway, possibly due to some currently unknown theory. We point out that the simulations use an iterative reconstruction method, the goal being to test conditions for reconstruction, and not the M-line method, whose role here is purely theoretical.

Similar results on using the M-line method to verify reconstruction theory for the cylindrical RFOV have been published for the case of a helical trajectory [4]. Note the helix is not a closed trajectory, which results in different considerations than for the proof that the 2-sin trajectory admits full (accurate and stable) reconstruction throughout $\Omega_0 \cap \text{RFOV}$.

II. Theory

A. Definitions and standard results

The unknown density function is denoted $f(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, x_2, x_3)$. For fixed height H and radius R ($H, R > 0$) and for $n \in \mathbb{N}, n > 1$ (usually n =2 or 3) we define the n-sin trajectory as

$$S_\Lambda^n = \{(R\cos\lambda, R\sin\lambda, H\cos(n\lambda))^T | \lambda \in \Lambda = [0, 2\pi)\}. \quad (1)$$

For each point $\boldsymbol{S}_\lambda$ on the source trajectory, we define the CB projection with the usual formula:

$$g(\lambda, \boldsymbol{\eta}) = \int_0^{+\infty} f(\boldsymbol{S}_\lambda + l\boldsymbol{\eta})\mathrm{d}l, \lambda \in \Lambda. \quad (2)$$

For non-truncated projections, we would have all values of $\boldsymbol{\eta}$ in the unit sphere, but here we only allow those $\boldsymbol{\eta}$ such that the half-line $L_{\boldsymbol{S}_\lambda, \boldsymbol{\eta}} = \{\boldsymbol{S}_\lambda + l\boldsymbol{\eta} | l \in [0, \infty]\}$ intersects the RFOV.

The operation of DBP takes a segment of the source trajectory and forms a 3D image by differentiation and weighted backprojection:

$$b_{1,2}(\boldsymbol{x}) = \frac{1}{\pi}\int_{\lambda_1}^{\lambda_2} \frac{1}{\|\boldsymbol{x} - \boldsymbol{S}_\lambda\|} \frac{\partial}{\partial\lambda} g(\lambda, \boldsymbol{\eta})\Big|_{\boldsymbol{\eta} = \frac{\boldsymbol{x} - \boldsymbol{S}_\lambda}{\|\boldsymbol{x} - \boldsymbol{S}_\lambda\|}} \mathrm{d}\lambda. \quad (3)$$

Note that $b_{1,2}(\boldsymbol{x})$ can always be calculated if $\boldsymbol{x} \in$ RFOV but if $\boldsymbol{x} \notin$ RFOV, then $b_{1,2}(\boldsymbol{x})$ is only available in rare fortuitous geometric circumstances, such as the entire subtrajectory $S_{[\lambda_1, \lambda_2]}$ located on the "other side" of the RFOV with respect to the location of $\boldsymbol{x}$.

Both the DBP-chord and M-line methods achieve image reconstruction using an intermediate Hilbert transform. For Hilbert transforms of 3D functions, we specify a direction $\boldsymbol{\theta}$ and use the following formula

$$\mathcal{H}_{\boldsymbol{\theta}} f(\boldsymbol{x}) = \int_{-\infty}^{+\infty} \frac{f(\boldsymbol{x} - s\boldsymbol{\theta})}{\pi s}\mathrm{d}s \quad (4)$$

where the integral in equation (4) is in the sense of the Cauchy principal value.

B. The DBP-chord method

The DBP-chord method is based on a link between the DBP image and the Hilbert transform. If $\boldsymbol{x} \in C_{i,j}$ (and $\boldsymbol{x} \in$ RFOV), then it is well-known (e.g. [5]) that

$$b_{i,j}(\boldsymbol{x}) = 2\mathcal{H}_{\boldsymbol{\eta}_{i,j}} f(\boldsymbol{x}) \quad (5)$$

where $\boldsymbol{\eta}_{i,j} = (\boldsymbol{S}_i - \boldsymbol{S}_j)/\|\boldsymbol{S}_i - \boldsymbol{S}_j\|$ in the direction from $\boldsymbol{S}_j$ to $\boldsymbol{S}_i$. "One dimensional" image reconstruction is performed along the chord segment $Q = C_{i,j} \cap (\Omega_0 \cup \text{RFOV})$, provided at least one endpoint of Q lies outside Ω_0 [6]. If both endpoints lie outside Ω_0, then an analytic formula can be used to invert $\mathcal{H}_{\eta_{i,j}} f$ to obtain f along Q (see [7]). If neither endpoint lies outside Ω_0, then the DBP-chord method fails for this chord. Fig. 3 illustrates these three possibilities.

The DBP-chord method has been known for at least 15 years, and provides a powerful method for CB reconstruction with truncated projections. The only requirement to reconstruct along a chord $C_{i,j}$ is that the chord remains visible (not truncated) in all the projections in the subtrajectory $\boldsymbol{S}_{[\lambda_i, \lambda_j]}$, and that the field-of-view is large enough that the chord extends beyond the object boundary.

C. The M-line method

For some λ_M and some direction $\boldsymbol{\eta}_M$, consider the M-line $L_{\boldsymbol{S}_M, \boldsymbol{\eta}_M}$ (where we have written $\boldsymbol{S}_M$ for $\boldsymbol{S}_{\lambda_M}$). If $\boldsymbol{x} \in L_{\boldsymbol{S}_M, \boldsymbol{\eta}_M}$ and $\boldsymbol{x}$ *also* lies on a chord $C_{1,2}$, then

$$b_{1,M}(\boldsymbol{x}) + b_{2,M}(\boldsymbol{x}) = 2\mathcal{H}_{\boldsymbol{\eta}_M} f(\boldsymbol{x}). \quad (6)$$

Note that the Hilbert direction is along the M-line, and not along the chord $C_{1,2}$.

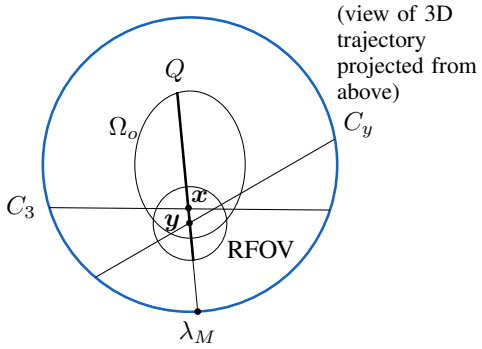


Fig. 4. The M-line method. The point $\boldsymbol{x}$ cannot be reconstructed by the DBP-chord method using its chord C_3. A suitable (one-endpoint) M-line is chosen passing through $\boldsymbol{x}$. If each $\boldsymbol{y} \in Q \cap \text{RFOV}$ lies on some chord C_y, then the Hilbert transform can be computed, and Hilbert inversion along Q is possible, to reconstruct f at $\boldsymbol{x}$ (and at all $\boldsymbol{y}$).

The M-line method is designed to handle reconstruction at a point $\boldsymbol{x}$ which lies on a chord not suitable for Hilbert inversion. In that case, an M-line through $\boldsymbol{x}$ is selected which also has the property that (at least) one endpoint of Q lies outside Ω_0, where $Q = L_{\boldsymbol{S}_M,\boldsymbol{\eta}_M} \cap (\Omega_0 \cup \text{RFOV})$. In order to invoke Hilbert inversion along Q, each point $\boldsymbol{y} \in Q \cap \text{RFOV}$ must lie on its own chord $C_{1,2}$, for in this case, (6) can be applied to obtain $\mathcal{H}_{\boldsymbol{\eta}_M} f(\boldsymbol{y})$. Once $\mathcal{H}_{\boldsymbol{\eta}_M} f(\boldsymbol{y})$ has been obtained for all $\boldsymbol{y} \in Q \cap RFOV$, then Hilbert inversion can be applied along Q to obtain $f(\boldsymbol{x})$. See Fig. 4 for an illustration of the concept.

The M-line method was introduced by Pack *et al.* [8], and although it is more complicated to apply, it is more powerful than the DBP-chord method due to the requirement that the Hilbert segment Q needs one endpoint outside Ω_0.

D. The 2-sin trajectory

The 2-sin trajectory, equation (1), also known as a saddle, has the property that its convex hull is equal to the union of all its chords [3], see Fig. 5. Thus every point in the Tuy region lies on a chord, so if there is only moderate truncation of the projections in the axial direction, all chords will be visible in the projections, and the DBP-chord method ensures reliable reconstruction throughout the Tuy region.

For transverse truncation dictated by the cylindrical RFOV, the situation is more complex, and the three cases shown in Fig. 3 arise. Any point $\boldsymbol{x}$ inside $\Omega_0 \cap \text{RFOV}$ will lie on a chord, which will be suitable for two-endpoint inversion (C_1 in Fig. 3), for one-endpoint inversion (C_2 in Fig. 3), or not suitable for Hilbert inversion (C_3). In this latter case, it is always possible to choose an M-line joining $\boldsymbol{x}$ to the trajecory, which exits Ω_0 while still inside RFOV. All points $\boldsymbol{y}$ inside the RFOV and on this M-line will lie on some chord, from which the required Hilbert transform can be calculated according to (6). Thus Hilbert inversion is possible in all three cases, and the entire region $\Omega_0 \cap \text{RFOV}$ can be reliably reconstructed in principle.

To the best of our knowledge, this is the first demonstration that region-of-interest reconstruction is possible with transverse truncation for the case of saddle trajectories. For a cylindrical RFOV intersecting the boundary of Ω_0, our argument clearly generalizes to any trajectory whose Tuy region is equal to the union of its chords. The restriction of a cylindrical RFOV can be relaxed also, but with care because combined axial and transverse truncation introduces new complications (the existence of a suitable M-line is not always ensured).

E. The 3-sin trajectory

Unlike the 2-sin trajectory, the Tuy region of the 3-sin trajectory contains points which do not lie on any chord. In the bottom row of Fig. 5, the convex hull (Tuy region) of the 3-sin curve consists of a triangular "roof" and triangular "floor" but there is a substantial triangular "hole" of locations which are not traversed by any chord, as shown bottom right of Fig. 5.

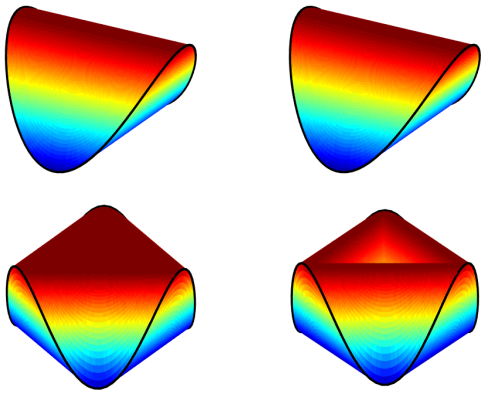


Fig. 5. The convex hull and the "chord hull" of two source trajectories. *Top*: the 2-sin trajectory. *Bottom*: the 3-sin trajectory. *Left*: the convex hull of the trajectory (the Tuy region). *Right*: the union of all chords in the trajectory.

For points in the Tuy region not lying on any chord, both the DBP-chord method and the M-line method fail, even if there were no projection truncation. However, other theory (still unknown) might indicate whether or not reconstruction is possible in regions where there are no chords. Our preliminary simulation experiments below suggest that, for cylindrical RFOV, reliable reconstruction might still be possible where chords are missing.

III. Simulations

We performed reconstructions on a 380 x 152 x 382 image array from analytically simulated projection data of the Forbild Thorax phantom. Our goal could be an investigation of the vertebrae. The reconstruction method used was the conjugate gradient method from RTK (minimizes $\|(Rf - p)\|_2^2 + \gamma\|\nabla f\|_2^2$ with R the forward projection operator and p the measured projections). In our case, $\gamma = 100$, 120 iterations were performed with 200 source positions. For both trajectories, the radius was 250 mm and the total height was 200 mm ($H = 100$ mm). The RFOV was centered on the spinal column, with a radius of 50 mm and a height of 380 mm (the height of the phantom). It is illustrated in Fig. 6. Since the phantom diameter is 300 mm, the RFOV causes substantial transverse truncations.

Two cross-sections of different heights are presented through the reconstructed phantom ($x_3 = 0$ mm and $x_3 = 90$ mm), as illustrated in Fig. 7 and Fig. 8. The aim is to see the influence

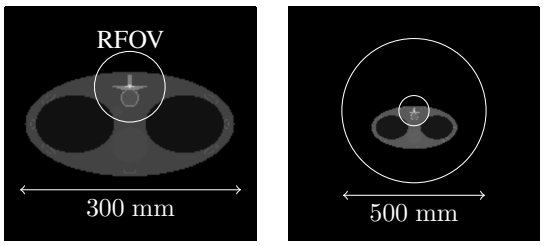


Fig. 6. Configuration in top view at $x_3 = 0$ for simulations. *Left*: the Forbild thorax phantom and RFOV. *Right*: the trajectory with the phantom in top view.

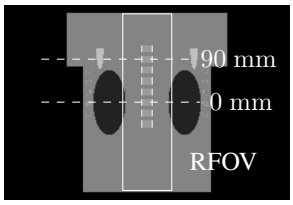


Fig. 7. Both cross-sections for the reconstruction in side view.

of the absence of chords. Specifically, for the 3-sin trajectory, the further vertically one moves away from the origin, the larger the area without chords. Results of the reconstructions are shown in Fig. 9. In the DBP theory, both cross-sections for the 2-sin trajectory can be exactly reconstructed, but for the 3-sin trajectory the cross-section $x_3 = 90$ mm cannot be reconstructed (no chord in the RFOV, the worst case). The cross-section $x_3 = 0$ mm is trickier: all points in the RFOV are traversed by a chord but there is no guarantee for the M-line method if it is needed (some points on the M-line must be intersected by a chord, see Section II-C above).

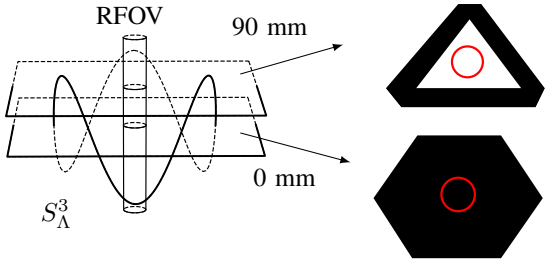


Fig. 8. *Left:* A 3-sin trajectory (not our dimensions, it is schematic) with a cylindrical RFOV and two horizontal cross-sections. *Right, above:* the set of chords (black, white means there is no chord) and the RFOV for the section $x_3 = 90$ mm, no point in the RFOV (red circle) for this section is traversed by a chord. *Right, below:* The set of chords (black) and the RFOV for the section $x_3 = 0$, all points in the RFOV for this section are traversed by a chord.

We note from Fig. 9 that all four images are of good quality and comparable quality to each other in the RFOV. Since the $x_3 = 90$ mm image for the 3-sin trajectory seemed similar to the same slice of the 2-sin trajectory, we tentatively consider that, due to some unknown reconstruction theory, this part of the Tuy region can be accurately reconstructed even though there are no chords passing through it. On the other hand, it is possible that our reconstructions are not fine enough to detect small artifacts or errors that might arise from an approximate reconstruction.

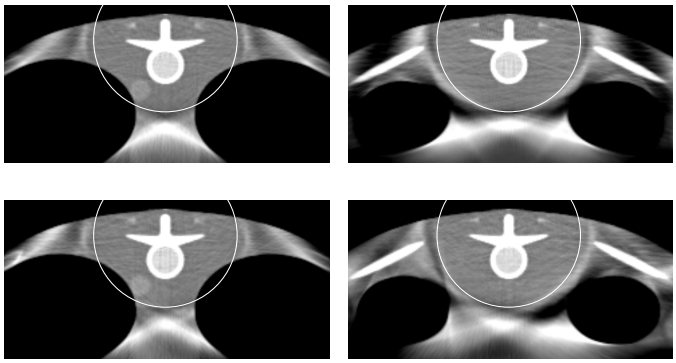


Fig. 9. Reconstruction for the 2-sin and 3-sin trajectories. *Left*: $x_3 = 0$. *Right*: $x_3 = 90$ mm. *Above*: $n = 2$. *Below*: $n = 3$. White circles represent the RFOV.

IV. Conclusion

We have shown that the M-line method can be used to reconstruct an ROI with some transverse truncation, in particular for 2-sin trajectories (sinusoid on a virtual cylinder), where only axial truncations are studied in the literature. This M-lines method has some conditions, for example each point of a ROI must be intersected by a chord.

Our numerical experiments using iterative reconstructions seemed to indicate that the 3-sin trajectory with transverse truncation also produces valid reconstructions, even where there are no chords and therefore the DBP-chord and M-lines methods both do not apply.

We note also, that for any n $>$ 2, the n-sin trajectory will have part of its Tuy region not intersected by chords. It seems likely that if reconstruction with transverse truncation is valid for 3-sin trajectories, that it will probably be valid for n-sin trajectories, for any n $>$ 3.

There is still CB reconstruction theory to be established to resolve these questions solidly, because it appears that reconstruction can be achieved for transaxial trunction with 3-sin, at least to the extent of the sensitivity of our simulations, and the authors know of no theory to explain it.

References

[1] H. K. Tuy, "An Inversion Formula for Cone-Beam Reconstruction," *SIAM Journal on Applied Mathematics*, vol. 43, no. 3, pp. 546–552, 2005.

[2] D. V. Finch, "Cone Beam Reconstruction with Sources on a Curve," *SIAM Journal on Applied Mathematics*, vol. 45, no. 4, pp. 665–673, 1985.

[3] J. D. Pack, F. Noo, and H. Kudo, "Investigation of saddle trajectories for cardiac CT imaging in cone-beam geometry," *Physics in Medicine and Biology*, vol. 49, no. 11, pp. 2317–2336, 2004.

[4] R. Clackdoyle, F. Noo, F. Momey, L. Desbat, and S. Rit, "Accurate Transaxial Region-of-Interest Reconstruction in Helical CT?," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 1, no. 4, pp. 334–345, 2017.

[5] R. Clackdoyle and M. Defrise, "Tomographic reconstruction in the 21st century," *IEEE Signal Processing Magazine*, no. July, pp. 60–80, 2010.

[6] M. Defrise, F. Noo, R. Clackdoyle, and H. Kudo, "Truncated Hilbert transform and image reconstruction from limited tomographic data," *Inverse Problems*, vol. 22, no. 3, pp. 1037–1053, 2006.

[7] F. Noo, R. Clackdoyle, and J. D. Pack, "A two-step Hilbert transform method for 2D image reconstruction," *Physics in Medicine and Biology*, vol. 49, no. 17, pp. 3903–3923, 2004.

[8] J. D. Pack, F. Noo, and R. Clackdoyle, "Cone-beam reconstruction using the backprojection of locally filtered projections," *IEEE Transactions on Medical Imaging*, vol. 24, no. 1, pp. 70–85, 2005.

Experimental Feasibility of a Residual-Modelled Material Decomposition Method for Spectral CT

Sen Wang, Yuxiang Xing, Hewei Gao, Marc Kachelrieß, Ao Zheng, Li Zhang*

***Abstract*—For photon counting based spectral x-ray CT imaging, basis material decomposition is a common practice for material density reconstruction. Compared to empirical look-up-table based decomposition methods, maximum-likelihood decomposition method yields better precision and noise property. However, it requires the knowledge of precise spectrum information and detector response calibration which may contain errors. If not compensated or corrected, errors in spectrum information or detector response will lead to decomposition bias which degrades the quantitative accuracy of reconstructed images. In this paper, a novel residual-modelled material decomposition method is proposed for spectral CT which fuses maximum likelihood decomposition and an empirical correction, aiming to achieve precise material decomposition. Maximum-likelihood decomposition functions as a primary estimation for material width. Based on primary decomposition results, decomposition residual is calculated and modelled by a polynomial model via a wedge calibration. The final estimation is the initial decomposition plus the residual correction.**

***Index Terms*—photon counting, material decomposition, maximum likelihood**

I. Introduction

DUAL energy x-ray imaging is capable of material discrimination and is free of beam hardening artifacts[1], [2], which is broadly used in clinical application, industry and security[3], [4]. Photon counting detectors (PCDs) can provide better and richer spectral information for computed tomography (CT) than conventional dual energy CT system facilitated with energy integrated detectors (EIDs)[5]–[10]. Material decomposition method is an effective way to utilize the spectral measurements, which is based on the assumption that common material attenuation coefficients can be linearly represented by two or more (if k-edge materials exist) functions[3], [5], [6], such as photoelectric term and Compton term or the combinations of different basis material attenuation coefficients[3]. In material decomposition processing procedure, the core is to estimate the line integral of magnitude of basis functions from spectral measurements[11], [12]. There are mainly two methods for the decomposition operation, maximum-likelihood (ML) decomposition and empirical methods. In this paper, we propose a residual-modelled material decomposition (RMMD) method that fuses the ML method and empirical correction for material decomposition. In the proposed RMMD method, ML decomposition functions as the primary estimator. Based on the primary ML decomposition results, an empirical residual correction model is reconstructed to correct or compensate the possible decomposition residuals induced by the errors in system model.

This work was supported partially by National key research and development program (No. 2016YFB1101203) and the National Natural Science Foundation of China (NNSFC) under grants 61527807 and 61771279.

Sen Wang was with Department of Engineering Physics, Tsinghua University, Beijing China (e-mail:senwang@stanford.edu).

Yuxiang Xing, Hewei Gao, Ao Zheng and Li Zhang are with Department of Engineering Physics, Tsinghua University, 100084, Beijing, China (e-mail: xingyx@tsinghua.edu.cn, hwgao@ tsinghua.edu.cn, zhenga16@mails.tsinghua.edu.cn, zli@tsinghua.edu.cn).

Marc Kachelrieß is with Medical Physics in Radiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany (e-mail: marc.kachelriess@dkfz.de).

II. Materials and Methods

A. Material decomposition with PCD

PCD discriminates each incident photon by the energy and records them in separate energy bins. $S_0(E)$ is the source spectrum generated by x-ray tube. After some attenuation in the forward path, source spectrum reaches the detector to become the incident spectrum for detector, $S_{\text{in}}(\text{E})$. The detected spectrum of PCD is the convolution result of incident spectrum and detector response $h(E;E')$.

$$\lambda_i = \int_{E_{\text{LTi}}}^{E_{\text{HTi}}} dE' \int_0^\infty S_0(E) \exp\left(-\int \mu(\mathbf{x},E)d\mathbf{x}\right) \cdot h\left(E';E\right)dE \quad (1)$$

, i=1,2,...,N

where λ_i denotes photon counts in the i-th energy bin, i=1,2,…,N. N is the total number of energy bins. $E_{\text{LT}i}$ and $E_{\text{HT}i}$ are the lower energy threshold and higher energy threshold, respectively. $\mu(\mathbf{x},E)$ is the attenuation coefficient in position $\mathbf{x}$.

For material decomposition, the fundamental assumption is that the attenuation coefficients can be finely approximated by linear combination of basis functions:

$$\mu(\mathbf{x},E) = \sum_{m=1}^{M} a_m(\mathbf{x}) f_m(E) \quad (2)$$

where $f_m(E)$ is the m-th basis function and $a_m(\mathbf{x})$ is the representation coefficient. M denotes the number of basis functions. With equation (2), we can reformulate equation (1) into:

$$\lambda_i(\mathbf{A}) = \int_{E_{\text{LTi}}}^{E_{\text{HTi}}} dE' \int_0^\infty S_0(E) \exp\left(-\sum A_m f_m(E)\right) \cdot h\left(E';E\right)dE \quad (3)$$

where

$$A_m = \int a_m(\mathbf{x})\, d\mathbf{x} \tag{4}$$

With a series of spectral measurements, i.e. λ_i, we can estimate the line integral A_m with material decomposition methods and $a_m(\mathbf{x})$ can be reconstructed with inverse Radon transform. For a specific CT system whose source spectrum and detector response are fixed, the energy bin counts λ_i are the function of line integral $\mathbf{A}$. ($\mathbf{A} = [A_m]^T$).

B. Residual-Modelled Material Decomposition (RMMD)

For ML decomposition and empirical decomposition, ML decomposition suffers from possible model error in source spectrum and detector response while empirical decomposition needs elaborated design to reach a good noise level[8], [12].

ML decomposition is both unbiased and efficient with precise system model (source spectrum and detector response). In real data decomposition, source spectrum and detector response are obtained with calibration methods[13], [14]. Decomposition bias might be induced by possible calibration inaccuracy. When calibrated source spectrum $S_0^{(c)}$ and detector response $h^{(c)}$ are used in ML decomposition, system model of (3) should be re-expressed:

$$\lambda_i\left(\mathbf{A} \mid S_0^{(c)}, h^{(c)}\right) = \int_{E_{\mathrm{LTi}}}^{E_{\mathrm{HTi}}} dE' \int_0^{\infty} S_0^{(c)}(E) \exp\left(-\int A_m f_m(E) \cdot d\mathbf{x}\right) h^{(c)}(E'; E)\, dE \tag{6}$$

The count difference from ground-truth is:

$$\Delta\lambda_i(\mathbf{A}) = \lambda_i\left(\mathbf{A} \mid S_0^{(c)}, h^{(c)}\right) - \lambda_i\left(\mathbf{A} \mid S_0^{*}, h^{*}\right) \tag{7}$$

where $\Delta\lambda_i$ denotes photon counts difference, S_0^* and h^* are the real source spectrum and detector response. From equation (7), we can find $\Delta\lambda_i$ is a function of $\mathbf{A}$ for a spectral CT system with source spectrum and detector response calibrated. As a result, the decomposition residual $\Delta\mathbf{A}$ is mathematically a function of $\mathbf{A}$ as well, regardless of the complex nonlinear processing of ML decomposition.

To ensure the good noise property of ML decomposition and correct or compensate the possible decomposition residual induced by model error, we try to fit the empirical mapping of $\Delta\mathbf{A}$ and ML decomposition results with polynomials via a wedge calibration.

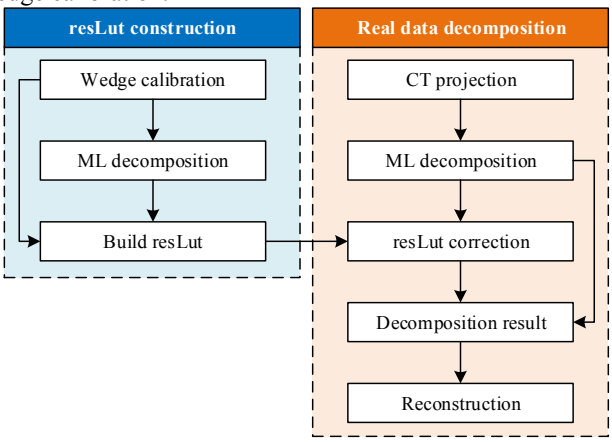


Fig. 1. Overall procedure of RMMD

Fig. 1 shows the overall procedure of the proposed RMMD method. There are two phases of the proposed method, residual correction model construction with wedge calibration phase that is denoted as resLUT construction in Fig. 1 and real data decomposition phase.

Wedge calibration provides a series of paired basis material width combinations $\mathbf{A}_{\mathrm{w}}=[A_{ij}]_{M\times K}$ and corresponding photon counts measurements $\mathbf{M}_{\mathrm{w}}=[m_{ij}]_{N\times K}$, The subscript M denotes the number of basis materials, N denotes the number of energy bins and K denotes the number of wedge tests. ML decomposition can give material decomposition result:

$$\mathbf{A}_{\mathrm{ML}} = D_{\mathrm{ML}}(\mathbf{M}_{\mathrm{w}}) \tag{8}$$

where $\mathbf{A}_{\mathrm{ML}}$ is ML decomposition results, D_{ML} denotes ML operation. If system model is accurate in ML decomposition, $\mathbf{A}_{\mathrm{ML}}$ should be equal to $\mathbf{A}_{\mathrm{w}}$. Otherwise, decomposition residual exists:

$$\Delta\mathbf{A} = \mathbf{A}_{\mathrm{w}} - \mathbf{A}_{\mathrm{ML}} = \left[\Delta A_{ij}\right]_{N\times K} \tag{9}$$

$\Delta\mathbf{A}$ is the residual matrix with the same dimension of $\mathbf{A}_{\mathrm{w}}$ and ΔA_{ij} is one of matrix elements in the i-th row and j-th column. From former analysis, $\Delta\mathbf{A}$ is the function of $\mathbf{A}_{\mathrm{w}}$, as well as the function of $\mathbf{A}_{\mathrm{ML}}$. We use polynomials to fit this relationship:

$$\Delta A_{ij} = \sum_{\substack{0\le j_1+\ldots+j_K\le t \\ j_1,\ldots,j_K\ge 0}} c_{j_1\cdots j_K}^{(i)} \prod_{k=1}^{M}\left(A_{kj}^{(\mathrm{ML})}\right)^{j_k} \tag{10}$$

where c is parameters of polynomials, i denotes i-th basis material. $A_{kj}^{(\mathrm{ML})}$ is the ML decomposition result, j_k is the power of ML decomposition result and t is the maximum power of polynomials. Parameter c in (10) can be determined by least-square fit for the residual look-up-table, resLut:

$$\Delta A_{\mathrm{resLut}} = \mathrm{resLut}(\mathbf{A}_{\mathrm{ML}}) \tag{11}$$

The final output of hybrid decomposition can then be formulated as follows:

$$\mathbf{A}_{\mathrm{RMMD}} = \mathrm{resLut}\left(D_{\mathrm{ML}}(\mathbf{M}_{\mathrm{w}})\right) + D_{\mathrm{ML}}(\mathbf{M}_{\mathrm{w}}) \tag{12}$$

III. Experiment Study

Physical experiments are carried out to verify the effectiveness and robustness of the proposed RMMD method. Fig. 2 shows the CT platform used in physical experiments. A VJ P051 x-ray source capable of up to 160 kVp and 3.12 mA, and an XCounter FLITE X1 photon counting detector(CdTe) with two energy channels were used in the experiment.

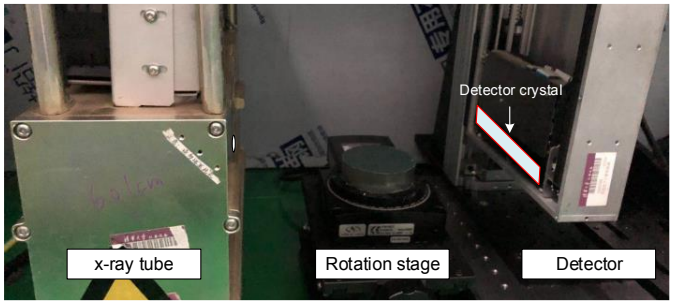


Fig. 2. Experiment CT platform with XCounter FLITE X1 PCD

Source spectrum is estimated with a weighted constraint spectral deconvolution method. Scattering is estimated with air value difference with and without attenuator[9]. Main configuration parameters of the system are listed in Table 1. We built the resLut through a wedge calibration with different combinations of Al and PMMA boards and tested the proposed method with two phantoms under different tube currents.

TABLE 1
MAIN PARAMETERS OF THE SYSTEM

Parameter	Value
Tube settings	100kVp, 0.2mA/0.35mA
Acquisition time	1s per frame
Geometry	Source to iso-center distance 346.2 mm Source to detector distance 207.2 mm
Detector	1536×128 pixels, 100×100 μm^2 per pixel
Energy bins	[24keV, 46keV], [46keV, 100keV]
Calibration wedge	Al: 0mm, 5.85mm, 11.70mm, 17.55mm PMMA: 0mm, 9.23mm, 18.46mm, 27.71mm, 36.86mm
Test phantoms	Al-PMMA phantom water-Teflon phantom

A fitting error is utilized as the bias metric for material decomposition accuracy in experimental tests:

$$\delta_\mu = \frac{1}{E_H - E_L} \cdot \int_{E_L}^{E_H} \frac{\left| \mu(E) - \sum_{k=1}^{M} a_k \mu_k(E) \right|}{\mu(E)} dE \qquad (13)$$

where δ_μ is the fitting error, $\mu(E)$ is attenuation coefficients of target materials which can be downloaded from NIST database[15]. M is the number of basis materials and E is the photon energy. In our simulation and experiment, M=2 and $E \in$ [24keV, 100keV].

A. Residual Correction of RMMD

We illustrate reconstruction results of basis images under 0.2mA tube current in Fig. 3. Fig. 3(a)(b) are results of ML and RMMD of a Al-PMMA phantom, respectively. Results of a water-Teflon phantom are shown in Fig. 3(c)(d).

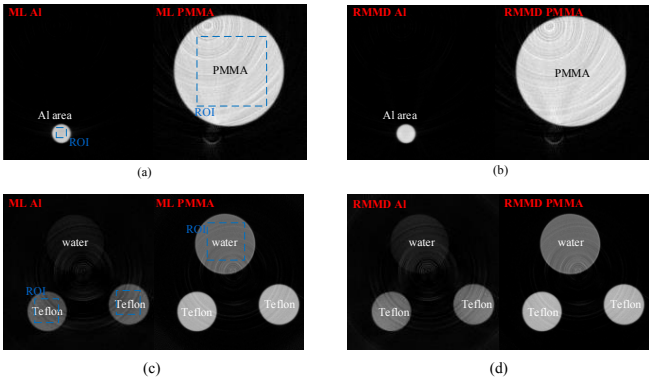
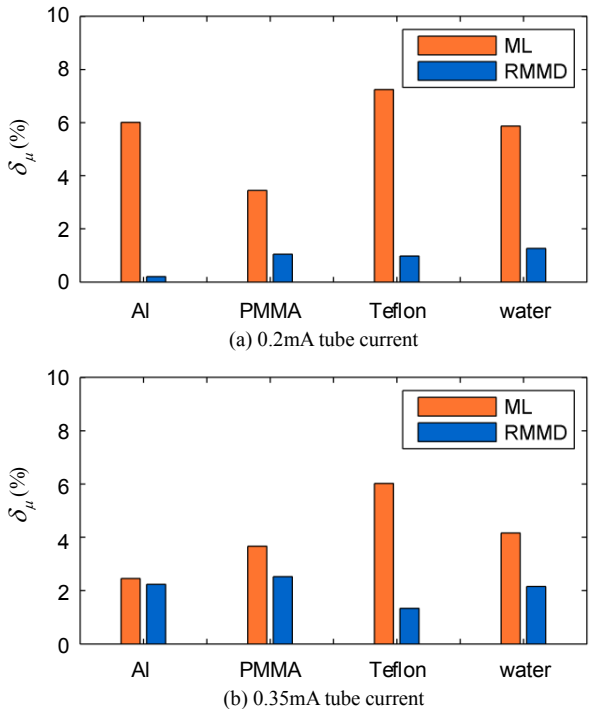


Fig. 3. Reconstruction of basis images in experimental tests. (a) ML results of Al-PMMA phantom, display window for all images is [0, 1.2]. (b) RMMD results of Al-PMMA phantom, display window for all images is [0, 1.2]. (c) ML results of water-Teflon phantom, display window is [0,0.3] for Al basis image and [0, 2] for PMMA basis image. (d) RMMD results of water-Teflon phantom, display window is [0,0.3] for Al basis image and [0, 2] for PMMA basis image.



(a) 0.2mA tube current

(b) 0.35mA tube current

Fig. 4. Fitting error of ROIs in experimental tests of different material areas under different tube currents. (a) 0.2mA tube current, (b) 0.35mA tube current

Fitting errors of different material ROIs under different tube currents are shown in Fig. 4. Fig. 4(a) and (b) denote the fitting error under 0.2mA tube current and 0.35mA tube current. For each material area under different tube currents, the fitting error of RMMD (blue bars) is smaller than that of ML results (orange bars). It demonstrates the effectiveness of the proposed RMMD method in modelling and correcting the decomposition residual.

B. Accuracy and noise property comparison

We compared the proposed RMMD method with an empirical decomposition method based on polynomials and the A-table method[16] for further evaluation.

Fig. 5 and Fig. 6 illustrate the comparison of fitting error and standard deviation of different methods. The green, purple, orange and blue bars denote results of polynomial based empirical decomposition (pLut), A-table method, ML decomposition and RMMD method, respectively.

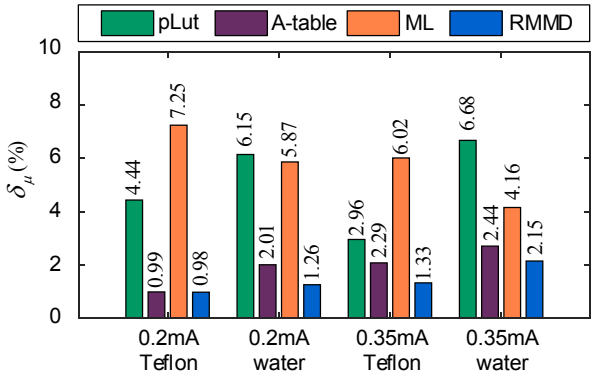
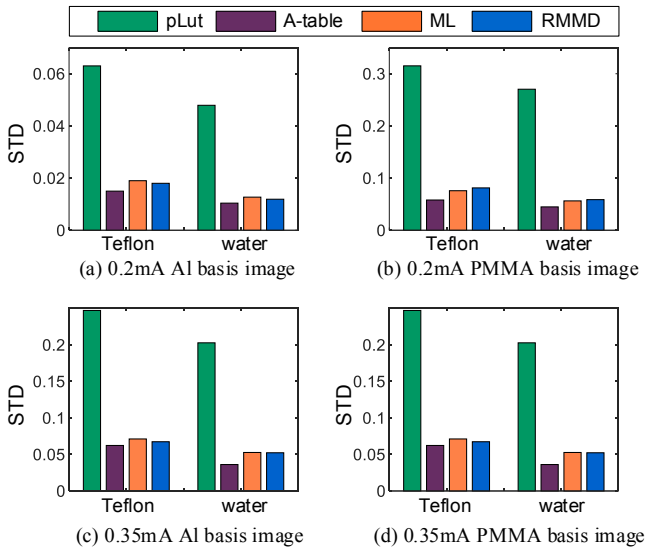


Fig. 5. Fitting error comparison

From Fig. 5, we can find due to the existence of system model bias, fitting errors of ML decomposition are sometimes worse than those of pLut but the proposed RMMD outperforms ML method, A-table and pLut in all material areas and tube currents. In Fig. 6, we can find noise level of pLut is the worst and RMMD inherits the noise property of ML while A-table shows a slight advantage to RMMD and ML.



(a) 0.2mA Al basis image (b) 0.2mA PMMA basis image

(c) 0.35mA Al basis image (d) 0.35mA PMMA basis image

Fig. 6. Standard deviation comparison

IV. Discussion and Conclusion

In this paper, a novel residual-modelled material decomposition (RMMD) method was proposed for quantitative imaging with PCDs. The proposed method uses ML method to generate primary decomposition results and corrects decomposition residuals with an empirical model, resLUT, to obtain the final decomposition results. The resLUT can be constructed in the form of polynomials via the wedge calibration. Physical experiment results demonstrate the proposed method can effectively improve the decomposition accuracy and inherits the good noise property of ML method.

Theoretically, ML decomposition residual should be induced by errors in system model (calibrated source spectrum and detector response). In our physical experiments, the detector response is calibrated with x-ray fluorescence measurements. Flux rate of x-ray fluorescence is relatively low without the risk of pulse pile-up (PPE) in PCDs while PPE might emerge under higher flux rate in practical CT scan[17], [18]. We carried out external experiments to see the decomposition results change under different x-ray tube currents and found ML decomposition becomes accurate with lower x-ray tube current. We think the system model error in this physical experiments is highly related to the absence of pulse pipe-up in our detector response model.

In other words, this implies our RMMD might be helpful to address the negative impact of pulse pile-up on material decomposition. Meanwhile, we are actively trying to model PPE in PCD and simulate its impact on material decomposition. The comparison between RMMD and ML with PPE model is worthy exploring.

Additionally, another important issue in real world CT application is scattering. Though we suppressed scattering influence with some scattering estimation method, we are to test the estimation accuracy and its influence in material decomposition as well in the future.

References

[1] T. R. Johnson *et al.*, "Material differentiation by dual energy CT: initial experience," *European radiology*, vol. 17, no. 6, pp. 1510–1517, 2007.

[2] A. Graser *et al.*, "Dual-energy CT in patients suspected of having renal masses: can virtual nonenhanced images replace true nonenhanced images?," *Radiology*, vol. 252, no. 2, pp. 433–440, 2009.

[3] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerised tomography," *Physics in Medicine & Biology*, vol. 21, no. 5, p. 733, 1976.

[4] G. Zhang, J. Cheng, L. Zhang, Z. Chen, and Y. Xing, "A practical reconstruction method for dual energy computed tomography," *Journal of X-ray Science and Technology*, vol. 16, no. 2, pp. 67–88, 2008.

[5] Jp. Schlomka *et al.*, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Physics in Medicine & Biology*, vol. 53, no. 15, p. 4031, 2008.

[6] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical physics*, vol. 40, no. 10, 2013.

[7] X. Wang, D. Meier, K. Taguchi, D. J. Wagenaar, B. E. Patt, and E. C. Frey, "Material separation in x-ray CT with energy resolved photon-counting detectors," *Medical physics*, vol. 38, no. 3, pp. 1534–1546, 2011.

[8] D. Wu, L. Zhang, X. Zhu, X. Xu, and S. Wang, "A weighted polynomial based material decomposition method for spectral x-ray CT imaging," *Physics in Medicine & Biology*, vol. 61, no. 10, p. 3749, 2016.

[9] X. Xu, Y. Xing, S. Wang, and L. Zhang, "Systematic implementation of spectral CT with a photon counting detector for liquid security inspection," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 893, pp. 99–108, 2018.

[10] R. Li, L. Li, and Z. Chen, "Feasible energy calibration for multi-threshold photon-counting detectors based on reconstructed XRF spectra," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 1, no. 2, pp. 109–120, 2016.

[11] D. Wu, X. Xu, L. Zhang, and S. Wang, "A hybrid Monte Carlo model for the energy response functions of X-ray photon counting detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 830, pp. 397–406, 2016.

[12] R. E. Alvarez, "Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis," *Medical physics*, vol. 38, no. 5, pp. 2324–2334, 2011.

[13] X. Xu, L. Zhang, D. Wu, and S. Wang, "Response function estimation for the XCounter Flite X1 photon counting detector using Monte Carlo method," in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, 2016, pp. 1–4.

[14] H. Ding, H.-M. Cho, W. C. Barber, J. S. Iwanczyk, and S. Molloi, "Characterization of energy response for photon-counting detectors using x-ray fluorescence," *Medical physics*, vol. 41, no. 12, 2014.

[15] J. Hubbell and S. Seltzer, "NIST standard reference database 126," *Gaithersburg, MD: National Institute of Standards and Technology*, 1996.

[16] Alvarez, Robert E. "Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis." Medical physics 38.5 (2011): 2324-2334.

[17] K. Taguchi, E. C. Frey, X. Wang, J. S. Iwanczyk, and W. C. Barber, "An analytical model of the effects of pulse pileup on the energy spectrum recorded by energy resolved photon counting x-ray detectors," *Medical physics*, vol. 37, no. 8, pp. 3957–3969, 2010.

[18] K. Taguchi *et al.*, "Modeling the performance of a photon counting x-ray detector for CT: Energy response and pulse pileup effects," *Medical physics*, vol. 38, no. 2, pp. 1089–1102, 2011.

An iterative reconstruction method for material decomposition of Dual Spectral Computed Tomography based on monochromatic images

Weibin Zhang, Shusen Zhao, Huiying Pan, and Xing Zhao*

***Abstract*—Compared with traditional computed tomography (CT), dual spectral CT (DSCT) has the capability to improve material differentiation and reduce beam hardening artifacts, thus has wide prospects of applications. However, the decomposition process is an illposed problem, which is sensitive to noise. Considering the structural consistency and the locally linear relationship of gray value between monochromatic image and polychromatic image, this paper proposes an iterative reconstruction method for material decomposition of DSCT based on monochromatic images without the knowledge of spectral information. Firstly, we calibrate the polychromatic projection function using polynomial functions between two monochromatic projections and polychromatic projections, which can reduce calibration error and improve the accuracy of decomposition. Secondly, we develop an alternately iterative method: the IFBP method is adopted to obtain monochromatic images; and then using high-energy image with high signal-to-noise ratio as a reference image, the guided image filtering is utilized to process the decomposed monochromatic images, reducing the impact of noise and further improving the image quality. The proposed method is validated with numerical experiments and the results confirm the effectiveness of our proposed approach.**

***Index Terms*—Dual spectral CT, material decomposition, monochrome images, iterative reconstruction method.**

I. Introduction

In dual spectral computed tomography (DSCT), the measured object is scanned with two different X-ray spectra. The collected polychromatic projection can be utilized to reconstruct electron density and effective atomic number images, energy-selective images, or material-selective images [1, 2]. Compared with conventional CT, DSCT has a better material distinguishability. Therefore, it has wide potential applications in medical and industrial domains, such as bone removal, PET attenuation correction, bone marrow composition analysis, nondestructive evaluation, security inspection and so on.

The existing methods to achieve decomposition of DSCT can be classified into three groups: image based methods, projection based methods, and iterative methods. Image based methods reconstruct each polychromatic projection dataset separately. Then, images from both spectra are linearly combined to obtain two decomposed images [3, 4]. The linear combinations do not include the product terms of images of spectra, hence it's impossible to perform an exact material separation without combining the polychromatic properties of both spectra, which leading the decomposition results will suffer from artifacts. Projection based methods combine the polychromatic projections of each spectrum to obtain two projections of basis material images, then the basis material images are directly reconstructed from their projections [5, 6]. Generally speaking, they can obtain better decomposition results than image based ones, but it requires geometrically consistent polychromatic projections. Several iterative reconstruction methods based on statistical models and non-linear optimizations were proposed [7-10]. By introducing prior knowledge or establishing an approximate model, these methods improve the decomposed image quality effectively. In particular, these iterative methods require precise knowledge of the detected X-ray spectra to implement the polychromatic forward-projection process. Although various methods have been developed to estimate the X-ray spectrum, the precise measurement of an X-ray spectrum is still a difficult task due to general calibration issues. Besides, the illposedness of the decomposition process causes it noticeably sensitive to noise, resulting in reduced signal-to-noise ratio.

The above methods directly obtain basis material images. Recently, some scholars have proposed some methods to obtain basis material images indirectly with monochromatic images [11-12].

Due to structural consistency between monochromatic image and polychromatic image, polychromatic forward projection based on monochromatic images has higher accuracy than that based on basis material images. Therefore, under the condition of the unknown X-ray spectrum, this paper proposes an iterative reconstruction method for material decomposition of DSCT based on monochromatic images. The method is divided into two steps: calibration step for polychromatic forward projection based on monochromatic images and alternately iterative solution. Firstly, the method calibrates the polychromatic forward projection function using polynomial functions between two monochromatic projections and polychromatic projections. Once the coefficients of the polynomial are determined by using a calibration scan of a phantom, the

This work was partly supported by the National Natural Science Foundation of China (grants: 61827809, 61671311, 61501310, 61371195), Beijing Commission of Education grant KM201710028002.

All authors are with the School of Mathematical Sciences, Capital Normal University, Beijing, 100048, Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, 100048, China.

Corresponding author: Xing Zhao (e-mail: zhaoxing_1999@126.com).

polynomial would be inverted via the IFBP method [10] to solve for monochromatic images. Furthermore, monochromatic image and high-energy image with high signal-to-noise ratio have a locally linear relationship of gray value. Through adding the prior information into the optimization model, the monochromatic images are processed by the guided image filtering technique [13]. Finally, the basis material images can be decomposed by the two monochromatic images. Numerical experiments show that the proposed method can achieve more accurate material decomposition of DSCT. Compared with other methods, our method can better suppress noise, eliminate hardening artifacts, and improve the image quality. At the same time, the method has faster convergence speed.

II. Method

In this section, the mathematical model of DSCT based on monochromatic images is presented and an iterative method is developed.

A. DSCT Reconstruction model based on monochromatic images

In DSCT, two projection data sets are acquired with two different X-ray spectra. By considering the facts that X-ray spectrum is polychromatic and scattering is negligible, we describe the physical process as follow:

$$q_k(L) = -\ln \int S_k(E) e^{-\int_L \mu(x,E)dl}\, dE\,, L \in \mathcal{L}_k,\ k = 1,\ 2. \tag{1}$$

where $\mu(x, E)$ represents the linear attenuation coefficient of the scanned object at point x and energy E, L represents one x-ray path, $\mathcal{L}$ represents the index sets of the x-ray paths, $S(E)$ represents the normalized effective x-ray spectra, k denotes low-energy and high-energy index 1, 2.

For DSCT, the linear attenuation coefficient $\mu(x, E)$ is decomposed with respect to variables E and x:

$$\mu(\boldsymbol{x}, E) = \phi(E) f(\boldsymbol{x}) + \theta(E) g(\boldsymbol{x}) \tag{2}$$

where $\phi(E)$ and $\theta(E)$ are two pre-defined basis functions of energy E, $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are the corresponding coefficient functions of spatial variable $\boldsymbol{x}$. In this paper, basis material decomposition is used, $\phi(E)$ and $\theta(E)$ are the mass attenuation coefficients of two selected basis materials, and the reconstructed $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are the mass densities of the two basis materials.

Substituting (2) into (1), we get:

$$q_k(L) = -\ln \int S_k(E) e^{-\phi(E)\int_L f(x)dl - \theta(E)\int_L g(x)dl}\, dE\,, L \in \mathcal{L}_k \tag{3}$$

Under two different X-ray spectra, monochromatic images $I_{E_1}(x)$ and $I_{E_2}(x)$ can be represented by the basis material images $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ as follow:

$$\begin{pmatrix} I_{E1}(x) \\ I_{E2}(x) \end{pmatrix} = \begin{pmatrix} \phi(E_1) & \theta(E_1) \\ \phi(E_2) & \theta(E_2) \end{pmatrix} \cdot \begin{pmatrix} f(\boldsymbol{x}) \\ g(\boldsymbol{x}) \end{pmatrix} \tag{4}$$

Equally, basis material images $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ can be represented:

$$\begin{pmatrix} f(\boldsymbol{x}) \\ g(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} \phi(E_1) & \theta(E_1) \\ \phi(E_2) & \theta(E_2) \end{pmatrix}^{-1} \begin{pmatrix} I_{E_1}(x) \\ I_{E_2}(x) \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} I_{E_1}(x) \\ I_{E_2}(x) \end{pmatrix} \tag{5}$$

Substituting (5) into (3), we establish polychromatic projection function based on monochromatic projections as follow:

$$q_k(L) = -\ln \int S_k(E) e^{-\Phi(E) p_{E_1}(L) - \Theta(E) p_{E_2}(L)}\, dE\,, \quad L \in \mathcal{L}_k \tag{6}$$

where $p_{E_1}(L) = \int_L I_{E_1}(x)\, dl$ and $p_{E_2}(\mathrm{L}) = \int_L I_{E_2}(x)\, dl$ denote the line integrals of $I_{E_1}(x)$ and $I_{E_2}(x)$ along the x-ray path L, $\Phi(E) = a_1\phi(E) + b_1\theta(E)$, $\Theta(E) = a_2\phi(E) + b_2\theta(E)$ denote linear combination of the mass attenuation coefficients of two selected basis materials.

Now, the DSCT reconstruction problem is to solve $I_{E_1}(x)$ and $I_{E_2}(x)$ from the obtained $q_k(L)$.

B. The calibration step for polychromatic forward projection

When the X ray spectra are unknown, the polychromatic forward projection in (6) cannot be determined. Contrary to the EDEC method, the proposed method is an iterative method. Therefore, it requires to calibrate a polychromatic forward projection that converts the monochromatic projections $p_{E_1}(L)$ and $p_{E_2}(L)$ to obtain virtual measurements $q_k(L)$. Here, a polynomial function is used to fit the polychromatic projection function in (6):

$$q_k(L) = C_k\big(\mathcal{R}(I_{E_1}), \mathcal{R}(I_{E_2}(x))\big) = \sum_{m=0}^{M}\sum_{n=0}^{N} c_{k,mn}\, p_{E_1}(L)^m p_{E_2}(L)^n \tag{7}$$

where the polychromatic projection $q_k(L)$ is expressed as a linear combination of the basis functions $p_{E_1}(L)^m p_{E_2}(L)^n$, $c_{k,mn}$ denotes polynomial coefficients, $\mathcal{R}$ represents the operator of line integrals.

EDEC method is used for the calculation of those coefficients. Once the polynomial coefficients are calibrated, the polynomial functions in (7) are determined. Then, the low-energy and high-energy polychromatic projections can be calculated by passing the monochromatic projections through the polynomial functions.

C. Iterative decomposition step for DSCT

Except for the difference of gray value, monochromatic image and polychromatic image are completely consistent in texture information and structure information. Although the high-energy polychromatic image has weaker capability for material distinguishing, its quality is significantly better than decomposed monochromatic images. By incorporating this constraint into an optimization model, the high-energy polychromatic image could work as a reference to effectively improve the smoothness and keep all the structures of decomposed results. Based on the above facts, we propose an optimization model for DSCT as follow:

$$\min_{I_{E_1}, I_{E_2}} \left\{ \begin{array}{c} \sum_{k=1}^{2} \left\| q_k(L)^* - C_k\big(\mathcal{R}(I_{E_1}(x)), \mathcal{R}(I_{E_2}(x))\big) \right\|_2^2 \\ +\lambda \sum_{k=1}^{2} R\big(I_{E_k}(x), I_{High}(x)\big) \end{array} \right\} \tag{8}$$

where $q_k(L)^*$ represents the measured polychromatic projection, λ is the regularization parameter, $R(\cdot)$ is the regularization term, in which guided image filtering is chosen, $I_{High}(x)$ is the reconstructed high-energy polychromatic image.

Considering the facts that the data terms are measured in projection domain and the regularized terms are measured in image domain, we adopt a simple altering minimization technique. Let $I_{E_k}^n$ denotes the solution after n-th iteration, then the update for $I_{E_k}^{n+1}$ can be split into two subproblems:

Subproblem 1:

$$I_{E_1}^{n+1/2}, I_{E_2}^{n+1/2} = \min_{I_{E_1}, I_{E_2}} \left\| q_1(L)^* - C_1\left(\mathcal{R}(I_{E_1}), \mathcal{R}(I_{E_2})\right) \right\|_2^2 + \left\| q_2(L)^* - C_2\left(\mathcal{R}(I_{E_1}), \mathcal{R}(I_{E_2})\right) \right\|_2^2 \quad (9)$$

Subproblem 2:

$$I_{E_k}^{n+1} = \min_{I_{E_k}} R\left(I_{E_k}(x), I_{High}(x)\right) \quad (10)$$

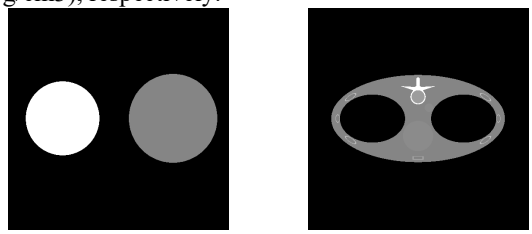
We can alternately obtain monochromatic images by using the IFBP method for (9), and the guided image filtering for (10). Finally, decomposition to basis material composition with monochromatic images can be obtained.

III. Experiments

To verify the effectiveness of the proposed method, numerical experiments with simulated data and real data are performed. For comparison, we have also implemented the EDEC [6] method and PFPC [7] method, since they are not only effective but also quite similar to the proposed method.

A. Simulated Data Experiments

The dual simulated data are generated with 80 kV and 140 kV X-ray spectra obtained by the software SpectrumGUI. A slice of the 3D FORBILD thorax phantom is used as test phantom, as shown in Fig.1(b). The calibration phantom is defined consisting of a big circular disc with a diameter of 60 mm and a small circular disc with a diameter of 48 mm (Fig. 1(a)), which are filled with water (1.0 g/cm3) and bone tissue (1.92 g/cm3), respectively.



(a) (b)

Fig. 1. The phantoms in the simulated data experiments. (a) The calibration phantom; (b) The test phantom.

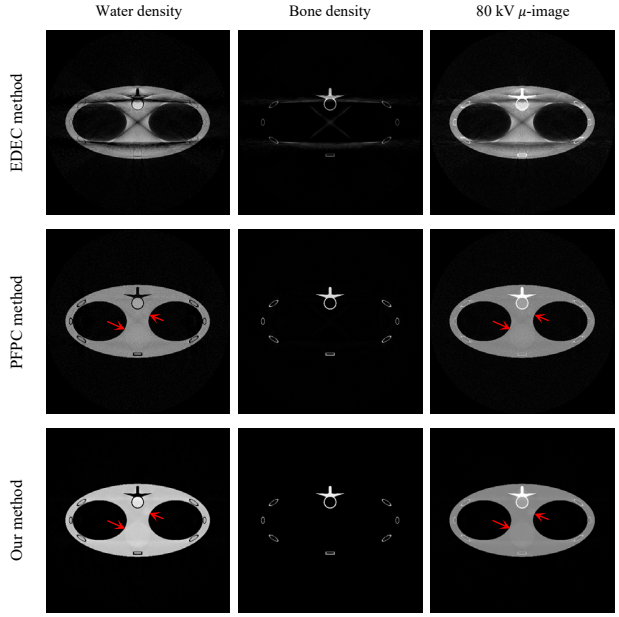


Fig. 2. The images reconstructed with EDEC method, PFPC method and the proposed method in the simulated data.

Fig. 2 shows the reconstruction results of EDEC method, PFPC method and the proposed method. We can see from the Fig. 2 that there are some radial artifacts in the reconstruction results of EDEC method, the basis materials are not separated completely. Because the EDEC method is a direct calibration process, which is greatly affected by noise, there is no iterative step to modify the reconstruction results. Although the PFPC method can achieve good reconstruction results and decompose the basis materials correctly, it can't clearly distinguish the areas where the density value is close (marked with red arrows). On the contrary, the proposed method not only accurately decompose the basis materials, but also distinguish the closed areas. It's due to the higher accuracy of the calibration based on the monochromatic images than that based on basis material images, thus results in smaller decomposition error. At the same time, the guided filtering technology is introduced as a constraint, effectively improving the smoothness and keeping all structures of decomposed results.

Fig. 3 illustrates the profiles of the central columns of the monochromatic images presented in Fig. 2. It can be seen from the profiles that the reconstructed density values of the proposed method are closer to the corresponding theoretical density values, which indicates that the proposed method achieves better reconstruction results.

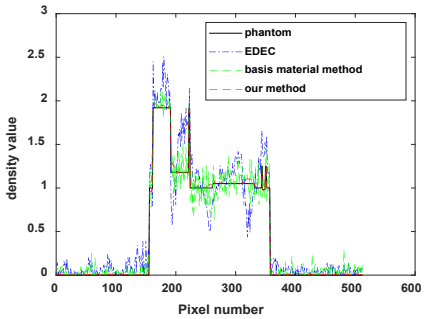


Fig. 3. Profiles of the central columns of the monochromatic images reconstructed shown in Fig. 2.

In addition, we also compare the convergence rate. Fig. 4 shows the NMAD of the decomposed images of the two methods based on basis material images and monochromatic images. From the convergence curve, we can see that the proposed method has faster convergence speed.

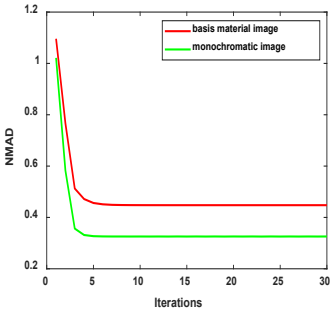
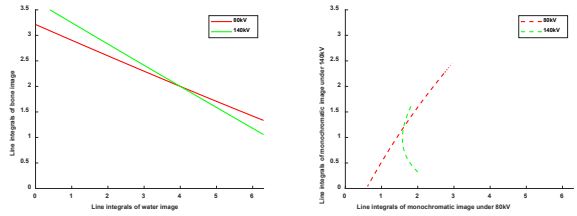


Fig. 4. The NMAD of the reconstructed images based on the basis material images and monochromatic images.



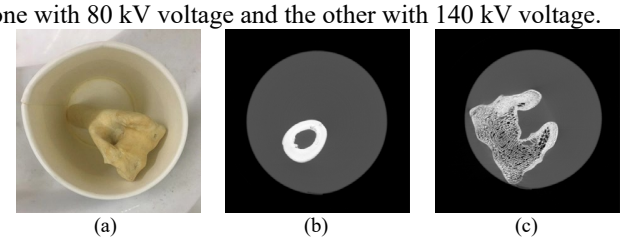
(a) (b)

Fig. 5. The comparison of intersection angle of polychromatic projection isopleth. (a) monochromatic images method, (b) basis material images method.

The reason is that intersection angle of polychromatic projection isopleth based on monochromatic images is larger than that based on basis material images, as shown in Fig. 5. In other word, the condition number of coefficient matrix is smaller, resulting in faster convergence rate.

B. Real Data Experiments

We also carried out a real data experiments on the industrial CT system of our laboratory. Fig. 6 shows a photograph of the scanned phantom, which is scanned twice with the CT system, one with 80 kV voltage and the other with 140 kV voltage.



(a) (b) (c)

Fig. 6. The phantoms in the real data experiments. (a) The scanned phantom; (b) The calibration phantom; (c) The test phantom.

The results are shown in Fig. 7. As we can see, the decomposed basis materials images with the proposed method are not only of good image quality but also with good accuracy than those decomposed with EDEC method and PFPC method. The proposed method effectively improves the accuracy and keeps more structures of decomposed results, especially clear bone trabecula structure.

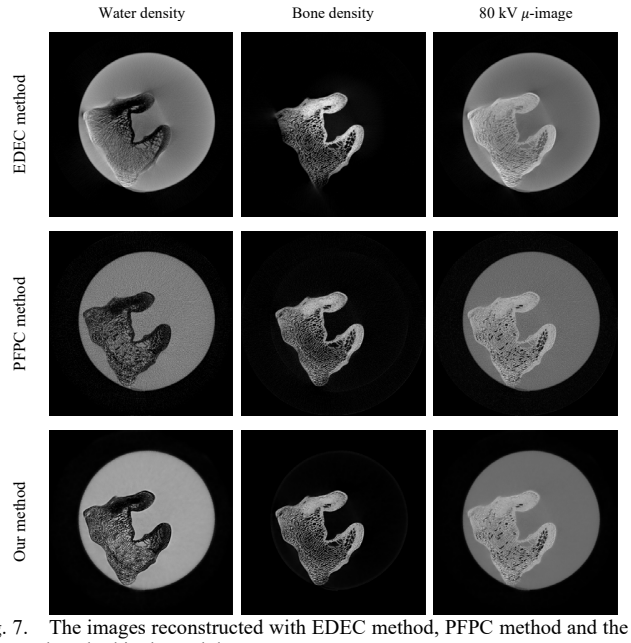


Fig. 7. The images reconstructed with EDEC method, PFPC method and the proposed method in the real data.

IV. Conclusion

In this paper, we have proposed an iterative reconstruction method for material decomposition of DSCT based on monochromatic images. In the calibration step, our method borrows the idea of polynomial fitting from EDEC. Different from the basis material image method, we establish the nonlinear relationship between polychromatic projections and monochromatic images, which can reduce the calibration error and improve the image quality. Besides, basis materials should be selective and the material-selective reconstructions are obtained in EDEC, while there is no limit of types of calibration materials and the monochromatic images are directly obtained in our method. In the iteration step, the guided filter technology is employed to constrain the monochromatic images, which can effectively improve the smoothness of decomposed results and keep the structure information. Further decomposition to basis material images with monochromatic images can be easily done.

Further work is to extend the iterative method to multiple energy CT for multiple material separations.

References

[1] R. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography," Phys. Med. Biol., vol. 21, no. 5, pp. 733–744, Feb, 1976.

[2] W. A. Kalender, W. H. Perman, J. R. Vetter, and E. Klotz, "Evaluation of a prototype dual-energy computed tomographic apparatus. I. Phantom studies," Medical physics, vol. 13, no. 3, pp. 334–339, May, 1986.

[3] C. Maaß, M. Baer and M Kachelrieß, "Image-based dual-energy CT using optimised precorrection functions: a practical new approach of material decomposition in the image domain," Med. Phys., vol. 36, no. 8, pp. 3818-3829, Jul, 2009.

[4] A. J. Coleman and M. Sinclair, "A beam-hardening correction using dual-energy computed tomography," Phys. Med. Biol., vol. 30, no. 11, pp. 1251–1256, May, 1985.

[5] K. S. Chuang and H. K. Huang, "Comparison of four dual energy image decomposition methods," Phys. Med. Bio., vol. 33, no. 4, pp. 455–466, 1988.

[6] P. Stenner, T. Berkus and M. Kachelrieß, "Empirical dual-energy calibration (EDEC) for cone beam computed tomography," Med. Phys., vol. 34, no. 9, pp. 3630-3641, Aug, 2007.

[7] P. Sukovic and N. H. Clinthorne, "Penalized weighted least-squares image reconstruction for dual energy X-ray transmission tomography," *IEEE trans.* Medical Imaging, vol. 19, no. 11, pp. 1075–1081, Nov, 2000.

[8] X. Zhao, J. J. Hu, Y. S. Zhao, "A novel iterative reconstruction method for dual-energy computed tomography based on polychromatic forward projection calibration," Insight, vol. 56, no. 10, Oct, 2014.

[9] Y. Zhao, X. Zhao, and P. Zhang, "An extended algebraic reconstruction technique (E-ART) for dual spectral CT," *IEEE trans.* Medical Imaging, vol. 34, no. 3, pp. 761–768, Nov, 2015.

[10] M. F Li, Y. S. Zhao, and P. Zhang, "Accurate iterative FBP reconstruction method for material decomposition of dual energy CT," *IEEE trans.* Medical Imaging, vol. 38, no. 3, pp. 802-812, Sep, 2017.

[11] C. Feng , "An empirical material decomposition method (EMDM) for spectral CT," in IEEE Nuclear Science Symp. Medical Imaging Conf. and Room-Temperature Semiconductor Detector Workshop, pp. 1–5, 2016.

[12] C. F. Feng, K. Kang, and Y. X. Xing, "Fully connected neural network for virtual monochromatic imaging in spectral computed tomography," *IEEE trans.* Medical Imaging, vol. 6, no. 1, Oct, 2019.

[13] K. M. He, J. Sun, and X. O. Tang, "Guided Image Filtering", *IEEE Trans.* Pattern Analysis And Machine Intelligence, vol. 35, no. 6, pp. 1397-1409, Jun, 2013.

Spatially varying regularization weights for one-step spectral CT with SQS

Pierre-Antoine Rodesch, Salim Si-Mohamed and Simon Rit

***Abstract*—Photon-counting detectors provide spectral information for x-ray acquisitions. Taking advantage of this information currently requires iterative algorithms to reconstruct basis material CT images. One-step reconstruction is the simultaneous inversion of the spectral distortion occurring in the detector and the geometrical projection. Separable quadratic surrogate (SQS) algorithms have been applied to this one-step problem with satisfactory convergence and material separation. However, this class of method leads to numerical instabilities stemming from voxels out of the field-of-view (FOV) which need to be included in the forward model for reconstructing the FOV. We aim at improving one-step spectral CT reconstruction by investigating two possible corrections of this effect: replacing the exponential in the forward model by a linear function for negative attenuations and spatially varying regularization depending on the geometrical conditioning. We demonstrate the efficiency of the second method using experimental data acquired on a clinical prototype CT scanner with a photon-counting detector.**

I. Introduction

In spectral computed tomography (CT), one-step approaches directly decompose photon counts in basis material density volumes. This class of methods is opposed to two-step methods, in which decomposition occurs before or after tomographic reconstruction, and which are less computationally expensive. The explicit one-step modeling enables regularization of material volumes while simultaneously inverting line integrals and the spectral distortion of the detector. This provides density maps with less correlated noise and can also be applied to K-edge imaging and unconventional scanning geometries [1].

Several algorithms have been proposed to inverse the one-step forward model. A combination of separable quadratic surrogate (SQS) and Nesterov's momentum methods have proved efficient in solving the one-step reconstruction problem of spectral CT with a K-edge material map from data acquired with photon-counting detectors [2]. In SQS approaches, each voxel is updated separately by minimizing a quadratic surrogate of the cost function at each iteration. Several SQSs have been developed for one-step spectral CT [3], [4], [6]. We chose the SQS proposed by Weidinger et al. [3] combined with ordered subsets (OS) and Nesterov's momentum technique as proposed in [4].

The problem addressed in this article are artifacts originating from voxels which are not in the field-of-view (FOV) but which are iteratively updated to accurately model the forward problem. These voxels are present in axial scans (the red stripped parts in Fig. 1) and helical scans. They cannot be accurately reconstructed due to limited angle data (since they are not in the FOV) but they influence the FOV because they are traversed by rays also going through the FOV (which is why they are reconstructed) and due to three-dimensional (3D) spatial regularization. This problem has already been described in [5] for SQS applied to mono-energetic CT in a helical acquisition case. The authors proposed to use a spatially varying update step based on their geometrical conditioning, i.e., the amount of projection data contributing to each voxel (which is less for voxels outside the FOV).

We have observed the same problem in spectral CT but the observed consequences are more severe. We focus here on the case of axial CT acquisitions. Since the forward problem generally involves the exponential function, negative values in these regions quickly translate to numbers which cannot be handled numerically. Unlike Kim et al [5], this behavior is observed without OS, but OS enhance the phenomenon. A potential solution would be the initialization of the one-step reconstruction with material volumes close to the solution [4], [6], e.g., a two-step reconstruction decomposition. We also investigate solutions for these numerical instabilities using a zero initialization to avoid the difficult question of the influence of the accuracy of the initialization.

This paper investigates two corrections. The first one is the replacement in the forward model of the exponential function by a soft-exponential as proposed by Sidky et al. [7]. The second one, inspired by [5] but formulated differently, is the

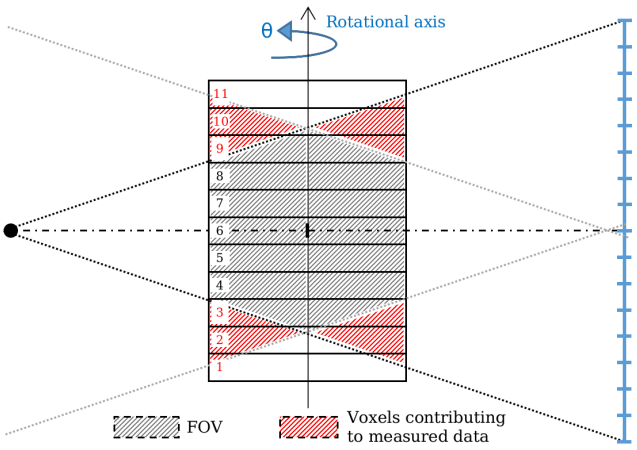


Fig. 1. Axial scan configuration with selected volume geometry. Displayed numbers correspond to slice indices used in the results section.

P.A Rodesch & S. Rit are with the Université de Lyon, CREATIS ; CNRS UMR5220 ; Inserm U1206 ; INSA-Lyon ; Université Lyon 1 ; Centre Léon Bérard, France.

Salim Si-mohamed is with the Université de Lyon, CREATIS ; CNRS UMR5220 ; Inserm U1206 ; INSA-Lyon ; Université Lyon 1, France and with the Radiology Department, Lyon University Hospital, Lyon, France.

implementation of regularization weights which vary in the volume according to the geometrical conditioning. We also study the combination of both and compare the results to the conventional SQS algorithm, with and without initialization.

II. Materials and Methods

A. SQS

We briefly summarize in this subsection common equations used in SQS techniques applied to the one-step problem. We explain the choice we made for our uncorrected method. As usual, the spectral discrete forward model is:

$$\overline{c}_{ib}(f) = \sum_{\epsilon=1}^{N_\epsilon} s_{ib}(\epsilon)\ e^{-t_i(f,\epsilon)} \tag{1}$$

where $\overline{c}_{ib}(f)$ is the expected counts in bin b for indexed detector pixel i, $s_{ib}(\epsilon)$ is the effective spectrum at energy ϵ and $t_i(f,\epsilon)$ the attenuation of the object expressed by:

$$t_i(f,\epsilon) = \sum_{m=1}^{N_m}\sum_{j=1}^{N_v} \mu_m(\epsilon) A_{ij} f_{mj} \tag{2}$$

with $\mu_m(\epsilon)$ the m-th basis (material) function, A the geometrical projection matrix and f the unknown basis volumes to be reconstructed with j the voxel index.

The problem is solved by minimizing the negative log-likelihood:

$$D_{\mathrm{PL}}(c,f) = \sum_{i=1}^{N_p}\sum_{b=1}^{N_b} \overline{c}_{ib}(f) - c_{ib}\ log\Big[\overline{c}_{ib}(f)\Big] \tag{3}$$

with c the measured photon counts.

A penalization term is added to the cost function with a spatial regularization applied independently on each basis volume:

$$R(f) = \sum_{m=1}^{N_m} \beta_m \sum_{j=1}^{N_v}\sum_{\xi\in\mathbb{V}_j}^{N_v} \phi(f_{mj} - f_{m\xi}) \tag{4}$$

with β_m the regularization weight per material and ϕ a twice differentiable convex function. In this study, we used the Green prior which, unlike others like the Huber function, does not require an extra hyperparameter. Only the β_m have to be tuned.

The SQS method for one-step reconstruction relies on the application of Newton's method at each voxel j. It enables the analytical minimization of Q^n, the SQS calculated at iteration n:

$$f_j^{n+1} = f_j^n - [H(Q^n)_j]^{-1}.[(\nabla Q^n)_j] \tag{5}$$

where $(\nabla Q^n)_j \in \mathbb{R}^{N_m}$ is the gradient of Q^n and $H(Q^n)_j \in \mathbb{R}^{N_m\times N_m}$ its Hessian. This update scheme is accelerated with OS and Nesterov's momentum technique as described by Kim et al. [8]. The gradient of the surrogate of the data fidelity term is:

$$(\nabla Q^n)_j = \sum_{i=1}^{N_p} A_{ij}\times \sum_{b=1}^{N_b}\Big(1 - \frac{c_{ib}}{\overline{c}_{ib}(f^n)}\Big)\sum_{\epsilon=1}^{N_\epsilon} s_{ib}(\epsilon) e^{-t_i(f^n,\epsilon)}\mu(\epsilon) \tag{6}$$

where $\mu(\epsilon) \in \mathbb{R}^{N_m}$ is the vector of attenuation coefficient at energy ϵ. Note that by definition of surrogates, $(\nabla Q^n) = (\nabla D_{\mathrm{PL}|f=f^n})$.

The Hessian of Weidinger's surrogate [3] is:

$$H(Q^n)_j = \sum_{i=1}^{N_p} A_{ij}\Big(\sum_{\xi=1}^{N_v} A_{i\xi}\Big)\times \sum_{b=1}^{N_b}\sum_{\epsilon=1}^{N_\epsilon} s_{ib}(\epsilon) e^{-t_i(f^n,\epsilon)}\ {}^T\mu(\epsilon)\otimes\mu(\epsilon) \tag{7}$$

where ${}^T\mu(\epsilon)\otimes\mu(\epsilon) \in \mathbb{R}^{N_m\times N_m}$.

As each basis material is regularized separately, we can directly apply the regularization surrogate of De Pierro [9] demonstrated in a mono-energetic case.

B. Proposed corrections

We now introduce modifications of the method described in the previous subsection. We investigated two solutions in order to stabilize voxels outside the FOV.

1) *Soft-exponential (Soft-E)*: Even though negative values of attenuation $t_i(f,\epsilon)$ are unphysical, they can occur, in particular in early iterations. The calculation of an exponential on a positive number will produce large gradient values causing divergence. To counteract this effect, we investigated the use of the (negative) soft-exponential (adapted from Sidky et al. [7]):

$$\mathrm{softexp}(t) = \begin{cases} e^{-t} & \text{if } t \geq 0, \\ 1 - t & \text{if } t < 0. \end{cases} \tag{8}$$

By replacing the exponential function in (1) with $\mathrm{softexp}(t)$, we expect to lower the effect of negative attenuation values and to prevent numerical instabilities. This leads to straightforward modifications of (6) and (7).

2) *Spatial regularization weights (SRW)*: there is not enough projection data to reconstruct voxels outside the FOV. Intuitively, their values should therefore be more impacted by the regularization with respect to voxels in the FOV. To realize this, we implemented regularization weights which vary according to the geometrical conditioning, formally defined by

$$\lambda_j = \frac{N_\theta}{\sum_{i=1}^{N_p} A_{ij}} \tag{9}$$

with N_θ the number of scanned views in the 360 degree axial scan configuration. The denominator represents the back-projection of a sinogram filled with ones. The number λ_j is equal to 1 in the FOV and is inversely proportional to the number of projections contributing to the voxel update. It is comparable to the scaling factor proposed by Kim et al. [5] to handle helical data with an SQS algorithm for mono-energetic data and used in their data attachment term. Instead, we make regularization stronger in regions out of the FOV. The penalization term (4) becomes

$$R_{\mathrm{SRW}}(f) = \sum_{m=1}^{N_m} \beta_m \sum_{j=1}^{N_v} \lambda_j \sum_{\xi\in\mathbb{V}_j}^{N_v} \phi(f_{mj} - f_{m\xi}). \tag{10}$$

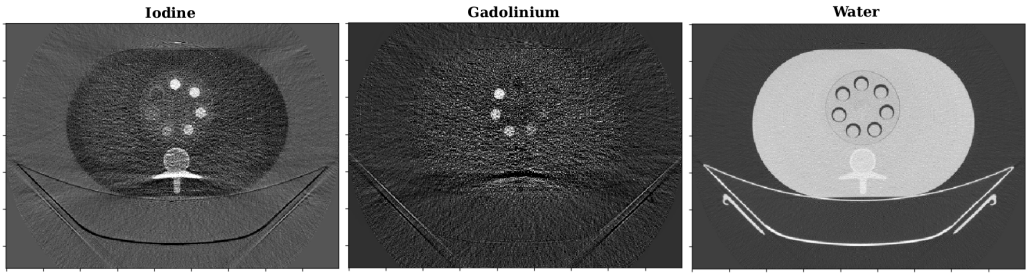


Fig. 2. Iodine, gadolinium and water maps coming from 2D reconstruction configuration. The display windows are [-5,10], [-3,7], and [-500,1500] mg/ml for iodine, gadolinium and water maps, respectively

C. Experiments

We tested the proposed solution on real spectral data of the abdomen QRM phantom acquired with an axial scan on a Philips' prototype clinical scanner recently installed in Lyon (France), which is an upgrade of [10] with mainly a larger photon counting detector. Iodine and gadolinium tubes at different concentrations were placed inside the QRM phantom. For this study, projection data were binned in 462 columns and 16 rows with 2.06 mm $\times$ 2.01 mm pixel size. The source voltage was 120 kV and 2400 views were acquired with a total 330 mA.s exposure.

The phantom was decomposed in iodine/gadolinium/water material maps. Voxel size was 1 mm $\times$ 1 mm $\times$ 3 mm. The FOV is 500 mm $\times$ 500 mm $\times$ 17.5 mm but only 450 $\times$ 330 voxels were reconstructed per slice for this object. A total of 11 slices were reconstructed, 5 central slices with only voxels in the FOV and 3 extra slices at both ends as shown in Fig. 1. The number of subsets was set to 8 and Nesterov's momentum was reset every 50 iterations (i.e. 400 voxel updates). Regularization uses a 27 voxels neighborhood. Reconstructed volumes were post-processed with an image-based correction for ring artifacts.

The material density volumes used for the non-zero initialization (NZI) were obtained with a projection-based material decomposition. The latter was calculated from a maximization of Poissonian likelihood [11] solved with 5000 iterations of the simplex algorithm of Nelder and Mead. The basis sinograms were 3D-median filtered and then reconstructed using a least squares data fidelity term and quadratic spatial regularization solved with the conjugate gradient algorithm. Voxels outside the FOV at each end were replaced from the closest ones in the FOV in the axial direction.

For comparison and to study convergence, the volume was also reconstructed in a simplified configuration. A 2D volume was reconstructed from the two middle rows of the detector averaged to form a one-row detector perfectly aligned with the source and one axial slice. This 2D axial slice was reconstructed with no subset and 1600 iterations. In this 2D configuration, we did not observe the numerical instabilities observed with the 3D configuration of Fig. 1 which is why it was used as a reference to evaluate the central slice of the 3D results (#6 in Fig. 1).

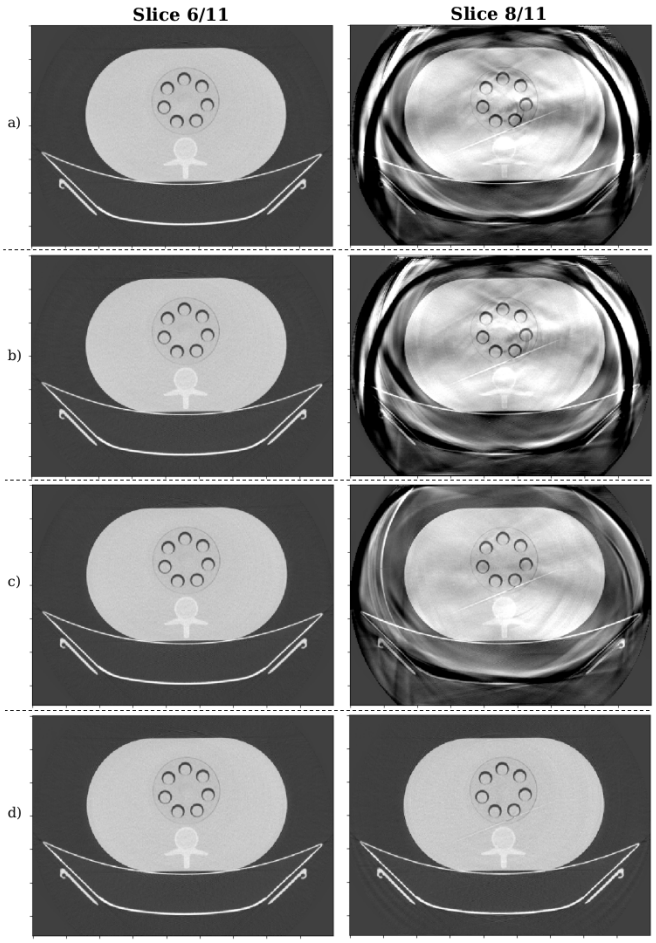


Fig. 3. Water volumes at 100 iterations, slice 6/11 is the central slice and slice 8/11 is the last top slice in FOV (cf Fig. 1. a) Uncorrected method. b) Uncorrected method with NZI. c) With Soft-E. d) With SRW. The display window is [-500,1500] mg/ml.

III. Results

The reference result is displayed in Fig. 2. The reconstruction of K-edge tubes with more than 2 mg/ml is visually satisfactory while the separation between iodine and gadolinium is subject to artifacts in the cortical part of the vertebra equivalent material.

We only show the water maps in Fig. 3, but other material decomposition images showed similar artifacts. This figure

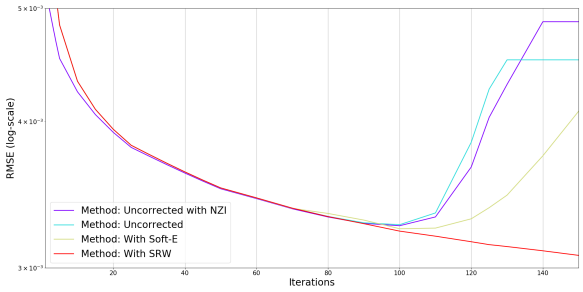


Fig. 4. RMSE between the central and the 2D volume for each methods. (Horizontal line expresses the numerical error appearance.)

corresponds to the reconstruction after 100 iterations. The process was continued until 200 iterations but faced numerical error (*not a number* (NaN) after taking the exponential of a too large number) at 130 iterations for the uncorrected method and 140 for the uncorrected method with NZI (using single precision floats). There was no such error with both the Soft-E and SRW corrections after 200 iterations. Note that we purposely chose large axial slice thickness because the numerical errors occur earlier in the iterations compared to isotropic 1 mm slice thickness but it also occurs with 1 mm slice thickness.

Fig. 3 clearly illustrates the negative effect of voxels out of the FOV and in the FOV with SQS algorithms. The NZI does not help to stabilize the behavior while the image quality with SRW is visually much more satisfactory with a minor extra computation cost. The Soft-E seems to slow down the divergent behavior of the reconstruction but cannot prevent it.

Fig. 4 shows the mean squared error between the central slice and the 2D reference volume:

$$\mathrm{RMSE} = \frac{1}{N_{vs} N_m} \sum_{j=1}^{N_{vs}} \sum_{m=1}^{N_m} \sqrt{\left(f_{bj} - f_{bj}^{2D}\right)^2} \qquad (11)$$

with N_{vs} the voxel number in a slice and f_{bj} the central slice values of each method.

Even if the central slice seems visually satisfactory after 100 iterations (Fig. 3), the RMSE reveals the divergent behavior of methods without SRW. We investigated the combination of Soft-E and SRW in the same algorithm but the contribution of adding Soft-E was negligible, both visually and on the RMSE.

The numerical instabilites outside the FOV have also been noticed without the OS implementation or the use of Nesterov's momentum technique. The SRW methods can be linked to the expression of the Hessian of the surrogate in SQS methods. This term depends on the number

$$\sum_{\xi=1}^{N_v} A_{i\xi} \qquad (12)$$

which corresponds to the projection of a volume filled with ones which appears in the SQS algorithm (7). It weights the influence of a pixel in the current update. Due to geometrical conditioning, this coefficient is smaller in voxels outside the FOV. This term is crucial to accelerate convergence in the FOV: without it, we have observed slower convergence. However, out of the FOV, it leads to larger update steps in Newton's scheme. The SRW is a simple way to preserve its effect in the FOV while guaranteeing convergence in the entire volume grid required in an iterative reconstruction.

Results were demonstrated in an axial configuration scan. We expect the helical case studied by Kim et al. [5] to be similar. The proposed correction can be applied to a helical source trajectory with a slight modification of the spatial weights λ_j.

IV. Conclusion

One-step reconstruction with an SQS algorithm leads to numerical instabilities stemming from voxels out of the FOV. Using a so-called 'soft exponential' only slightly reduces this instability at a significant computational cost. We propose to correct this effect by implementing regularization weights which balance the larger SQS update steps in voxels out of the FOV. This SRW solution is computationally efficient and preserves convergence in the FOV. It can also be used in helical scan configurations where a similar behavior is observed.

Acknowledgment

This work was performed within the framework of the EU's H2020 research and innovation program under the grant agreement No. 633937, the SIRIC LYric Grant INCa-DGOS-4664, and the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR- 11-IDEX-0007) operated by the ANR.

References

[1] B. Chen, Z. Zhang, D. Xia, E. Y. Sidky and X. Pan, "Algorithm-enabled partial-angular-scan configurations for dual-energy CT", *Med. Phys.* vol. 45, no. 5, pp. 1857-1870, May 2018.

[2] C. Mory, B. Sixou, S. Si-Mohamed, L. Boussel and S. Rit, "Comparison of five one-step reconstruction algorithms for spectral CT", *Phys. Med. Biol.*, vol. 63, no. 23, pp. 235001, Nov. 2018.

[3] T. Weidinger, T. M. Buzug, T. Flohr, S. Kappler and K. Stierstorfer, "Polychromatic Iterative Statistical Material Image Reconstruction for Photon-Counting Computed Tomography", *International Journal of Bio. Imag.*, vol. 2016, pp. 1–15, 2016.

[4] K. Mechlem et al., "Joint statistical iterative material image reconstruction for spectral computed tomography using a semi-empirical forward model", *IEEE Trans. on Med. Imag.*, vol. 37 pp. 68-80, July 2017.

[5] D. Kim, D. Pal, J-B. Thibaul, and J. A. Fessler, "Accelerating ordered subsets image reconstruction for X-ray CT using spatially non-uniform optimization transfer", *Trans. Med. Imag.*, vol. 32, no. 11, pp. 1965–1978, Nov. 2013.

[6] S. Tilley II, W. Zbijewski and J. W. Stayman, "Model-based material decomposition with a penalized nonlinear least-squares CT reconstruction algorithm", *Phys. Med. Biol.*, vol. 64, no. 3, pp. 1-23, Jan. 2019.

[7] E. Y. Sidky, R. F. Barber, T. Gilat-Schmidt and X. Pan, "Three material decomposition for spectral computed tomography enabled by block-diagonal step-preconditioning", *Proc. 5th Int. Meeting Image Formation X-ray CT*, pp. 268-271, 2018.

[8] D. Kim, S. Ramani and J. A. Fessler, "Combining Ordered Subsets and Momentum for Accelerated X-ray CT Image Reconstruction", *Trans. Med. Imag.*, vol. 34, no. 1, pp. 167–178, Jan. 2015.

[9] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography", *IEEE Trans. Med. Imag.*, vol. 12, pp. 328-333, June 1993.

[10] D P. Cormode, S. Si-Mohamed et al., "Multicolor spectral photon-counting computed tomography: in vivo dual contrast imaging with a high count rate scanner", *Sci. Rep.*, vol. 7, no. 4784, pp. 1-11, May 2017.

[11] J. P. Schlomka et al., "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography", *Phys. in Med. and Biol.*, vol. 53, no. 15, p. 4031-4047, 2008.

High-Resolution Model-based Material Decomposition for Multi-Layer Flat-Panel Detectors

Yiqun Q. Ma[1], Wenying Wang[1], Matt Tivnan[1], Junyuan Li[1], Minghui Lu[2], Jin Zhang[2], Josh Star-Lack[2], Richard E. Colbeth[2], Wojciech Zbijewski[1], J. Webster Stayman[1]

***Abstract*—Flat-panel detectors have found application in a plethora of dedicated and interventional systems. In part, their success has been to their high-resolution capability in a compact form factor. In this work, we seek to drive improved high-resolution cone-beam CT (CBCT) performance for a new class of multi-layer flat-panel detectors. Such multi-layer detectors are capable of multiple distinct spectral channels which enables spectral CT and material decomposition. In this work, we investigate the use of one-step model-based material decomposition (MBMD) to improve the imaging performance of CBCT based on such detectors. Specifically, we compare traditional filtered back-projection (FBP) followed by image-domain decomposition, idealized MBMD with no blur modeling (iMBMD), and a relatively new MBMD approach that includes system blur modeling (bMBMD). Physical experiments were conducted using a prototype dual-layer detector and a custom high-resolution iodine-enhanced line-pair phantom. We find that both MBMD methods yielded higher resolution decompositions with lower noise than the FBP method, and that bMBMD further improves spatial resolution over iMBMD due to the additional blur modeling. These results demonstrate the advantages of MBMD in resolution performance and noise control over traditional methods for spectral CT - suggesting the potential for high-resolution spectral CT applications at finer resolutions than would be traditionally achieved.**

***Index Terms*—One-step reconstruction and decomposition, Spectral CT, Model-based iterative reconstruction, Quantitative CT, Deblurred CT, Dual-layer flat-panel detectors**

I. Introduction

SPECTRAL CT has found increased application in diagnostic imaging with its ability to improve detection of contrast-enhanced lesions, provide virtual non-contrast images, estimate quantitative density measures of endogenous tissues and contrast agents, and reduce artifacts from beam hardening and metallic implants. Recent advances in flat-panel detector design which use multiple detection layers[1] have the potential to extend these capabilities to cone-beam CT (CBCT) systems including dedicated (e.g., head and neck, extremities, dental, etc.) and interventional CBCT (e.g. mobile and fixed C-arms).

Multi-layer detectors offer many opportunities and challenges. Prototype flat-panel detectors with multiple layers enable energy discrimination in a similar packaging and format as traditional flat-panels by stacking two or more scintillator and semiconductor layers. Often additional filtration is put between the layers to change the x-ray spectrum arriving at the second (or deeper) layers. There are many trade-offs in the design of the scintillator thickness for each layer and the material and thickness of interstitial filters. For example, thicker filtration layers help to provide increased spectral separation between detection layers, which tends to improve material decomposition performance; however, thicker layers result in decreased detection efficiency and, thus, increased noise. Similarly, differences in scintillator thickness can also be used to customize spectral sensitivity; however, varied scintillator thickness can also lead to different spatial resolution in each spectral channel resulting in mismatches that can degrade quantification or lead to edge artifacts in material decomposition.

High spatial resolution is often one of the drivers for using flat-panel detectors. Scintillator thickness presents additional trade-offs for maintaining fine resolution capability with thicker scintillators generally yielding lower spatial resolution. Another potential resolution loss in multi-layer detectors arises from incomplete registration of the layers. That is, the pixel grids of the first and second layer are not perfectly aligned due to manufacturing constraint, or, more generally, are impossible to align exactly due to divergent beam effects. Traditional data processing that resamples data onto a common grid involves interpolation and consequent resolution degradation.

Model-based reconstruction and decomposition offers potential solutions to the challenges of multi-layer flat-panel CBCT. In particular, in contrast to projection-domain[2] and image-domain decomposition[3] in which material density estimates are computed in a staged decompose-then-reconstruct or reconstruct-then-decompose process, there are a growing number of so-called one-step approaches[4;5;6], wherein material density estimates are computed directly from projection data. Not only can these approaches explicitly model the differences in sampling and geometry of the detection layers without the need for data interpolation; these approaches can be modified (as in Wang et al.[7]) to model detector blur effects.

In this work, we adapt model-based material decomposition approaches to a prototype dual-layer flat-panel CBCT system. Both model-based approaches with and without blur modeling are applied and the high-resolution capability of the approaches is investigated and compared to traditional image-domain data processing.

II. Method

A. Model-based Material Decomposition (MBMD)

The following is a brief overview of our implementation of MBMD. We begin with the following forward model:

$$\bar{y}_n = \sum_i b_{in} g_i \sum_m s_{im} \exp\Big(- \sum_k q_{mk} \sum_j a_{ij} \rho_{jk} \Big), \quad (1)$$

This work supported in part by NIH grant R21EB026849 and R01EB025470, and an academic-industry partnership between Johns Hopkins University and Varex Imaging Corporation.

[1]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21205

[2]Varex Imaging Corporation, 683 River Oaks Parkway, San Jose, CA 95134

where the mean measurements, $\bar{y}_n$, are indexed by n (over both detector pixels and rotation angles), and are related to the material density parameters, ρ_{jk}, which are indexed over voxel j and material k. This relationship takes into account the energy-dependent mass attenuation coefficients q_{mk} where m is the energy index; and the coefficients, s_{im}, which model all the elements contributing to spectral sensitivity including incident beam spectrum, scintillator type and thickness, energy weighting, etc. Note that we differentiate between pre-detector blur indices, i, and post-blur indices, n; where the blur kernel is given by $b_i n$. This blur model is general; however, we will focus on independent shift-invariant blur for each detection layer. Also, note that the projection operation, specified by a_{ij} permits arbitrary system geometries including different geometries for each detection layer. Lastly, we permit modeling of additional gain effects (e.g. automatic exposure control) through g_i. This forward model may be written compactly using matrix-vector form as

$$\bar{y} = \mathbf{BGS}\exp\{-\mathbf{QA}\rho\}. \tag{2}$$

We choose to estimate ρ from the measurements y using the following regularized objective function:

$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}} \frac{1}{2}(y-\bar{y})^T K_y^{-1}(y-\bar{y}) + R(\rho) \tag{3}$$

This objective was previously reported in Wang et al.[7] and allows for specification of the inverse of the data covariance and an arbitrary regularizer $R(\rho)$. In this work, we restrict ourselves to weighting by the inverse of the variance of the data and a quadratic regularizer. The objective function was solved using a preconditioned method described in Tivnan et al.[8].

Note that this system model is general and permits the investigation of both algorithms with and without blur modeling. In the case of no blur modeling, the operator $\mathbf{B}$ becomes the identity matrix. In this work, we compared three material decomposition methods: filtered-backprojection followed by image-domain decomposition, idealized MBMD with no detector blur modeling (iMBMD), and model-based material decomposition with system blur model (bMBMD).

B. High-Resolution Material Decomposition Experiments

Physical investigations were conducted on a prototype spectral CBCT test-bench (Figure. 2). The system consists of a rad/fluoro X-ray source (Rad-94, Varex Imaging, San Jose, CA), a rotary stage (Alio Industries, Arvada, CO), and a prototype dual-layer a-Si detector provided by Varex Imaging. Both the "low-energy" flat panel (top layer) and the "high-energy" panel (bottom layer) have 2880 x 2880, 150 µm pixels. The low-energy layer has a 200 µm CsI scintillator, and the high-energy layer has a 550 µm CsI scintillator. Between the two layers exists a 1 mm interstitial copper filter. The detector Modulation Transfer Functions (MTFs) of each layer have been measured in previous work[1] and are shown in Figure 1.

To investigate the high-resolution capabilities of the prototype system and the effect of different data processing

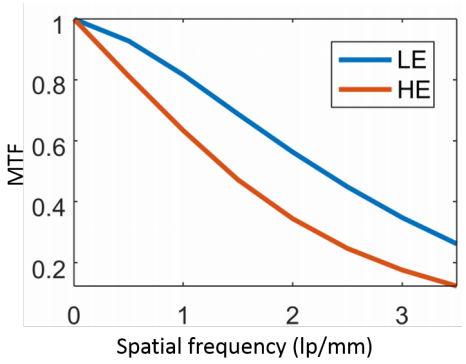


Fig. 1: MTFs of the low energy (LE) and high energy (HE) channels of the detector

schemes, a custom iodine-enhanced phantom was constructed. The phantom consists of nine groups of 3D-printed line pairs immersed in 50 mg/mL iodine solution (Figure. 2). The phantom is designed with variably sized line pairs from 1-5 lp/mm in 0.5 lp/mm increments. Due to resolution limits of the 3D printer used in this work, the Peopoly Moai SLA printer, not all line pairs are accurate representations of the original design. In particular, finer line pairs are printed somewhat thinner and the smallest line pairs are completely fused together. Thus, while this phantom exhibits high-resolution features that help probe the resolution limits of the system, the line pair frequencies themselves are not quantitative.

Projection data were acquired using an X-ray technique of 90 kVp and 1 mAs per frame. The phantom was rotated 360° and 360 projections were acquired with uniform angular sampling. Native 150 µm pixels were used without binning. The source-to-axis distance (SAD) was 828.9 mm, and the average source-to-detector distance (SDD) for the two layers was 1129.2 mm. Each detector layer was calibrated individually using the geometric calibration method described in Cho et al.[9].

As discussed above, three separate reconstructions/material decompositions were performed: FBP and image-domain decomposition, iMBMD, and bMBMD. The blur operator B in the bMBMD forward model accounts for the source and detector blur in each of the two spectral channels. The reconstructed volumes used 400 x 400 x 3, 0.11 mm cubic voxels. For the two MBMD approaches, the iterative algorithm in Tivnan et al.[8] was used with 60 main iterations, each with 5 subiterations for inverse Hessian application. Quadratic regularization strength was chosen to be in a regime where iMBMD resolution was limited by system blur (e.g. smaller regularization values did not improve spatial resolution) and bMBMD regularization strength was chosen to illustrate improved resolution while limiting overall noise. The FBP reconstruction consisted of separate reconstructions of the high- and low-energy channels into volumes with a common coordinate system. Image domain material decomposition was performed using an empirically selected linear combination of the high- and low-energy reconstructions with a normalization to approximate proper concentrations.

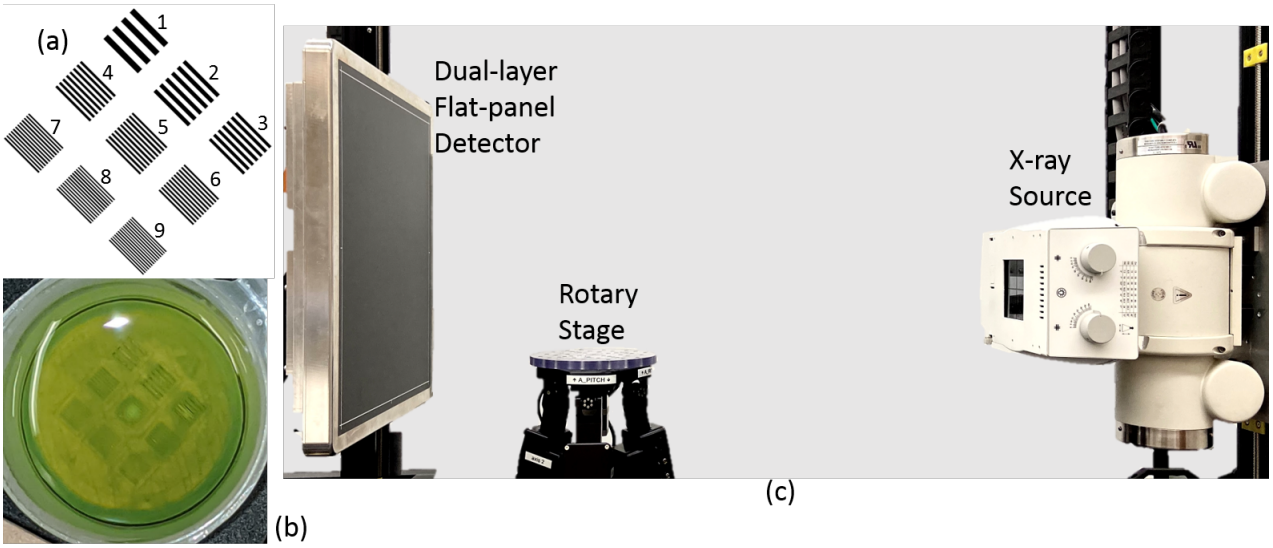


Fig. 2: (a) A high-resolution phantom design and 3D-printed line-pair phantom. Line pair densities of groups 1 - 9 range from 1 lp/mm to 5 lp/mm in 0.5 lp/mm increments (however, not all line pairs appear due to finite spatial resolution of the printer). (b) Photo of the 3D printed phantom immersed in 50 mg/mL iodine solution. (c) Illustration of the experimental dual-layer detector spectral CT test-bench used in physical experiments.

III. Results

Reconstructions are shown in Figure 3. The FBP result suffers from visibly higher noise than the other two methods. This noise was dominated by higher noise in the high-energy channel that propagated through to the decomposition via the linear combination of channels. The iMBMD approaches exhibits less noise in the reconstruction - likely to the improved system model and statistical modeling of noise. Advantages are most prominent in the zoomed images of the iodine basis material. (Note that line pair groups 7 - 9 do not appear in the reconstructions due to 3D-printing restrictions not data processing constraints.) It appears there is also a subtle improvement in spatial resolution in iMBMD over FBP. In the bMBMD images, we see further improved spatial resolution (due to the blur modeling); however, there does appear to be somewhat increased noise. These results are consistent with prior work[10] in model-based iterative reconstruction where blur modeling methods outperform methods without blur modeling in that they extend the feasible range of noise-resolution tradeoffs. That is, where the iMBMD method is limited by the unmodeled system blur, the bMBMD method can achieve higher resolution albeit at increased noise.

IV. Discussion

In this work we have explored the use of one-step model-based material decomposition methods for a dual-layer flat panel prototype system. We have applied standard data processing techniques as well as model-based iterative methods with and without blur modeling. These initial studies suggest that the model-based approaches have the ability to provide increased imaging performance - e.g. decreased noise at similar resolution and/or higher spatial resolution through detailed modeling of the measurement noise, sampling, and system blurs. This work illustrates that model-based processing has the potential to recover imaging performance and help balance the inherent trade-offs present with the design of multi-layer detectors.

In ongoing work, we are refining the imaging targets used for these kinds of high-resolution studies including the use of alternate phantom construction methods and the use of micro-CT to obtain high-resolution ground truth for comparison. Similarly, quantitative analysis of these approaches over a range of regularization strength is underway to illustrate the limiting performance of iMBMD as compared with bMBMD. Moreover, we are exploring alternate regularization schemes to provide increased advantages in the trade-off between noise and resolution.

References

[1] M. Lu, A. Wang, E. Shapiro, A. Shiroma, J. Zhang, J. Steiger, and J. Star-Lack, "Dual energy imaging with a dual-layer flat panel detector," vol. 1094815, no. March, p. 40, 2019.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerised tomography," *Physics in Medicine & Biology*, vol. 21, no. 5, p. 733, 1976.

[3] X. Liu, L. Yu, A. N. Primak, and C. H. McCollough, "Quantitative imaging of element composition and mass fraction using dual-energy ct: Three-material decomposition," *Medical physics*, vol. 36, no. 5, pp. 1602–1609, 2009.

[4] T. G. Schmidt, R. F. Barber, and E. Y. Sidky, "A spectral ct method to directly estimate basis material maps from experimental photon-counting data," *IEEE transactions on medical imaging*, vol. 36, no. 9, pp. 1808–1819, 2017.

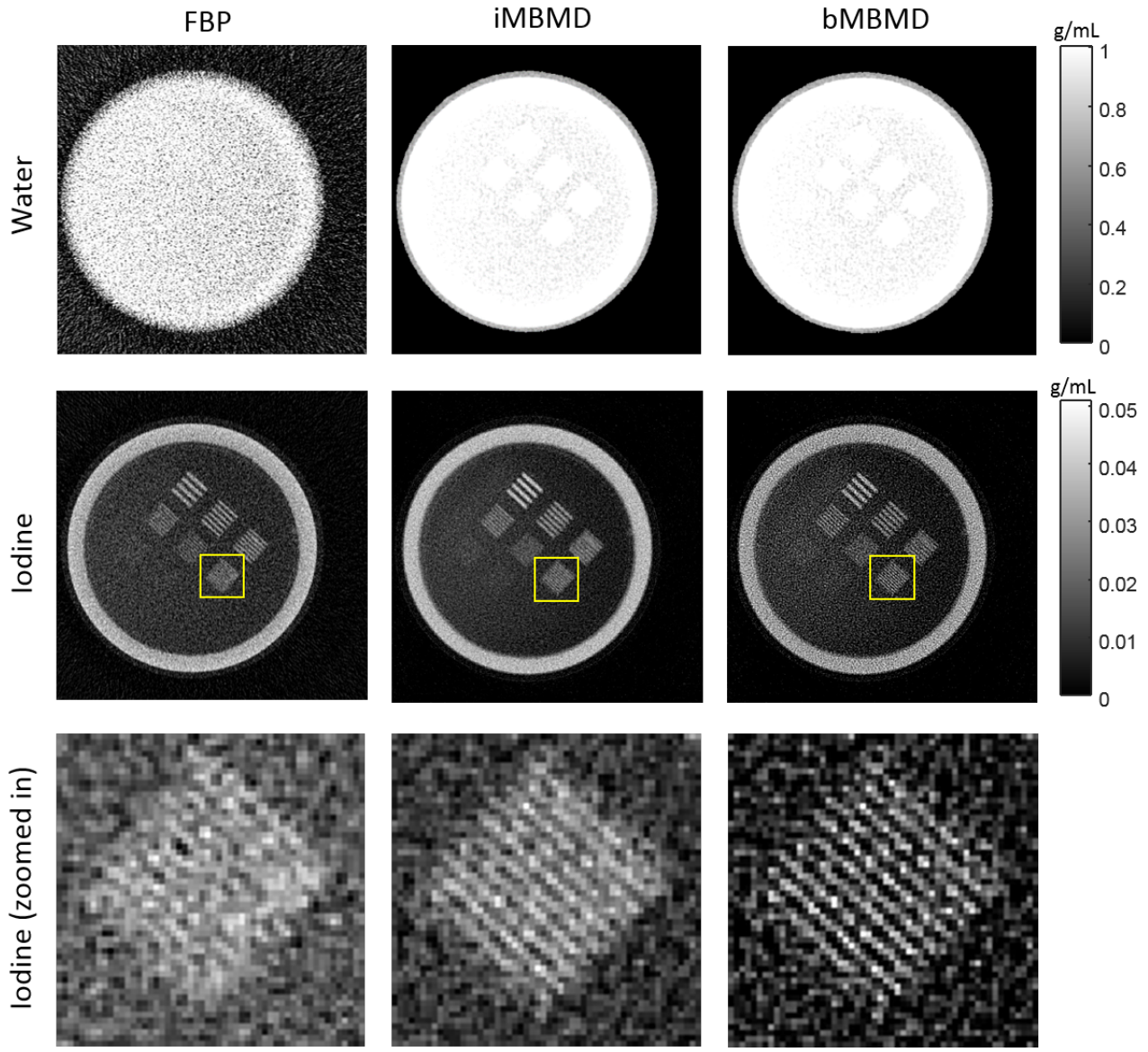


Fig. 3: Reconstruction and material decomposition of the iodinated line-pair phantom using FBP, iMBMD, and bMBMD.

[5] K. Mechlem, S. Ehn, T. Sellerer, E. Braig, D. Münzel, F. Pfeiffer, and P. B. Noël, "Joint statistical iterative material image reconstruction for spectral computed tomography using a semi-empirical forward model," *IEEE transactions on medical imaging*, vol. 37, no. 1, pp. 68–80, 2017.

[6] S. Tilley II, W. Zbijewski, and J. W. Stayman, "Model-based material decomposition with a penalized nonlinear least- squares CT reconstruction algorithm," *Phys Med Biol*, vol. 64, no. 3, 2019.

[7] W. Wang, M. Tivnan, G. J. Gang, Y. Ma, Q. Cao, M. Lu, J. Star-Lack, R. E. Colbeth, W. Zbijewsku, and J. W. Stayman, "Model-based material decomposition with system blur modeling," in *Medical Imaging 2020: Physics of Medical Imaging*. SPIE, 2020.

[8] M. Tivnan, W. Wang, and J. W. Stayman, "Preconditioned algorithms for model-based iterative reconstruction and material decomposition from spectral ct data," *arXiv:3010843*, 2020.

[9] Y. Cho, D. J. Moseley, J. H. Siewerdsen, and D. A. Jaffray, "Accurate technique for complete geometric calibration of cone-beam computed tomography systems," *American Association of Physicists in Medicine (AAPM)*, Mar 2005.

[10] S. Tilley, M. W. Jacobson, Q. Cao, M. Brehler, A. Sisniega, W. Zbijewski, and J. W. Stayman, "Penalized-likelihood reconstruction with high-fidelity measurement models for high-resolution cone-beam ct imaging," *IEEE Transactions on Medical Imaging*, vol. 37, no. 4, pp. 988–999, 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/29621002 https://ieeexplore.ieee.org/document/8125700/

Triple-Energy X-Ray CT Using Spectral Modulator with Flying Focal Spot: Modulator Design and Scatter-Modeled Material Decomposition

Yifan Deng, and Hewei Gao*

***Abstract*—For X-ray computed tomography (CT), spectral modulator with flying focal spot (SMFFS) technology could be a promising low-cost approach to accurately solving the X-ray scattering problem and physically enabling multi-energy imaging in a unified framework, with no significant sparsity or misalignment in data sampling. In this work, we present a design of the spectral modulator to essentially generate a triple-energy CT imaging with the focal spot deflecting at both X and Z directions, and develop a scatter-modeled material decomposition (SMMD) for such a triple-energy SMFFS. Taking advantage of a similarity among scatter distributions in spectral data using SMFFS, the SMMD algorithm aims at directly removing scatter's impact during the material decomposition process. As a feasibility study, energy-separation of the triple-energy projection rays is first analyzed with optimal modulator material identified; a Monte Carlo simulation is then conducted to re-confirm that the X-ray scatter distributions of scanned object have strong correlations among the varying focal spot positions; after that preliminary evaluations of a custom-manufactured 2D Molybdenum modulator on an experimental X-ray system validate our design of the spectral modulator; and finally SMMD results from numerical simulation of SMFFS with multi iodine contrasts further demonstrate that accurate material decomposition is achievable when the scatter-to-primary rater is as high as 150%.**

***Index Terms*—Material Decomposition, Multi-Energy CT, Spectral Modulator, Flying Focal Spot**

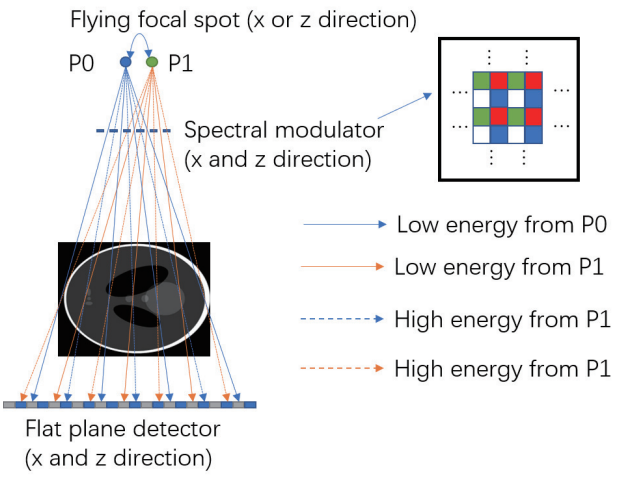


Fig. 1. Triple-energy data sampling enabled by a mixed 2D modulator and two-directional focal spot deflections. The mixed 2D modulator can be made of two 1D modulator overlapped together. The "white", "green", "blue", and "red" blockers (sub-cell areas) represent four levels of filtration to the X-ray beam passing through them. When the focal spot deflects in both X and Z directions sequentially, three levels of effective energies of projection rays can be collected at roughly the same sampling position.

I. INTRODUCTION

In recent years, modulating the X-ray source in computed tomography (CT) has advanced greatly in both academia and industry, such as the technologies using the primary modulator for scatter correction and dual-energy CT [1], [2], the splitting filter in twin-beam dual-energy CT by Siemens [3], the dynamic bowtie filter [4], and the dynamic fluence field modulation [5] and the associated spatial-spectral modulator [6]. Among them, the primary modulator and the splitting filter are stationary with respect to the X-ray source, having advantages of simplicity in mechanical design and practical implementations. However, a stationary modulator is known to suffer from data sparsity or misalignment of projection rays in terms of spectral data sampling, especially for circular scan trajectories such as cone-beam CT imaging.

Recently, a new concept of multi-energy imaging using a spectral modulator and flying focal spot (SMFFS) technology has been proposed by our group [7]. It has no significant sparsity or misalignment in data sampling and could potentially enable full-scale cone-beam CT spectral imaging and simultaneous scatter correction capabilities.

In this work, we present a design of the spectral modulator that can essentially generate a triple-energy CT imaging with the focal spot deflecting at both X and Z directions. To improve the performance of quantitative imaging in presence of high scatter-to-primary ratio (SPR), a scatter-modeled material decomposition (SMMD) is therefore proposed for such a triple-energy SMFFS to correct for scatter during the material decomposition process. In SMMD, a key assumption on scatter distribution from SMFFS is leveraged: the X-ray scatter from the scanned object is mostly dominated by low-frequency components and there exists strong correlation among scatter distributions at varying source positions as long as the distance of focal spot deflection is relatively small.

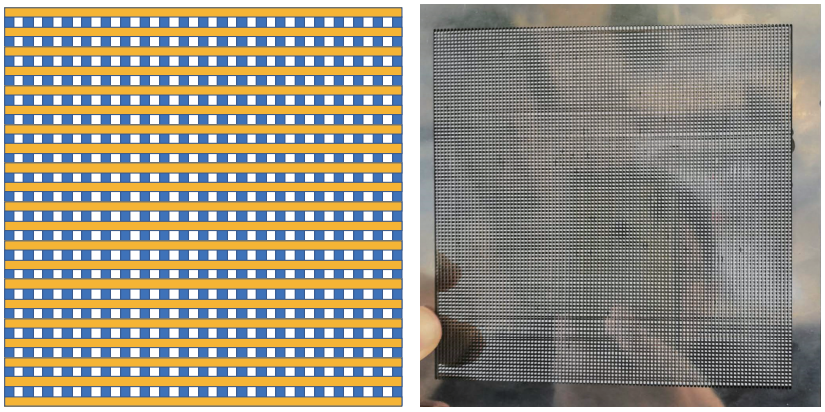
II. METHOD

A. Design of a Mixed two-dimensional (2D) Modulator

A spectral modulator is made of partially attenuating blockers and placed between the X-ray source and the scanned

This project was supported in part by the New Faculty Startup Fund from Tsinghua University, Beijing, China (53331100120). Yifan Deng and Hewei Gao are with the Department of Engineering Physics, Tsinghua University, Beijing 100084, China.

*Author to whom correspondence should be addressed. Email: hwgao@tsinghua.edu.cn.



(a) (b)

Fig. 2. A mixed 2D spectral modulator made of two 1D modulator stacking together that can enable triple-energy imaging for SMFFS. (a) A conceptual design; and (b) a manufactured modulator from the "Mo+Mo" combination.

object, which can attenuate the intensities for some of the X-ray beam and filters their spectra essentially. Along with the focal spot deflection, multi-energy imaging can be realized with no significant sparsity or misalignment in data sampling. As illustrated in Fig. 1, triple-energy imaging can be generated by SMFFS if the spectral modulator contains at least three levels of filtration (including no filtration). Ideally, one would like to directly build a modulator whose material and effective thickness can vary from one blocker (a sub-cell area) to another. However, such a modulator is too difficult to make in reality. For ease in manufacturing, we therefore stack two one-dimensional (1D) strip modulators together to form a mixed two-dimensional (2D) mesh modulator as shown in Fig. 2. Flexibility given by making two 1D modulators separately allows us to further optimize the modulator design, especially to maximize the energy separation at a given attenuation rate.

B. Energy Separation from Triple-Energy SMFFS

A spectral modulator shares a lot of common properties with the primary modulator that has been developed for scatter correction. One thing that is very different from the regular primary modulator for scatter correction is that instead of minimizing the spectral change caused by the modulator one want to do the opposite way.

For material decomposition, energy separation is an important factor to be considered. Here, we calculate the energy separation at given photon flux attenuation (α) for a variety of materials for the two 1D modulators. The energy separation between two different spectra ($S_1(E)$ and $S_2(E)$) can be calculated as

$$\Delta E_{\text{sep}} = abs(\frac{\int ES_1(E)\mathrm{d}E}{\int S_1(E)\mathrm{d}E} - \frac{\int ES_2(E)\mathrm{d}E}{\int S_2(E)\mathrm{d}E}). \quad (1)$$

Here, the difference of the mean energies of two separate spectra is taken as the energy separation. In this study, the attenuation of the first layer is set as 40%, total attenuation is 32%, the X-ray source is operated at 140 kVp, S_i represents the energy spectrum of X-ray passing through the i_{th} sub-cell vertically. Here $i = 0$ represents sub-cell of air, $i = 1$ represents the sub-cell of layer1 material, $i = 2$ represents sub-cell of layer2 material, and $i = 3$ represents sub-cell of layer1 and layer2 material overlapped. Some typical metal materials that may be used to build the modulator are analyzed as in Table. I. It is observed that the best choice in terms of modulator material should be relatively low-Z element (e.g., Molybdenum) for the first layer, and high-Z element (e.g., Gold) for the second layer.

TABLE I
Energy separation (keV) for a variety of materials for layer 1 (MAT-1) and layer 2 (MAT-2), including the energy separation from the residual spectra in the triple-energy SMFFS ($S_0 - S_3$,$S_1 - S_3$), and between regular spectra (S_0,S_1;S_2,S_3).

MAT-1	MAT-2	$[S_0 - S_3, S_1 - S_3]$	$[S_0, S_1]$	$[S_2, S_3]$
Al(13)	Au(79)	8.26	7.74	7.12
Cu(29)	Au(79)	19.14	14.12	12.39
Mo(42)	Au(79)	20.20	14.84	13.10
Au(79)	Au(79)	5.07	5.45	4.06
Mo(42)	Al(13)	17.32	14.84	13.29
Mo(42)	Mo(42)	10.11	14.84	11.40
Mo(42)	Ta(73)	18.44	14.84	14.22
Mo(42)	Pb(82)	19.93	14.84	12.59

C. Scatter-Modeled Material Decomposition

A quite unique property of multi-energy SMFFS is that, theoretically, one can calculate the scatter distribution during basis material decomposition as long as the number of measurements are more than that of unknowns. Details are as follows.

The polychromatic measurement at a detector pixel can be expressed as

$$I_{t_k} = \int_0^{E_{\max}} S_k(E) e^{-\mu_1(E)L_1 - \mu_2(E)L_2} \mathrm{d}E + I_{s_k}, \quad (2)$$

where, I_{t_k} is the total signal at k_{th}-energy received by the detector with the effective spectrum for primary being $S_k(E)$; $\mu_1(E)$, $\mu_2(E)$, L_1 and L_2 denote the attenuation coefficients

and path-lengths for basis material 1 and 2, respectively; and I_{s_k} is the scatter signal at the k_{th}-energy received by detector.

Based on the fact that scatter is mostly low-frequency dominated, it is reasonable to say that before and after moving of the focal spot by a relatively small distance, the scattering distributions are very much similar. For example, if we simply assume they are basically the same, eliminating the scatter term becomes extremely easy, which can be done by a subtraction as

$$I_{t_i} - I_{t_k} = \int_0^{E_{\max}} [S_i(E) - S_k(E)] e^{-\mu_1(E)L_1 - \mu_2(E)L_2} \mathrm{d}E. \quad (3)$$

After the subtraction, $I_{t_i} - I_{t_k}$ become a new set of CT data, corresponding to a residual spectrum $S_i(E) - S_k(E)$. Given three signals are collected at three energy levels from a triple-energy SMFFS scan, one can get a pair of new spectral projections without scatter that are enough for a two-material decomposition, a typical case in a dual-energy CT imaging. As a result, conventional material decomposition such as projection-domain basis material decomposition can be adopted here.

For triple-energy SMFFS, $I_{t_1} - I_{t_3}$, $I_{t_2} - I_{t_3}$ can be used, with I_{t_1}, I_{t_2} and I_{t_3} correspond to low-, medium-, and high-energy projection rays, respectively. According to analysis in Table I, their energy separation depends on how to the modulator design, in particular the modulator materials and thicknesses. Nevertheless, a triple-energy SMFFS enables a dual-energy material decomposition and scatter correction in a unified framework. Of course, using the outcome of the basis material decomposition, one can also get the scatter distribution I_s if needed.

Also, it is worth noting that given the low-frequency property of X-ray scatter distribution, triple-energy SMFFS by two-directional focal spot deflection can also be done sparsely in view direction (to get the scatter distribution from few views), while keep doing a dual-energy SMFFS scan by one-directional deflection of focal spot at a regular view sampling.

D. MC Simulation of Scatter Distribution

To validate the key assumption on the similarity of scatter distribution in triple-energy SMFFS, we conducted a Monte Carlo (MC) simulation of a water cylinder of 20 cm in diameter and 160 mm in height at 120 kVp using the GEANT4 software. In our MC simulation, a "Mo+Au" double-layer modulator is chosen. The first layer is 0.3 mm thick of Mo, while the second layer is 0.025 mm of Au. Both the interval and width of the strips are set as 0.55mm. Three detector signals can be collected by moving X-ray source point by 0.75 mm in X or Z direction from the origin, respectively.

III. RESULT

The MC simulation of scatter distribution for triple-energy FFS is shown in Fig. 3. The total photon number of per detector pixel is about $\sim 10^5$. We summed up the photons at a few detector pixels together and plot the scatter profiles in both vertical and horizontal directional as illustrated in Fig. 3 as well. Overall, it is observed that scatter distributions at the three focal spot positions looks very similar, validating our assumptions.

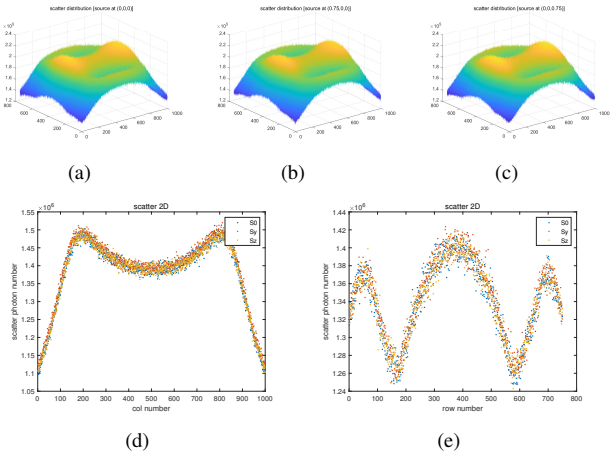


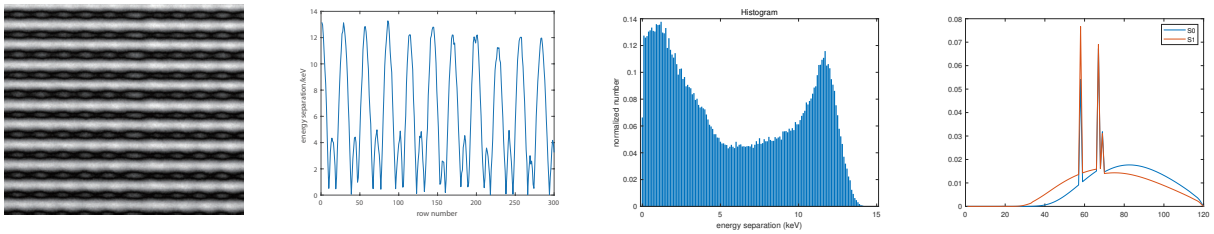
Fig. 3. Scatter distribution of a water cylinder by the MC simulation. (a) scatter distribution at S0 position [focal spot located at (0,0,0) mm]; (b) scatter distribution at Sy position [focal spot located at (0.75,0,0) mm]; (c) scatter distribution at Sz position [focal spot located at (0,0,0.75) mm]; (d) a 1D profile of scatter summed up over rows from 322 to 328; and (e) a 1D profile of scatter summed up over columns from 497 to 503

Based on the analysis in Table. I, and taken the manufacturing difficulty into account, we successfully made a mixed 2D modulator with "Mo + Mo" combination as a compromise in this study, whose first layer is 0.4 mm thick and second layer blocker is 0.1 mm thick. We measured the transmission of a manufactured "Mo+Mo" modulator on an X-ray system operated at 120 kVp (the maximum kVp available), from which the effective modulator thickness, the corresponding effective spectrum and the related energy separations at each detector pixel were calculated using an approach developed in [8]. As shown in Fig. 4, it is observed that the highest energy separation is about 13 keV, roughly consistent with our theoretical calculations.

Moreover, a numerical simulation of a water cylinder that consists of four iodine contrasts (5, 10, 15 and 20 mg/cc iodine solution) was conducted to generate data from a triple-energy SMFFS scan. The scatter to primary ratio(SPR) was set at 150% without any noise. After material decomposition using our proposed SMMD method, we picked a portion of projections with relatively high energy separation for reconstruction using the standard filtered-backprojection algorithm. The material decomposition images are shown in Fig. 5, where iodine and water basis images are successfully separated, despite that the reconstruction needs further improvement to remove image artifacts.

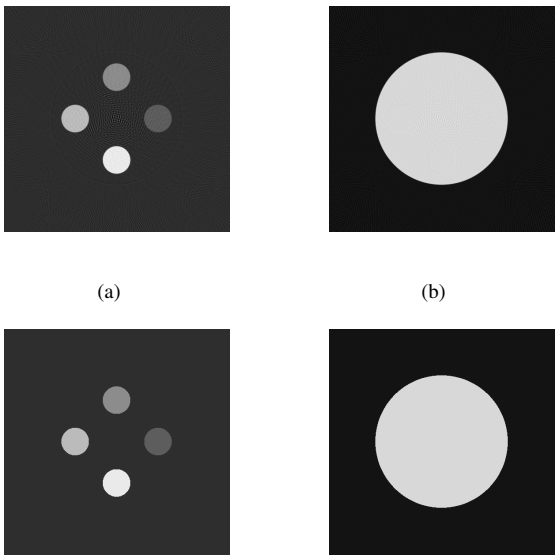
IV. CONCLUSION

In this work, we presented a design of the spectral modulator to essentially generate a triple-energy CT imaging with the focal spot deflecting at both X and Z directions, and proposed a scatter-modeled material decomposition for such a triple-energy CT scan using a spectral modulator with the flying focal spot technology. A triple-energy spectral modulator made of "Mo+Mo" combination has been manufactured and



(a) (b) (c) (d)

Fig. 4. Measured energy separation and the associated effective spectra from a manufactured "Mo+Mo" 2D modulator. (a) Mean energy separation between two focal spot positions; (b) a vertical profile of the mean energy separation; (c) the histogram of energy separation from the entire scan field; and (d) the effective spectra estimated.



(a) (b)

(c) (d)

Fig. 5. The material decomposition images by the SMMD for a triple-energy SMFFS simulation. (a) Decomposed iodine basis image; (b) decomposed water basis image; (c) iodine reference image; and (d) water reference image.

evaluated. Monte Carlo and numerical simulations validate our theory and method to be effective.

In our future studies, physical experiments on triple-energy SMFFS scans will be conducted. We will also explore iterative image reconstruction algorithms to improve image quality after material decomposition.

REFERENCES

[1] L. Zhu, N. R. Bennett, and R. Fahrig, "Scatter correction method for x-ray ct using primary modulation: Theory and preliminary results," *IEEE transactions on medical imaging*, vol. 25, no. 12, pp. 1573–1587, 2006.

[2] M. Petrongolo and L. Zhu, "Single-scan dual-energy ct using primary modulation," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1799–1808, 2018.

[3] I. P. Almeida, L. E. J. R. Schyns, M. C. Öllers, W. van Elmpt, K. Parodi, G. Landry, and F. Verhaegen, "Dual-energy ct quantitative imaging: a comparison study between twin-beam and dual-source ct scanners," *Medical Physics*, vol. 44, no. 1, pp. 171–179, 2017.

[4] P. Shunhavanich, S. S. Hsieh, and N. J. Pelc, "Fluid-filled dynamic bowtie filter: Description and comparison with other modulators," *Medical Physics*, vol. 46, no. 1, pp. 127–139, 2019.

[5] G. J. Gang, A. Mao, W. Wang, J. H. Siewerdsen, A. Mathews, S. Kawamoto, R. Levinson, and J. W. Stayman, "Dynamic fluence field modulation in computed tomography using multiple aperture devices," *Phys. Med. Biol.*, vol. 64, no. 10, p. 105024, 2019.

[6] M. Tivnan, S. T. II, and J. W. Stayman, "Physical modeling and performance of spatial-spectral filters for ct material decomposition," *Proc. SPIE*, vol. 10948, 2019.

[7] H. Gao, H. Zhou, L. Zhu, N. Pelc, N. R. Bennett, and A. Wang, "Spectral modulator with flying focal spot for cone-beam ct: A feasibility study," in *Medical Imaging 2020: Physics of Medical Imaging*, vol. 11312. International Society for Optics and Photonics, 2020, p. 113121P.

[8] H. Gao, L. Zhang, R. Grimmer, and R. Fahrig, "Physics-based spectral compensation algorithm for x-ray ct with primary modulator," *Physics in Medicine & Biology*, vol. 64, no. 12, p. 125006, 2019.

Data-Domain Decomposition Strategy in Short-Scan Configurations for 2D Dual-Energy CT

Daxin Shi

***Abstract*– In this work, we provide a data-domain decomposition strategy in a short-scan configuration for two-dimensional (2D) fan-beam dual-energy CT. The short-scan configuration employed in dual-energy CT was recently proposed in the literature. It is not very clear how to apply data-domain decomposition methods to this scanning protocol in three-dimensional (3D) imaging tasks. In general, the data redundancy condition in conventional 2D fan-beam CT does not hold any more due to the poly-chromaticity of the x-ray source spectrum. However, for the data domain decomposition problem of the 2D fan-beam dual-energy CT, the only thing we need is to have two distinct energy measurements of the same ray-sum. It turns out we can still use the complementary ray information to propose a data-domain decomposition strategy for the short-scan configuration. In this work, we investigate such a strategy and conduct simulation studies to validate our proposed method.**

I. Introduction

DUAL-energy CT regained its research interest due to its various clinical applications and the revolution of hardware techniques such as fast-kV switching and dual-source techniques. However, high-end CT facilities equipped with these new hardwares are generally expensive. To realize the dual energy imaging functionality, the traditional CT scanner employs two full scans with distinct voltage in each scan which is referred to as the slow-kV switching technique [1]. The slow-kV switching technique was generally complained for less efficiency and high dose to patients and also susceptible to subject motion. In this paper, we will call the slow-kV switching technique as the full-scan technique for short. To mitigate these disadvantages of the full-scan technique, a short-scan configuration has been proposed recently [2]. In the short-scan configuration a conventional circular scanning trajectory was employed and total $2\pi + 2\gamma_m$ angular ranges of the trajectory was illuminated with two distinct voltages where each voltage was responsible for each $\pi + \gamma_m$ angular ranges. Here γ_m denotes the usual full fan angle of a CT scanner. The belief was that the short-scan configuration can readily be realized on standard CT scanners without additional hardware simply by use of the standard single-kV switch technique [2]. Compared with the full-scan technique which generally covers 4π angular range, the short-scan configuration might have a chance to improve imaging efficiency and lower dose delivered to patients.

Accompanied with CT facilities which provide the dual-energy imaging functionality are the reconstruction methods which can address the nonlinear spectral response. In general reconstruction methods to handle dual energy CT data can be roughly classified as two categories which are two-step approach and one-step approach, respectively. In the two-step approach there exist two different types of methods according to the order when the nonlinear spectral response was handled. When the nonlinear spectral response is processed first in the projection domain and followed by an image reconstruction method, one generally refers to this procedure as data domain decomposition approach [1] [3]. When images were reconstructed first and followed by nonlinear spectral response processing techniques, one generally refers to this type of procedure as image domain decomposition method [4-5]. In the one-step approach, the nonlinear spectral response was generally built into the reconstruction procedure [2]. Usually analytic methods were employed in the reconstruction stage in the two-step approach while iterative methods were involved in the one-step approach.

A one-step approach has been proposed for the short-scan configuration for dual energy CT recently [2]. Unlike the full-scan technique where data domain decomposition methods can be developed in a straightforward manner [3], it is not very clear how to develop a data domain decomposition method for the short-scan configuration since ray overlapping condition is not satisfied in general. For the short-scan configuration, one could employ the image domain decomposition method since the angular coverage is $2\pi + 2\gamma_m$. A possible approach is to reconstruct two sets of image volumes by projection data collected in each $\pi + \gamma_m$ angular range for each energy and followed by nonlinear spectral response processing. In the reconstruction stage, the well-known Parker weighting scheme can be employed [5]. Although data-domain decomposition difficulties lie in the general 3D dual-energy CT imaging tasks with short-scan configuration, we propose in this work a data-domain decomposition strategy for the 2D dual-energy CT imaging task. Our approach is based on the data redundancy conditions inherent in the 2D fan-beam CT reconstruction problem. After identification of the redundancy conditions in dual energy CT data, we can employ the well-established data domain decomposition method [3] to process the nonlinear spectral response followed by a conventional filtered backprojection algorithm to yield desired monochromatic images. When the cone-angle is relatively small, the strategy proposed in this work can be served as an approximation of the reconstruction problem. The method proposed in this work can also be employed to yield a seed image for the one-step iterative reconstruction methods.

D. Shi is with Campo Medical Imaging Research Center USA, L.L.C. and Liaoning Campo Medical Imaging Technology Co., Ltd. No. 18, Shennong Street, Hi-tech Industrial Development Zone, Benxi, Liaoning, PRC, 117000, (e-mail: shidx@campo-imaging.com).

II. The Dual-energy Projection Data and Data Domain Decomposition Technique

Projection data acquisition procedures for fan-beam dual-energy CT are schematically illustrated in Fig. 1 (a) and (b) for the full-scan and short-scan configurations, respectively. In the full-scan technique, patients are scanned twice by a low kVp settings followed by a high kVp settings where both scan cover 2π angular range. Fig. 1(b) describes the short-scan configuration where low and high kVp settings are used in the angular range $[0, \pi + \gamma_m)$ and $[\pi + \gamma_m, 2\pi + 2\gamma_m)$, respectively. The light and heavy circles or arcs denote scanning trajectories with the low and high kVp settings, respectively, in Fig. 1(a) and (b). Here we use the quantities β and γ to denote the view angle and fan angle, respectively.

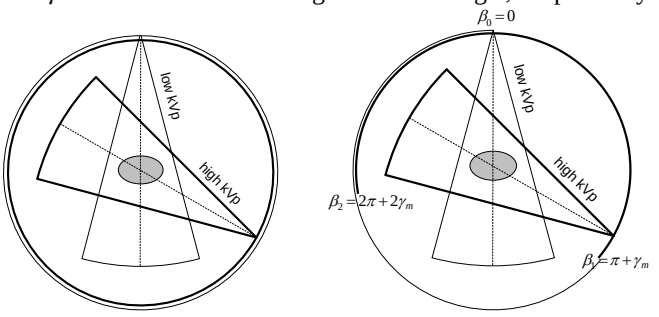


Fig.1. Traditional slow-kV switch scanning protocol (left) full-scan configuration and recently proposed (right) short-scan configuration for dual-energy CT.

The dual-energy projection data g can be described by the following data model,

$$g = -\log\left\{\int S(E)\mathrm{e}^{-\int \mu(E,\vec{r})\mathrm{d}l}\,\mathrm{d}E\right\} \quad (1)$$

where $S(E)$ is the normalized effective spectra after wedge which also accounts for the detector response. Due to the existence of the wedge filter, $S(E)$ is also a function of channel γ which we will add in later. Similarly the projection data g is a function of β and γ where β denotes the view angle. The linear attenuation coefficient of the object (patient), denoted by $\mu(E,\vec{r})$, is a function of position $\vec{r}$ and energy E. The basic assumption in dual-energy CT is that the linear attenuation coefficient $\mu(E,\vec{r})$ can be decomposed by two bases functions, i.e.,

$$\mu(E,\vec{r}) = \mu_1(E)c_1(\vec{r}) + \mu_2(E)c_2(\vec{r}) \quad (2)$$

where the bases functions $\mu_1(E)$ and $\mu_2(E)$ are typically chosen as water and bone linear attenuation coefficients, respectively. Hence the decomposition coefficients $c_1(\vec{r})$ and $c_2(\vec{r})$ are called water- and bone-basis images, respectively, which we will call water- and bone-images, respectively, for short in this paper. This approach is generally known as material-based decomposition. The imaging task in the slow-kV switch dual-energy CT is to reconstruct the water- and bone-images by a collection of projection data with distinct kVp settings. For simplicity, we will call the projection data acquired with high- and low-kVp settings as high- and low-energy data, respectively. Once the water- and bone-images are available, we can synthesize monochromatic images with Eqn. (2) at any given energy E.

On substitution of Eqn. (2) into Eqn. (1), we have

$$g = -\log\left\{\int S(E)\cdot \mathrm{e}^{-\mu_1(E)L_1-\mu_2(E)L_2}\cdot \mathrm{d}E\right\} \quad (3)$$

where

$$L_1 = \int c_1(\vec{r})\,\mathrm{d}l \quad (4)$$

and

$$L_2 = \int c_2(\vec{r})\,\mathrm{d}l\,. \quad (5)$$

Again, we have omitted the β and γ dependence of the quantities in Eqns. (3)-(5) and will put them back later. One can find that L_1 and L_2 are x-ray transforms of the basis images $c_1(\vec{r})$ and $c_2(\vec{r})$, respectively. Once we have data L_1 and L_2 available for each ray, we can employ any convenient reconstruction algorithm to yield the basis images. Now the key is to recover the quantities L_1 and L_2 from Eqn. (3) and this is why we need two distinct energy measurements for each ray. When high- and low-energy datasets are available, we have a system of two non-linear equations with two unknowns,

$$g_H(\beta,\gamma) = -\log\left\{\int S_H(E,\gamma)\mathrm{e}^{-\mu_1(E)L_1(\beta,\gamma)-\mu_2(E)L_2(\beta,\gamma)}\,\mathrm{d}E\right\} \quad (6)$$

and

$$g_L(\beta,\gamma) = -\log\left\{\int S_L(E,\gamma)\mathrm{e}^{-\mu_1(E)L_1(\beta,\gamma)-\mu_2(E)L_2(\beta,\gamma)}\,\mathrm{d}E\right\} \quad (7)$$

where we use $g_H(\beta,\gamma)$ and $g_L(\beta,\gamma)$ to denote high- and low-energy data collected at view angle β and fan angle γ, respectively. Similary, the normalized effective spectrum with low- and high-kVp settings are denoted by $S_L(E,\gamma)$ and $S_H(E,\gamma)$, respectively. In the full-scan configuration, $g_H(\beta,\gamma)$ and $g_L(\beta,\gamma)$ are overlapped for each β and γ. Data-domain decomposition algorithms are available [3] to solve for $L_1(\beta,\gamma)$ and $L_2(\beta,\gamma)$ and hence reconstructions for the basis images $c_1(\vec{r})$ and $c_2(\vec{r})$.

III. Data-Domain Decomposition Strategy in Short-Scan Configuration

A key requirement of data-domain decomposition technique is that each ray must overlap for every measurement with distinct energy settings. In the short-scan configuration in Fig. 1(right), rays in the angular range $\beta \in [0, 2\gamma_m]$ are overlapped so that we can employ a data-domain decomposition method [3] to recover $L_1(\beta,\gamma)$ and $L_2(\beta,\gamma)$. On the hand, the ray overlapping condition is not satisfied in the angular range $\beta \in (2\gamma_m, 2\pi)$. It seems that one will get stuck with the data-domain decomposition technique since the usual data redundancy condition $g(\beta,\gamma) = g(\beta+\pi+2\gamma,-\gamma)$ does *not* hold any more due to the fact that poly-chromaticity of the x-ray source spectrum is now considered. However, for the 2D fan-beam case considered in this paper, complementary data in the 2D dual-energy CT problem can still be employed to remedy the situation to enable the data-domain decomposition technique.

The observation is that the quantities $L_1(\beta,\gamma)$ and $L_2(\beta,\gamma)$ in Eqns. (4) and (5) are x-ray transforms of the basis images $c_1(\vec{r})$ and $c_2(\vec{r})$, respectively, which obey the following well-known data redundancy conditions

$$L_1(\beta,\gamma) = L_1(\beta+\pi+2\gamma,-\gamma) \quad (8)$$

and

$$L_2(\beta,\gamma) = L_2(\beta+\pi+2\gamma,-\gamma)\,. \quad (9)$$

We can employ these conditions to develop a data-domain decomposition strategy for the short-scan configuration.

For $\beta \in (2\gamma_m, \pi + \gamma_m)$, we have only low-energy data $g_L(\beta, \gamma)$ available. However, we can evaluate Eqn. (6) at $(\beta_H = \beta + \pi + 2\gamma, -\gamma)$ to have the following extra equation:

$$g_H(\beta_H, -\gamma) = -\log\left\{\int S_H(E, -\gamma) \mathrm{e}^{-\mu_1(E)L_1(\beta,\gamma) - \mu_2(E)L_2(\beta,\gamma)} \, \mathrm{d}E\right\} \quad (10)$$

where the data redundancy conditions in Eqns. (8) and (9) have been used. Now Eqns. (10) and (7) will form a system of two nonlinear equations with two unknowns which can be solved for the quantities $L_1(\beta, \gamma)$ and $L_2(\beta, \gamma)$ in the angular range $\beta \in (2\gamma_m, \pi + \gamma_m)$. We should check the high-energy data availability for this strategy when $\beta = 2\gamma_m$ and $\gamma = -\frac{\gamma_m}{2}$ which is the minimum angular requirement for the high-energy data. Indeed high-energy data at view angle $\beta_H = \pi + \gamma_m$ are available when $\beta = 2\gamma_m$ and $\gamma = -\frac{\gamma_m}{2}$. Similarly, high energy data are also available at view angle $\beta_H = 2\pi + 2\gamma_m$ which is the maximum angular requirement when $\beta = \pi + \gamma_m$ and $\gamma = \frac{\gamma_m}{2}$.

For $\beta \in [\pi + \gamma_m, 2\pi)$, we have only high energy data $g_H(\beta, \gamma)$ available. But we can use the same strategy to enable data-domain decomposition technique. Let Eqn. (7) be evaluated at $(\beta_L = \beta + \pi + 2\gamma, -\gamma)$. One can get

$$g_L(\beta_L, -\gamma) = -\log\left\{\int S_L(E, -\gamma) \mathrm{e}^{-\mu_1(E)L_1(\beta,\gamma) - \mu_2(E)L_2(\beta,\gamma)} \, \mathrm{d}E\right\} \quad (11)$$

where again Eqns. (8) and (9) have been used. Eqns. (11) and (6) will form a system of equations to enable data-domain decomposition technique to yield $L_1(\beta, \gamma)$ and $L_2(\beta, \gamma)$. The minimum angular requirement of low energy data availability can be checked when $\beta = \pi + \gamma_m$ and $\gamma = -\frac{\gamma_m}{2}$. We have $\beta_L = 2\pi$ in which case the low energy data are indeed available. The maximum angular requirement of low energy data availability can also be checked at view angle $\beta_L = \pi + \gamma_m - \Delta$ when $\beta = 2\pi - \Delta$ and $\gamma = \frac{\gamma_m}{2}$. The quantity Δ means that we need not include the view angle 2π which had already been taken care of in the range $\beta \in [0, 2\gamma_m]$. By use of this strategy, we have a 2π coverage of the quantities $L_1(\beta, \gamma)$ and $L_2(\beta, \gamma)$ which can be used to reconstruct basis images $c_1(\vec{r})$ and $c_2(\vec{r})$, respectively.

IV. Simulation Study

Simulation studies were conducted to validate the strategy we proposed in this paper. Polychromatic projection data were computed by use of Eqn. (1) for the full-scan configuration where high- and low-energy were set to 80 kVp and 135 kVp, respectively. The short-scan projection data were synthesized by low energy data in the range $[0, \pi + \gamma_m)$ and high energy data in the range$[\pi + \gamma_m, 2\pi + 2\gamma_m]$ extracted from the full-scan dataset.

The reconstructed images from 80 and 135 kVp polychromatic full-scan projection data are shown in Fig. 2 in which the beam-hardening artifacts can be seen between the two bony structures. The images from left to right in Fig. 3 are the water- and bone-image respectively, reconstructed from the full-scan data with the data-domain decomposition method. Figure 4 shows the corresponding water- and bone-images reconstructed from the short-scan data with the data-domain decomposition strategy proposed in this paper. In Fig. 4, one can find that there are some noticeable discrepancies in the water- and bone-images of short-scan configuration compared with the full-scan results. Figure 5 shows the profile comparison between basis images reconstructed from decomposed full-scan and half-scan projection data. The profiles are plot through the central lines of the reconstructed water- and bone-images (Figs 3 and 4, respectively). One can also see in Fig. 5 that results from short-scan data are noisier than those from full-scan data. These discrepancies are due to the interpolation error in addition to decomposition error and short-scan reconstruction algorithm. Since the projection data in Eqns. (10) and (11) employed in the decomposition stage are generally not measured directly, we have to interpolate between two successive projection views which will introduce data inconsistency hence more severe decomposition error than the full-scan configuration. It is also well-known that there exit some discrepancies between full-scan and short-scan reconstruction results because of the fewer views used in the short-scan reconstruction algorithm. In the method proposed in this paper, one could decompose only the data between the angular range $[0, \pi + \gamma_m]$ and employ half-scan reconstruction algorithm to yield images of $c_1(\vec{r})$ and $c_2(\vec{r})$.

Figure 6 from left to right shows the synthesized monochromatic image at 70keV reconstructed from the decomposed full-scan data and half-scan data respectively. It can be observed that the beam-hardening artifacts are suppressed in the synthesized monochromatic image for both the full-scan and half-scan cases.

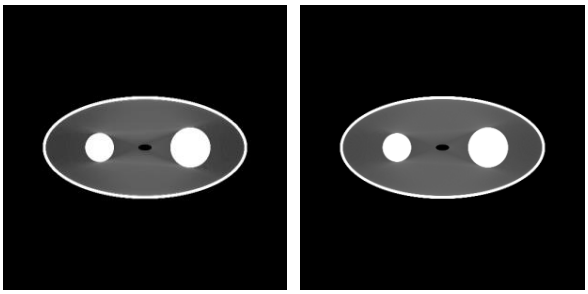


Fig.2. Images reconstructed from polychromatic data. Left: 80 kVp, right: 140 kVp.

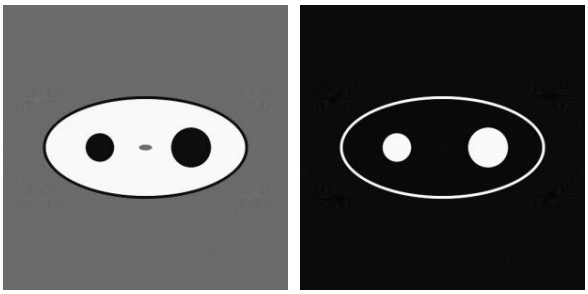


Fig.3. Basis images reconstructed from full-scan data. Left: water-image, right: bone-image, reconstructed from decomposed full-scan projection data.

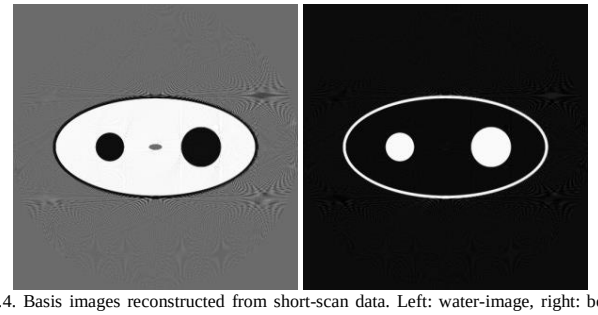


Fig.4. Basis images reconstructed from short-scan data. Left: water-image, right: bone-image, reconstructed from the decomposed short-scan projection data with the decomposition strategy proposed in this paper.

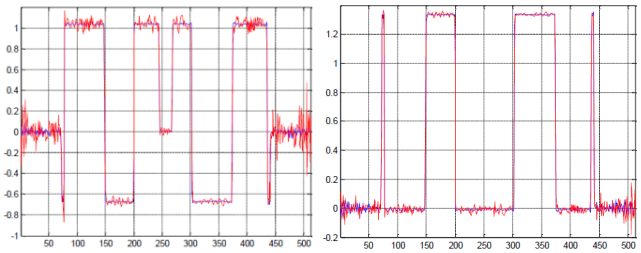


Fig.5. (left): Profiles through central line of full-scan (blue) water-image (Fig. 3 left) and short-scan (red) water-image (Fig. 4 left); (right): Profiles through central line of full-scan (blue) bone-image (Fig. 3 right) and short-scan (red) bone-image (Fig. 4 right).

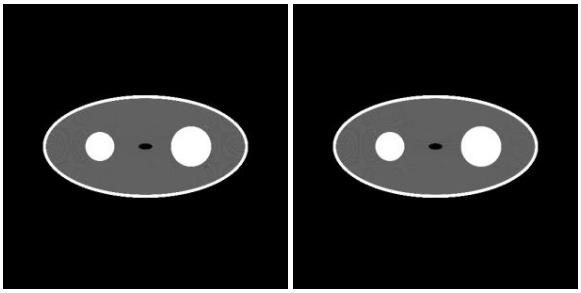


Fig.6. Synthesized monochromatic image from the basis images in Figs. 3 and 4 using Eqn. (2) at 70 keV. Left: full-scan, right: half-scan.

V. SUMMARY

In this work, we proposed a strategy to perform the decomposition in data domain for the 2D dual-energy CT short-scan configuration. The usual data redundancy condition in conventional 2D fan-bam CT reconstruction problem does not hold any more due to the poly-chromaticity of the x-ray source spectrum. However, complementary ray-sums can still be employed to develop data domain decomposition algorithm due to the fact that data redundancy condition still hold for the x-ray transform of the coefficients of water- and bone-images which are now the sought-after quantities in dual-energy CT decomposition problem.

REFERENCES

[1] R. Alvarez and A. Macovski, "Energy-selective Reconstructions in X-ray Computerized Tomography," *Phys. Med. Biol.* vol. 21, pp. 733-744, 1976.

[2] B. Chen *et. al.*, "Algorithm-enabled partial-angular-scan configurations for dual-energy CT," *Med. Phys.* vol. 45, pp. 1857-1870, 2018.

[3] Y.Zou and M. Silver, "Analysis of fast kV-switching in dual energy CTusing a prereconstruction decomposition technique," *SPIE,* vol. 6913, 691313, 2008.

[4] C. Maass *et. al.*, "Image-based dual energy CT using optimized pre correction functions," *Med. Phsy.* vol. 36, pp. 3818-3829, 2009.

[5] B. Heismann *et. al.*, "Density and atomic number measurements with spectral x-rayattenuation method," *J. Appl. Phys*., vol. 94, pp. 2073-2079, 2003.

[6] D. Parker, "Optimal short scan convolution reconstruction for fanbeam CT," Med. Phys., vol. 9, pp. 254-257, 1982.

A Pre-Filter that Incorporates the Noise Model

Gengsheng L. Zeng

***Abstract*—A linear denoising filter is usually of lowpass type, and the main parameter in a lowpass filter is the cutoff frequency. The lowpass filters are normally shift invariant and can be implemented as convolution in the spatial domain or as multiplication in the Fourier domain. This paper presents a linear filter that is not characterized by its cutoff frequency but is characterized by the noise model. An example of such a linear filter is presented for low-dose X-ray computed tomography (CT).**

***Index Terms*—Low-dose CT, denoising, linear filter, nonstationary filter.**

I. Introduction

Nowadays, the iterative image reconstruction algorithms have become more popular than the analytic filtered backprojection (FBP) algorithm [1]-[5]. The main motivation of using an iterative algorithm is that it is much easier to incorporate the noise model into the iterative algorithm than into an analytic algorithm such as FBP [6][7]. Many noise filtering techniques are available for the FBP algorithm [8]-[11]. In [9], a non-linear adaptive filter was proposed, in which a threshold value was assigned to determine whether the measurements were required to be filtered.

In [11] and [12], a linear FBP algorithm was proposed to mimic the iterative Landweber algorithm [13] and to incorporate the projection noise in a window function in the Fourier domain. The ray-by-ray noise-weighting scheme is spatially variant. Its implementation was to decompose the spatially variant filter into 10 spatially invariant filters. The 10 versions of the filtered measurements were then combined into one and used in the backprojector.

Instead of mimicking an iterative algorithm to filter the noise, this paper takes a different approach and develops a convolution lookalike linear filter for the projection measurements. The filter being designed is shift variant and is based on the noise model. We treat each measurement as a random variable. By noise model, we mean the relationship between the measurement value and its standard deviation. For emission tomography measurements, the mean value is approximated by its measurement. The Poisson noise model is that the mean value and the variance are the same. For the transmission tomographic measurements, the logarithm has been taken. The variance of the line-integral measurement is proportional to the exponential function of the line-integral measurement.

II. Methods

This section develops a nonstationary convolution lookalike filter for the line-integral projection measurement data. Let p be the unfiltered line-integral projection measurement and h be the spatial-domain filter kernel. In other words, p is the logarithm of the detected photon current from a CT scanner bin at an angle. If h is a shift-invariant convolution kernel, the conventional linear shift-invariant filtering procedure can be expressed as a convolution integral below

$$q(t,\theta) = \iint h(\hat{t}-t, \hat{\theta}-\theta)\, p(\hat{t},\hat{\theta})\, d\hat{t}d\hat{\theta} \qquad (1)$$

where $q(t,\theta)$ is the filtered projection. In (1), the kernel h is shift-invariant. In other words, the shift-invariant filter blurs the projections $p(t,\theta)$ with the same kernel h everywhere. If the kernel h varies from location to location, (1) can be modified to

$$q(t,\theta) = \iint h(\hat{t},\hat{\theta};t,\theta)\, p(\hat{t},\hat{\theta})\, d\hat{t}d\hat{\theta}. \qquad (2)$$

Equation (2) is no longer in the form of convolution. However, the computational complexity of (2) is almost the same as the complexity of (1), except that in (2) the kernel h must be evaluated differently for different locations (t,θ). In this paper, we assume the filter kernel h to be a 2D Gaussian function with a 'standard deviation' $\sigma_h(t,\theta)$. Our general strategy is to relate this $\sigma_h(t,\theta)$ with the noise standard deviation, $\sigma_p(t,\theta)$, of the noisy measurement $p(t,\theta)$.

For emission measurements, the projection $p(t,\theta)$ can be used to approximate the measurement noise variance according the Poisson distribution. Thus, $\sigma_p(t,\theta) \approx \sqrt{p(t,\theta)}$.

For Transmission measurements, the noise variance of the post-log measurement $p(t,\theta)$ can be approximated as the exponential function of the line-integral $p(t,\theta)$. Thus, $\sigma_p(t,\theta) \approx \sqrt{\exp(p(t,\theta))}$.

Our general strategy is to use a large kernel size of h if the corresponding $\sigma_p(t,\theta)$ value is large and to use a small kernel size if the corresponding $\sigma_p(t,\theta)$ value is small. The kernel size $\sigma_h(t,\theta)$ is thus a monotonic function of $\sigma_p(t,\theta)$. Here we propose this monotonic function to have the form of

$$\sigma_h = a \times \sigma_p^b + c. \qquad (3)$$

For transmission tomography, (3) becomes

$$\sigma_h = a \times (\sqrt{e^{p(t,\theta)}})^b + c. \qquad (4)$$

After combining the square-root and parameter b into a single parameter b, (4) becomes

$$\sigma_h = a \times e^{b \times p(t,\theta)} + c, \qquad (5)$$

with user defined parameters a, b, and c.

G. L. Zeng is with the Department of Engineering, Utah Valley University, Orem, UT 84058 USA and Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, UT 84108 USA (e-mail: larry.zeng@uvu.edu).

As a special case of $a = 0$, h has a constant σ_h, the filter is then shift-invariant, and the convolution lookalike expression (2) reduces to the true convolution (1).

III. Results

To illustrate the feasibility of the proposed linear nonstationary filter, a cadaver torso was scanned using an X-ray CT scanner with a low-dose setting. The projections were first filtered by the proposed filter using (2) and then the conventional FBP algorithm was used to reconstruct the final image.

The cadaver data was collected with a diagnostic scanner (Aquilion ONE™, Toshiba America Medical Systems, Tustin, CA, USA; raw data courtesy of Leiden University Medical Center). The imaging geometry was cone-beam, the X-ray source trajectory was a circle of radius 600 mm. The detector had 320 rows, the row-height was 0.5 mm, each row had 896 channels, and the fan angle was 49.2°. A low-dose noisy scan was carried out. The tube voltage was 120 kV and current was 60 mA. There were 1200 views uniformly sampled over 360°.

Figure 1 shows the FBP reconstruction from the low-dose CT data. It is observed that some horizontal streaking artifacts pass through both arms and the torso. The X-rays passing through both arms suffer from photon starvation and create large noise.

Figures 2 to 6 show the FBP reconstructions from the pre-filtered low-dose CT data. The pre-filters were nonstationary, the associated kernel h was a Gaussian function, whose 'standard deviation' value σ_h was defined in (5). They shared the same values of $a = 2.5$ and $c = -0.9999 \times 2.5$. The b value was different, with $b = 4$, 3, 2, 1, and 0.5, respectively. It seems that a large b value helps removing the streaking artifacts.

Figures 7 and 8 show the FBP reconstructions from the pre-filtered low-dose CT data. This time, the pre-filters were stationary, that is, the kernel h was shift-invariant by setting $a = 0$. The parameter c controls the degree of smoothing during filtering. In Figures 7 and 8, $c = 2$ and 4, respectively. For a stationary filter, the entire image is smoothed at the same amount. It is observed that the stationary filters are not effective in removing streaking artifacts. The image is already blurry, while the streaking artifacts are still there.

Figure 2. FBP reconstruction using a nonstationary Gaussian pre-filter h whose σ_h is defined in (5) with $a = 2.5$, $b = 4$, and $c = -0.9999 \times 2.5$.

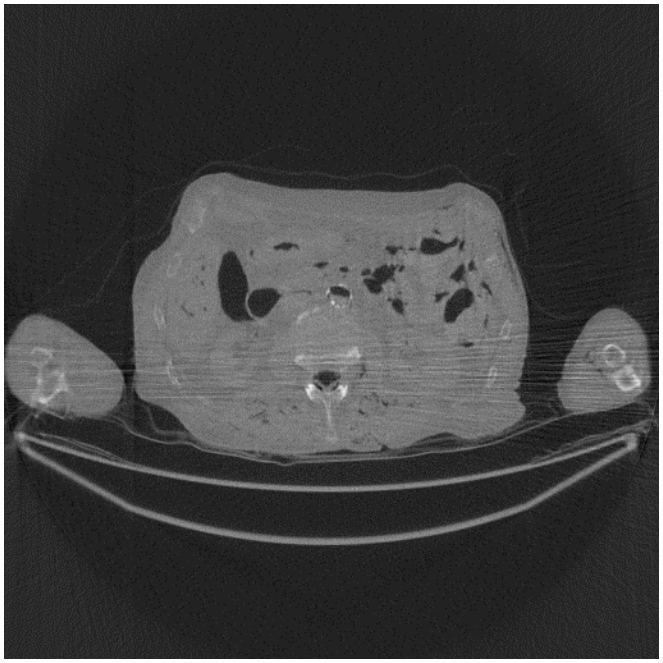


Figure 1. FBP reconstruction using the low-dose data without pre-filtering.

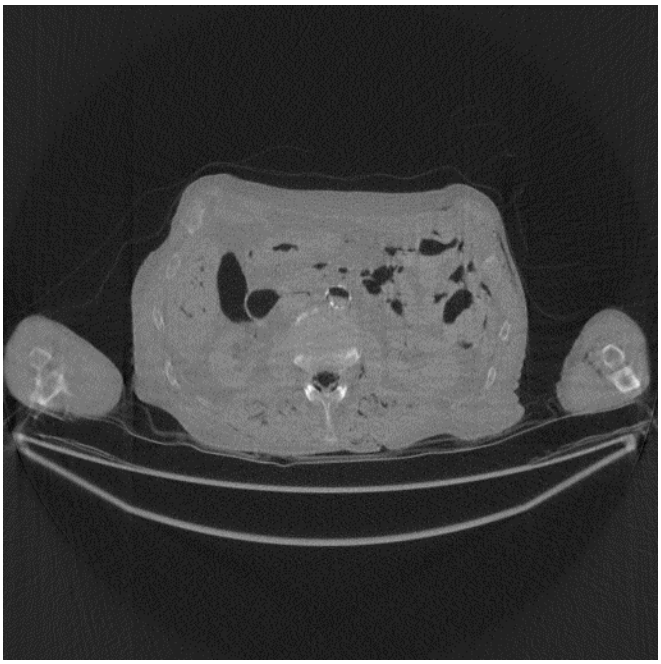


Figure 3. FBP reconstruction using a nonstationary Gaussian pre-filter h whose σ_h is defined in (5) with $a = 2.5$, $b = 3$, and $c = -0.9999 \times 2.5$.

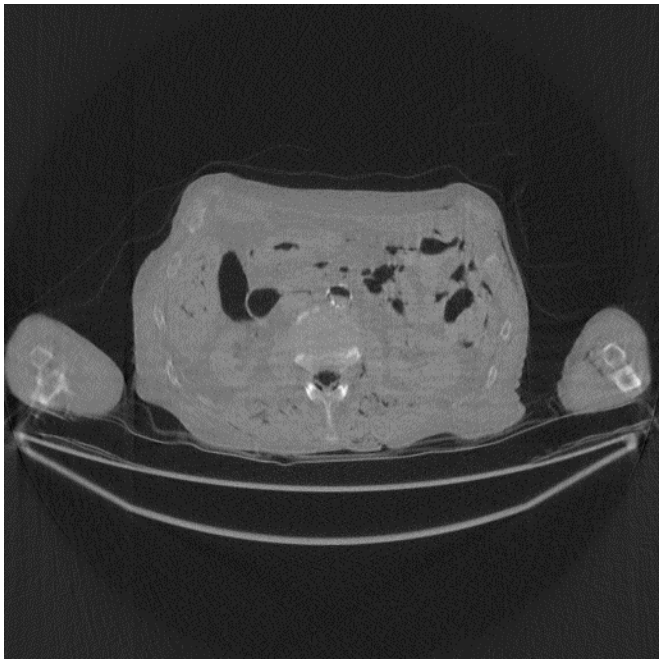


Figure 4. FBP reconstruction using a nonstationary Gaussian pre-filter h whose σ_h is defined in (5) with a = 2.5, b = 2, and c = -0.9999 × 2.5.

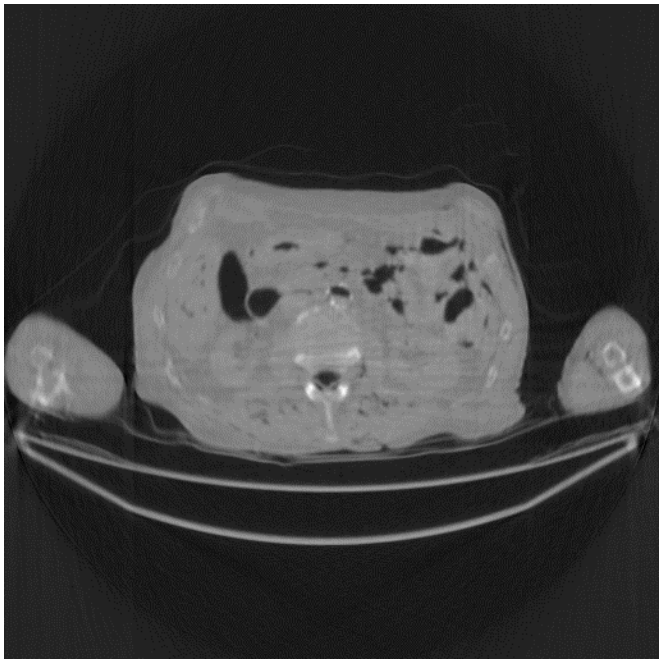


Figure 6. FBP reconstruction using a nonstationary Gaussian pre-filter h whose σ_h is defined in (5) with a = 2.5, b = 0.5, and c = -0.9999 × 2.5.

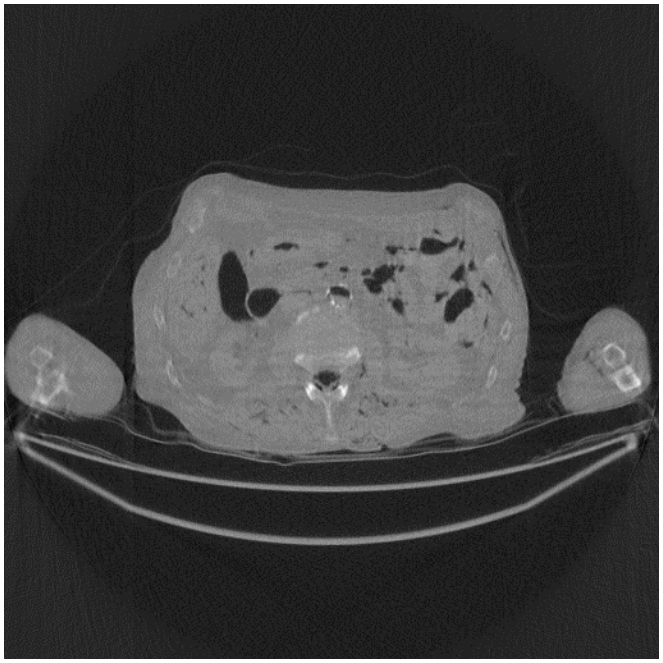


Figure 5. FBP reconstruction using a nonstationary Gaussian pre-filter h whose σ_h is defined in (5) with a = 2.5, b = 1, and c = -0.9999 × 2.5.

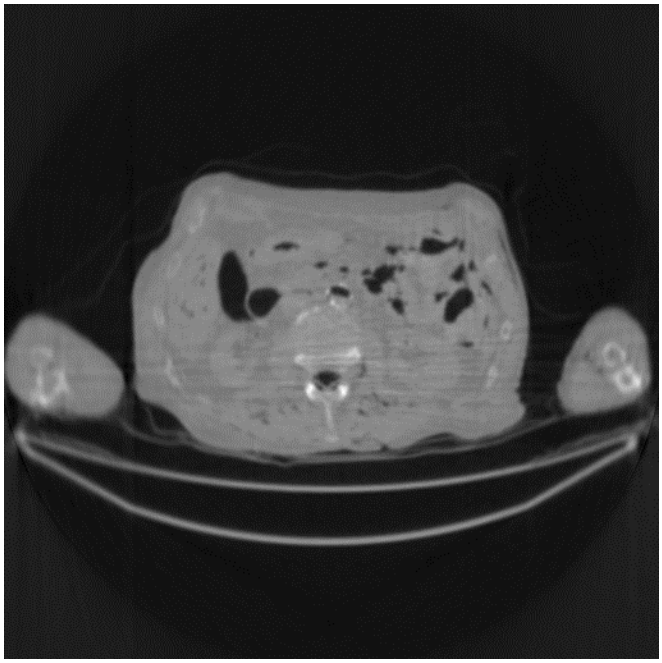


Figure 7. FBP reconstruction using a stationary Gaussian pre-filter h whose σ_h is defined in (5) with a = 0 and c = 2.

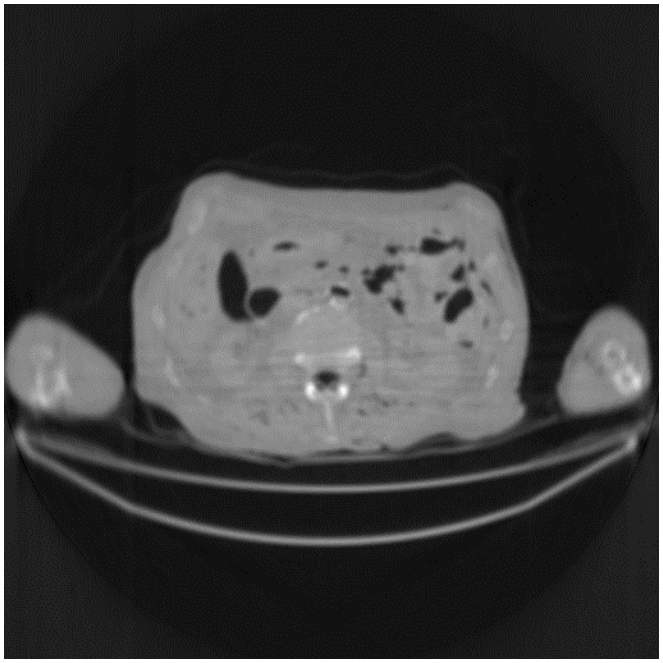


Figure 8. FBP reconstruction using a stationary Gaussian pre-filter h whose σ_h is defined in (5) with $a = 0$ and $c = 4$.

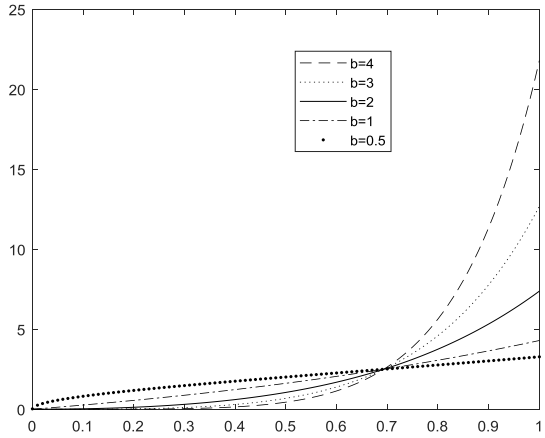


Figure 9. The functional relationship (5) of 5 cases as in Figures 2 to 6, respectively. The horizontal axis represents the normalized projection value, and the vertical axis represents the size σ_h of the kernel h.

Figure 9 shows 5 curves according to the relationship (5). The 5 curves correspond to the 5 cases with different b values as shown in Figures 2 to 6, respectively. The horizontal axis represents the normalized projection value, and the vertical axis represents the size σ_h of the kernel h.

When the parameter b is small such as 0.5, the kernel size σ_h increases gently with the projection $p(t,\theta)$. When the parameter b is large such as 4, the kernel size σ_h increases dramatically with the projection $p(t,\theta)$. Intuitively speaking, the streaking artifacts are caused by few projections with the largest values, whose rays pass through both arms. Using a large parameter b (e.g., $b = 4$) effectively separates the projections with large values from other projections. Only those large projection values are smoothed, while other smaller projection values are almost unchanged. The overall image resolution is unchanged.

IV. Conclusion

The proposed pre-filter is designed for the FBP algorithm, aiming to reduce the photon starvation streaking artifacts. This is a linear filter as suggested by (2). It has a convolution lookalike form, with similar computational complexity as convolution.

The most important property of the proposed filter is its nonstationary characteristic, which allows the filter to smooth the data with specified range of values. As a result, some streaking artifacts can be reduced. Stationary filters do not have this ability.

A low-dose CT study is used to test the feasibility of the proposed pre-filter. The relationship between the kernel size and the noise variance is not uniquely determined. The model (5) is only one functional form that can be utilized. The user is encouraged to adopt other forms to label the undesired projection values to be smoothed out.

Acknowledgment

The author thanks Raoul M.S. Joemai of Leiden University Medical Center for collecting and providing us raw data of the cadaver CT scan.

References

[1] S. German and D. E. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Internat. Statist. Inst*., vol. LII-4, pp. 5–21, 1987.

[2] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser. B (Methodol)*, vol. 39B, pp. 1–38, 1977.

[3] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag*., vol. 1, pp. 113–122 1982.

[4] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag*., vol. 13, pp. 601–609, 1994.

[5] K. Langer and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," J. Comput. Assist. Tomogr. **8**, 302–316 (1984).

[6] G. L. Zeng, "Emission expectation-maximization look-alike algorithms for x-ray CT and other applications," *Medical Physics*, vol. 45, pp. 3721-3727, 2018.

[7] G. L. Zeng and Y. Li, "Extension of emission expectation maximization lookalike algorithms to Bayesian algorithms," *Visual Computing for Industry, Biomedicine, and Art*, vol. 2, article 14, 2019, open access, https://doi.org/10.1186/s42492-019-0027-4

[8] J. Hsieh, "Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise," *Med. Phys*. vol. 25, pp. 2139–2147, 1998.

[9] M. Kachelrieß, "Branchless vectorized median filtering," *IEEE Medical Imaging Conference Record, Workshop on High Performance Medical Imaging* (IEEE, Orlando, Florida, 2009), vol. HP3-5, pp. 4099–4105, 2009.

[10] M. Kachelrieß, O. Watzke, and W. A. Kalender, "Generalized multidimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT," Med. Phys. vol. 28, pp. 475–490 2001.

[11] G. L. Zeng and A. Zamyatin, "A filtered backprojection algorithm with ray-by-ray noise weighting," *Med. Phys.*, vol. 40, 031113; http://dx.doi.org/10.1118/1.4790696 (7 pages), 2013.

[12] G. L. Zeng, *Image Reconstruction: Applications in Medical Sciences*, ISBN 978-3-110-50059-2, Walter de Gruyter GmbH, Berlin/Boston, 2017.

[13] L. Landweber, "An iteration formula for Fredholm integral equations of the first kind," *Amer. J. Math*., vol. 73, pp. 615–624, 1951.

Cone Beam Artifacts Correction in Multi-detector Computed Tomography Using Deep Neural Networks

Hao Zhao, Yanyan Liu, and Guotao Quan*,
United Imaging Healthcare, 2258 Chengbei Rd, Shanghai, China, 201800

***Abstract*— The cone beam artifacts are commonly seen in multi-detector computed tomography with large coverage, especially in axial geometry where the acquired data are insufficient for analytic reconstruction outside the acquisition plane. The artifacts show up as shadows and bright regions from the axial view. Subsequent streak artifacts can be observed from sagittal and coronal views. Cone beam artifacts can be reduced by various modified Feldkamp/Davis/Kress (FDK)-like reconstruction methods. Additionally, they can be suppressed by the use of iterative reconstruction. These methods cannot completely reduce the cone-beam artifact or have unpractically long time consumption. In this paper, we proposed a deep neural network to construct a novel correction, which can efficiently eliminate the cone beam artifacts in head images. We construct the target images by eliminating the low frequency artifacts in original FDK images using model based iterative reconstructed images. To take advantage of the cone angle information, we incorporated a 'cone angle map' as an additional feature, expecting the network to be able to recognize the severity of cone beam artifacts in a certain image. The trained neural network outperformed FDK reconstruction within a relatively low computational cost.**

***Index Terms*—computed tomography, cone beam artifacts, deep learning.**

I. Introduction

With an increased volume coverage, multi-detector computed tomography (CT) is able to scan a large volume rapidly [2]. As the coverage goes wider, the cone beam artifact, which is caused by insufficient data at cone angle plane [7], becomes worse. The cone beam artifact causes severe image distortion at the boundary of the scan volume. Much effort has been made to reduce cone angle artifacts using various modified Feldkamp/Davis/Kress (FDK) reconstruction methods and iterative reconstruction [1][5][6][8][9][10]. The analytical methods can only partially reduce the artifact, while the converged iterative methods are rarely applied in clinical setting for their relatively long computation time.

Deep neural networks (DNN) developed rapidly in recent years, and declared great success in many areas including image processing [3]. In this paper, we proposed a modified U-net model to construct a novel network that can recognize and eliminate cone beam artifacts in head images. The application of this network can eliminate cone beam artifacts quickly and effectively, without introduction of additional artifacts.

II. Data and Methods

A. Data

Clinical data of 33 patients were collected from a new 16cm coverage CT scanner made by United Imaging Healthcare (Shanghai, China) during the clinical validation study. The wide detector coverage guarantees the ability to scan the whole head within a single axial shot. The images were reconstructed with thickness of 0.5mm and increment of 0.5mm.

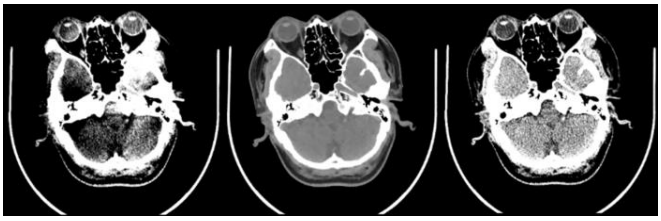


Fig. 1. The construction of the training dataset. Left: FDK image served as the input of the network; Middle: iterative reconstruction image; Right: image constructed from FDK and iterative reconstruction images, served as the learning target.

The original FDK images, which were distorted by cone beam artifacts, were served as the inputs of the network. Model based iterative reconstruction (MBIR) provided artifact free images, albeit with the long processing time. We extracted the 'cone beam artifacts images' by directly subtracting the MBIR reconstruction images from the corresponding FDK images. A Gaussian low pass filter was applied to the 'cone beam artifacts images' to filter out the noise. We believe that the remaining low frequency information in artifacts images is attributed to be the cone beam artifacts. The blurred negative of the cone artifacts was added to the original FDK images to get a final artifact free images, which were served as the target images. As the MBIR images always show up with a modified noise pattern compared with FDK images, the low pass filtering kept the artifacts only, without modifying the FDK noise pattern. We depicted the appearance of the training datasets in Fig.1. The

Hao. Zhao, Yanyan. Liu and Guotao. Quan are with the Department of Computed Tomography, United Imaging Healthcare, 2258 Chengbei Rd, Shanghai, China.

*Corresponding author, guotao.quan@united-imaing.com

FDK image showed severe cone beam artifacts. The MBIR image showed artifact free head structure, but with an over smoothed noise pattern. The processed learning target showed no cone beam artifacts, while maintaining the similar noise pattern to the FDK image.

B. Method

The proposed deep neural network in this work is a modified U-net model [4]. The network structure is shown in Fig.2. Each rectangle corresponds to a multi-channel feature map. The blank rectangles represent the learned feature map while the grey rectangles represent the copied feature map. Basically, U-net encodes the image to deep features then decodes it to its original size. The encoding part consists of four-part down-sampling process, with two convolutional layers containing 3x3 kernel between each of them. The down sampling process is executed using a 2x2 maximum pooling. The decoding part using up sampling and convolutional layers to produce a whole resolution image. Cropped image from the encoding part is added on the decoding matrix directly.

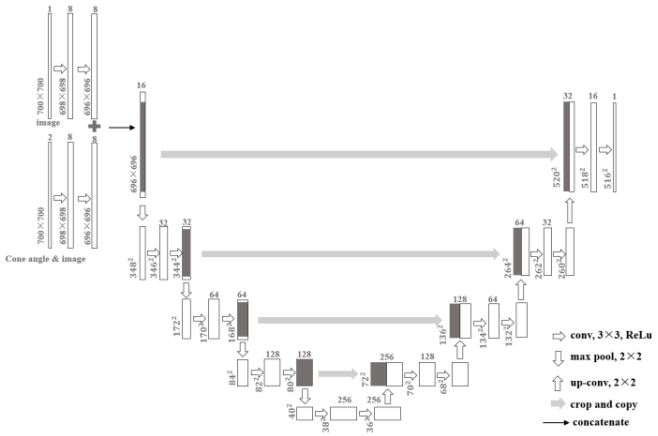


Fig. 2. The modified U-net structure. A 'cone angle map' feature is added in another branch at the beginning of the net to make it recognize the severity of the cone angle effect.

To make the network recognize the severity of the cone angle effect, we added an additional 'cone angle map' as another input of the network in addition to the original image. The 'cone angle map' has the same size to the input map, but with a constant value related to corresponding cone angle. The 'cone angle map' together with the original greyscale image passed through two convolutional layers, each of them followed by a batch normalization layer and a pooling layer. The image was then linked with another branch where the original input image also passed through two similar convolutional layers. The concatenated image features are fed into a standard U-net structure.

We used Mean Square Error (MSE) as the loss function,

$$MSE\ Loss(x_i, y_i) = (x_i - y_i)^2 \qquad (1)$$

Where x_i and y_i stand for input and output image of the network respectively. To reduce the instability of the Hounsfield Unit (HU) value in the output image, we added another loss function to calculate the error of mean HU value between network output and target image in soft tissue regions,

$$HU\ Loss(\mathrm{x}, \mathrm{y}) = |mean(x_i) - mean(y_i)|^2 \qquad (2)$$

Where i is the pixel index of soft tissue. Total loss is the addition of MSE loss and HU loss multiplying a scale factor,

$$Total\ loss = MSE\ Loss + \gamma * HU\ Loss \qquad (3)$$

Where γ a balance term set to 10 according to the values of two loss term during training. The optimizer is chosen as the adaptive moment estimation optimizer (Adam). We trained the network on an NVidia Tesla V100. The total loss converged after more than 400 epochs. Total training time is about 4 days.

III. RESULT

Additional images of 10 patients were chosen to test the performance of the trained network. We randomly picked one set of testing images to show the artifact reduction. The comparison was depicted in Fig.3 and Fig.4, from both axial and sagittal view. The improvement of the image quality is remarkable at first sight as the low frequency dark and bright artifacts were greatly reduced. The target images in the right panel are used for comparison only. In Fig.3 we showed three images with different image location from the edge to the middle plane. All images processed by the network showed great artifact reduction considering different cone angle. As shown in the bottom panel in Fig.3, for the acquisition plane of the scan volume, the network did not add any artifacts. This is expected, since this image originally has good quality without cone beam artifacts. That illustrates the validity of the additional cone angle map. MBIR images can sometimes have cone beam artifacts as well. With careful selection of training datasets, the network can possibly get a better result than the MBIR images when the artifacts are present in the test data, as indicated in our illustration.

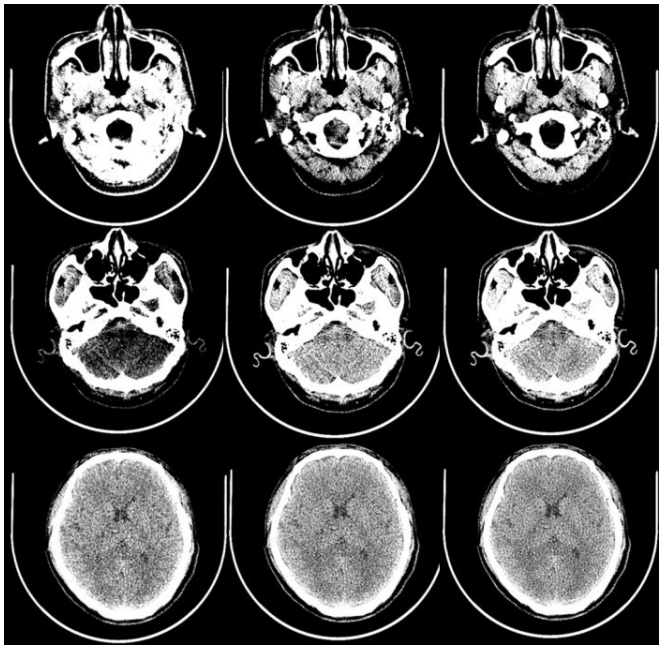


Fig. 3. The performance on clinical FDK images. These three rows represent images of different slice location, from the edge to the middle plane. The left, middle and right columns are original FDK images, output images of the trained network and the target images constructed from FDK and MBIR images. All images are displayed with a window level of 35 HU and a window width of 80 HU.

The sagittal view in Fig. 4 exhibits that the discontinuity along the row direction almost disappeared. The network did not disturb any existing structures in FDK images in our evaluation. It proved that the network can process the artifacts according to the severity of the cone beam effect, which we believed benefited from the addition of cone angle information.

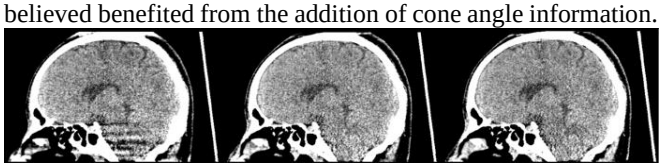


Fig. 4. The performance on clinical FDK images in sagittal view. Left: original FDK image; Middle: output image of the network; Right: target image.

We tested the computational cost of this network on NVidia Tesla V100. The total time consumption of processing 320 images of one axial scan is less than 3 seconds.

IV. DISCUSSION

The application of deep neural network eliminated cone beam artifacts efficiently and effectively in head images. The quality of artifact free images is crucial to network training. The training set should also be sufficient in number, while containing various artifact forms that might appear in different structures.

It is foreseeable that the cone beam artifacts on images of other body parts, like cardiac or abdomen, can also be reduced by deep learning method, which is a topic of our future research.

REFERENCES

[1] Feldkamp, L. A., Davis, L. C., & Kress, J. W. (1984). Practical cone-beam algorithm. Josa a, 1(6), 612-619.
[2] Hsieh, J. (2009, November). Computed tomography: principles, design, artifacts, and recent advances. Bellingham, WA: SPIE.
[3] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436.
[4] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
[5] Tang X, Hsieh J, Hagiwara A, Nilsen R A, Thibault J-B and Drapkin E 2005 A 3D weighted cone beam filtered back projection (CB-FBP) algorithm for image reconstruction in volumetric CT under a circular source trajectory Phys. Med. Biol. 50 3889–905
[6] Thibault J-B, Sauer K D, Bouman C A and Hsieh J 2007 A 3D statistical approach to improved image quality for multislice helical CT Med. Phys. 34 4526
[7] Tuy, H. K. (1983). An inversion formula for cone-beam reconstruction. SIAM Journal on Applied Mathematics, 43(3), 546-552.
[8] Stanislav Žabić, Zhicong Yu, Wenjing Cao, Lu Wang, "Low frequency recovery in 16cm coverage axial multi-detector computed tomography," Proc. SPIE 11072, 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 110720V (28 May 2019);
[9] Zeng K, Chen Z, Zhang L and Wang G 2004 An error-reduction-based algorithm for cone-beam computed tomography Med. Phys. 31 3206
[10] Mori, S., Endo, M., Komatsu, S., Kandatsu, S., Yashiro, T., & Baba, M. (2006). A combination-weighted Feldkamp-based reconstruction algorithm for cone-beam CT. Physics in Medicine & Biology, 51(16), 3953.

Cardiac Motion Correction of Computed Tomography (CT) with Spatial Transformer Network

Guotao Quan, Jiao Tian, Yi Wang

***Abstract*—Even with the latest computed tomography (CT) scanner, cardiac motion artifact is still a major factor of misinterpretations of coronary artery disease. In order to reduce the motion artifact of cardiac more efficiently, deep learning with convolutional neural networks (CNN) is introduced into cardiac motion correction algorithm. In essence, motion correction is a spatial transformation of images of coronary artery. However, spatial variability in images of objects is still a major challenge of CNN. Therefore, how to achieve motion correction of cardiac completely with CNN is still an open issue. In this paper, a spatial transformer network (STN) is embedded into U-net to fulfill an end to end motion correction of cardiac without any traditional algorithm. The STN will address the CNN's deficiency in spatial transformation. With four sequential phases of cardiac images as input, the deep learning network can directly generate the cardiac images with rare motion artifacts. Real case and simulated case with fixed motion trajectory are used to prove that this method can reduce severe motion artifacts and achieve effective temporal resolution of 12.5ms**

***Index Terms*— Cardia Motion Correction, Convolutional Neural Networks, Deep Learning, Spatial Transformer Network**

I. Introduction

REDUCING cardiac motion artifacts and improving the temporal resolution of coronal artery angiography of Computed Tomography(CT) is still a hot topic in CT reconstruction[1]. Because even with the latest CT scanner with 0.2-0.3s rotation speed, the motion artifact of coronal artery is still a major impact of misinterpretations of coronary artery disease[2].One solution to reduce the cardiac motion artifacts is using iterative reconstruction with motion estimation and compensation integrated into the penalty of iterative algorithm[3], but the speed of this method limits its application in routine clinic practice. The other solution is estimating the motion vector from sequential phase of cardiac[4] or reconstructed images with partial angle[5], and compensating the motion with motion vector in image or projection data[6]. As the motion of cardiac contains not only the movement of coronal artery but also atria and ventricle, segmentation and extraction of center line of coronary artery are needed to focus the motion estimation and compensation on the coronary artery. These methods are popular and effective, but they need a processing chain with lots of image processing algorithms, error of any algorithm in this chain will be accumulated and effect the final result, also it need a huge computation time for all of these image processing algorithms, so it will not practical for routing clinic application. Benefit from the performance and accuracy, deep learning with convolution neural networks (CNN) outperforms many traditional algorithms in image processing [7]. Many steps in motion correction algorithm, such as coronary segmentation or extraction of the center line et al., are replaced by deep learning with CNN. Other key steps like image registration[8], motion estimation and motion compensation are still using traditional algorithms, since the CNN lack the capability to implement spatial transformation efficiently. So, in this paper, a spatial transformer network (STN)[9] is embedded into U-net[10] to fulfill an end to end deep learning based motion correction(deepMC). The structure of network will be introduced first (Sec. II.A), then data preparing and training parameters will be shown (Sec. II.B), after that, how to simulate with real cardiac images and quantify the effective temporal resolution will be described (Sec. II.C), real cases will show the performance of this method, and simulated data will quantify the effective temporal resolution (Sec. III).

.

II. Materials and Methods

The structure of U-net with spatial transformer network (STN) embedded is introduced first, then training and testing dataset preparation will be described. At last how to simulate cardiac movement with real cardiac images and quantify the effective temporal resolution will be shown.

A. Structure of Deep Learning Network

The structure of U-net with STN embedded is shown in Fig. 1. The network contains encoder and decoder part. The output of each layer of encoder is feed as the input of each layer in decoder by skip connection. STN is embedded in the end of each layer of encoder. Inputs of this network are 4 cardiac images with sequential phase time, each image is random

Guotao Quan is with the Shanghai United Imaging Healthcare Co., Ltd, 2258 Chengbei Rd., Jiading District, Shanghai China (e-mail: guotao.quan@united-imaging.com)

Jiao Tian and Yi Wang is with the Shanghai United Imaging Healthcare Co., Ltd, 2258 Chengbei Rd., Jiading District, Shanghai China

divided into 256x256 patch. In encoder part, the size of image will decrease 2 times in each layer because of the max pool layer, and the channel number will increase from 4 to 512, but in decoder part, the size of image will increase 2 times in each layer because of the up-convolution layer, and the channel number will decrease from 512 to 1.

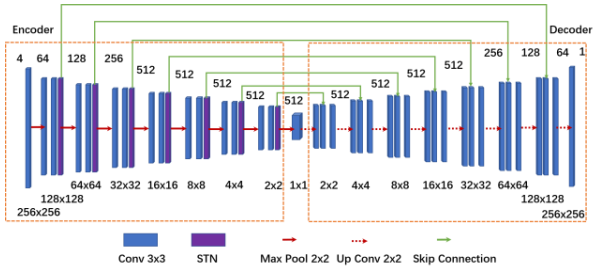


Fig. 1 Structure of STN embedded U-net

The structure of STN is shown in Fig.2. it contains localization network which analyze the input's feature and output the coefficients θ. For rigid transformer, θ is a 3x2 matrix. For detail please see this reference[9].

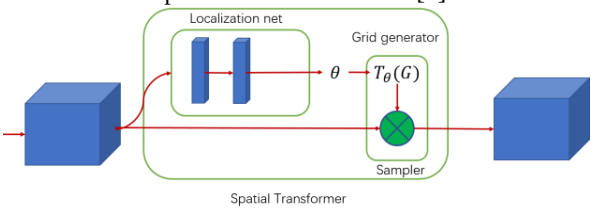


Fig. 2 Structure of spatial transformer[9]

B. *Dataset preparation*

The training dataset contains 50 cases. Each case contains 5 adjacent phases (with an increment of 5% of the R-R interval) as the input, each phase contains 280 slices of cardiac images (scanning with uCT960 of United Imaging, a 320-detector row CT) with the pixel size of 512x512. For each case, one of the phases (or the middle phase) is processed by the traditional motion correction algorithm and is taken as the target to learn.

The loss function is mean square (MSE), and optimization function is Adam with 0.1 decay.

C. *Effective temporal resolution quantification*

According to the reference[11, 12], the quantification of effective temporal resolution can be analysis, however, they only study the effective temporal resolution on simple and ideal cases like movement of a small ball[13]. In order to simulate more realistic movement, images from uCT960 of United-imaging were adopted. Radiologist identified one case I_g with 60 heart beats per minute as the golden standard, which exhibited no coronary motion artifact. With this case, the coronary was segmented, then, the center line C_g of the coronary was extracted. A set of center line $C_{t_0}, C_{t_1} \dots C_{t_N}$ were simulated under a linear motion model from time t_0 to t_N. The simulated coronary moved with constant velocity(v) along different direction and had a rest phase of variable duration T_{rest}, the longer of T_{rest} the less motion artifact would be presented. The speed of 56 mm/s was chosen according to these reference[11, 12]. Motion vector field (MVF) around centerline was calculated by point registration to approximate the motion between the reference time t_g and the time t_n. By using this MVF, the I_g could be warped into any I_{t_n}. 3D forward projection was applied into each I_{t_n} to generate the sinogram of CT, the forward projection parameters were the same as uCT960 of United-imaging (0.25s/rotation, SID 600mm, SDD 1118mm, image thickness 0.5mm, detector row number 280, thickness of detector in row direction 0.5mm, channel number of detector 936), at last a 3D filter back-projection(FBP) was applied on the sinogram to get the images $I_{T_{rest}}$ with motion artifacts.

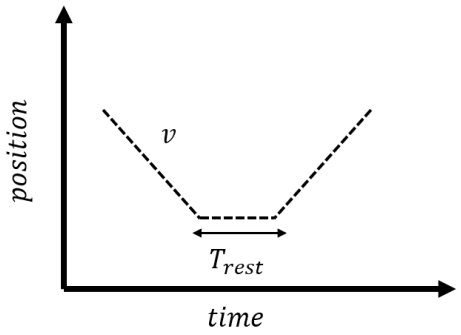


Fig. 3 Simulated cardiac movement curve[12]

In order to quantify the effective temporal resolution, an accuracy ratio was proposed in this reference[11], but this accuracy ratio proposed in this paper could cover the simple and ideal case of motion for small ball, in order to make this accuracy ratio suitable for complex and realistic motion case. A new accuracy ratio is proposed.

$$\alpha(T_{rest}, p, d, s) = \frac{\sum_i I_{T_{rest}}(i,p,d,s)\cdot C_{mask}(i,p,d,s)\cdot I_g(i,p,d,s)\cdot C_{mask}(i,p,d,s)}{\sum_i I_g(i,p,d,s)\cdot C_{mask}(i,p,d,s)\cdot I_g(i,p,d,s)\cdot C_{mask}(i,p,d,s)} \quad (1)$$

Where i denotes the sum over all pixel in image, p is the simulated phase index, d is the simulated direction index, s is the slice number of images in volume. C_{mask} is the mask of coronary, this accuracy is the ratio of the shift-invariant cross-correlation between I_g and $I_{T_{rest}}$.

Four kinds of T_{rest}, five sequential phases and five kinds of movement direction were simulated, where $T_{rest} = [125.00, 62.50, 31.25, 12.50]$ ms, Movement direction $D = [0°, 45°, 90°, 135°, 180°]$. Five sequential phases phase1-5.

The effective temporal resolution could be expressed as:

$$T = argmin_{T_{rest}}(\alpha(T_{rest}) - \alpha(T_{rest} = 125ms))^2 \quad (2)$$

This threshold in Eq.2 was not 0.98 as describe in reference[11], because the rotation speed of uCT960 of United-imaging was 0.25s/rotation, so the temporal resolution of this system was 125ms, if $\alpha(T_{rest})$ with deepMC was close to $\alpha(T_{rest} = 125ms)$, that mean the effective temporal resolution was T_{rest}.

III. RESULT

A. *Image Quality evaluation*

One simulated case with 90mm/s cardiac movement, $T_{rest} = 25ms$ was shown below, the reconstruction parameters are: 512*512 matrix size, 280 slices, 200mm FoV, 4 phases images

of 40%, 45% ,55% ,60% were reconstructed as the input of deep learning network.

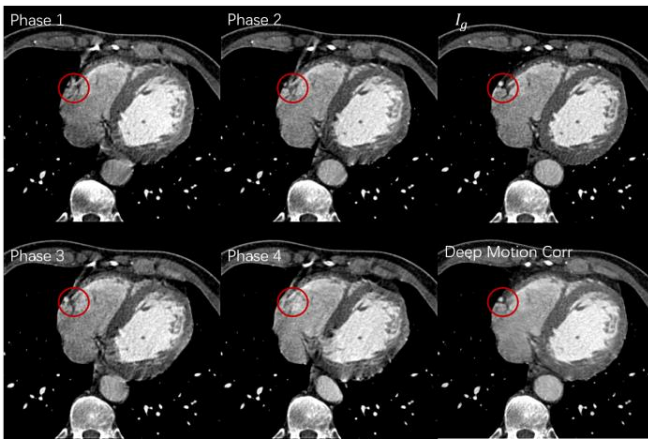


Fig. 4 Simulated case, the input of deep learning network is the images from phase 1, 2, 3 and 4, deep motion corr means the output of deep learning network, I_g is the gold standard of simulation.

One real case with 90 heart beats per second was used to demonstrate the capability of this method, the reconstruction parameters are: 512*512 matrix size, 280 slices, 185mm FoV, 4 phases images of 40%,45%,55%,60% were reconstructed as the input of deep learning network.

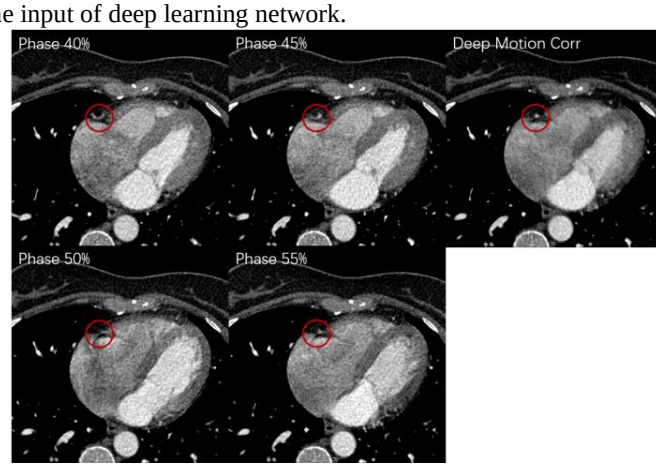


Fig. 5 Real case, the input of deep learning network is the images of 40%, 45% ,55% and 60%, deep motion corr means the output of deep learning network

In Fig. 4 and 5, the cardiac motion artifact was severe, some coronaries can't even be seen clearly in some images, but after deepMC, the coronary could be found without motion artifacts.

B. Effective temporal resolution

Four different of T_{rest} , each with five sequential phases and five kinds of movement direction were simulated. Average value is calculated to evaluate the effective temporal resolution. First, the average was done in movement direction and slice number, as shown in Eq. 3

$$\alpha(T_{rest}, p) = \sum_{d,s} \alpha(T_{rest}, p, d, s) \quad (3)$$

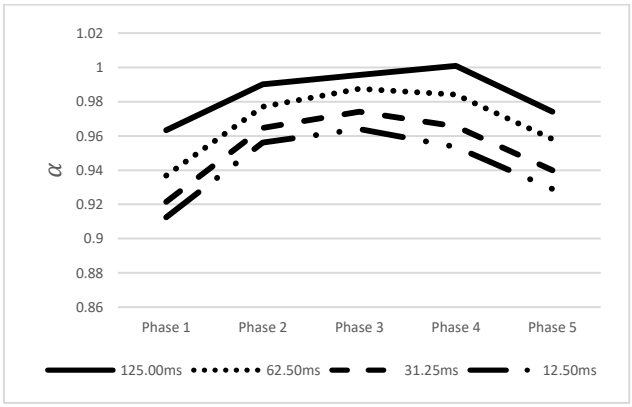


Fig. 6 accuracy ratio of FBP

The accuracy ratio is decreased with the reduction of the T_{rest} , which proves as the relationship between T_{rest} , and temporal resolution. This trend conformed to the real situation that with the decrease of T_{rest} the temporal resolution reduced.

The deepMC was applied into the simulated images with 62.5ms, 31.25ms and 12.5ms to find the effective temporal resolution for this method. The accuracy ratios of FBP and deep MC were calculated according to Eq. 3, but only phase 3 of FBP was selected to be compared, because the time of phase of deepMC's was the same as phase 3.

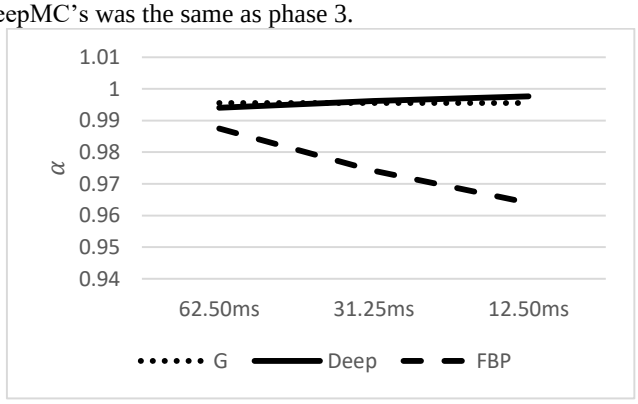


Fig. 7 Accuracy ratio between FBP, deepMC (Deep) and FBP with $T_{rest} = 125ms$ denoted by G.

	62.50ms	31.25ms	12.50ms
T test result between Deep and G	0	0	0
T test result between FBP and G	1	1	1

Tab. 1 T test result of value in Fig. 7 at 5% significance level

In Fig. 7 the accuracy ratio decreased as T_{rest} reduction, but with deepMC, the accuracy ratio didn't change a lot. T test between α value of deepMC, FBP with $T_{rest} = 125ms$, and FBP with other T_{rest} was applied to prove that there was no significant difference between α value of deepMC and FBP with $T_{rest} = 125ms$, but there was significant difference between FBP with other T_{rest} and FBP with $T_{rest} = 125ms$. So, the effective temporal resolution of deepMC was 12.5ms.

IV. Conclusion

In order to realize end to end motion correction with deep learning of CNN，U-net with STN embedded is proposed in this paper. A cardiac movement simulation with real cardiac images is presented to quantify the temporal resolution of this method. The result show 12.5ms temporal resolution can be achieved. At last, the real case of cardiac with fast heart beats prove the validation of this method.

References

[1] K. Lackner, and P. J. R. Thurn, “Computed tomography of the heart: ECG-gated and continuous scans,” vol. 140, no. 2, pp. 413-420, 1981.

[2] D. P. Boyd, and M. J. J. P. o. t. I. Lipton, “Cardiac computed tomography,” vol. 71, no. 3, pp. 298-307, 1983.

[3] M. J. C. C. I. R. Kachelrieß, “Iterative reconstruction techniques: What do they mean for cardiac CT?,” vol. 6, no. 3, pp. 268-281, 2013.

[4] L. Fan, J. Zhang, D. Xu, Z. Dong, X. Li, and L. J. J. o. c. a. t. Zhang, “CTCA image quality improvement by using snapshot freeze technique under prospective and retrospective electrocardiographic gating,” vol. 39, no. 2, pp. 202-206, 2015.

[5] J. Hahn, H. Bruder, T. Allmendinger, K. Stierstorfer, T. Flohr, and M. Kachelriess, "Reduction of motion artifacts in cardiac CT based on partial angle reconstructions from short scan data." p. 97831A.

[6] S. Kim, Y. Chang, and J. B. J. M. p. Ra, “Cardiac motion correction based on partial angle reconstructed images in x-ray CT,” vol. 42, no. 5, pp. 2560-2571, 2015.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks." pp. 1097-1105.

[8] R. Bhagalia, J. D. Pack, J. V. Miller, and M. J. M. p. Iatrou, “Nonrigid registration-based coronary artery motion correction for cardiac computed tomography,” vol. 39, no. 7Part1, pp. 4245-4254, 2012.

[9] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks." pp. 2017-2025.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation." pp. 234-241.

[11] C. Maaß, and M. Kachelrieß, "Quantification of temporal resolution and its reliability in the context of TRI-PICCS and dual source CT." p. 79611M.

[12] B. E. Nett, J. D. Pack, and D. Okerlund, "Task based assessment of a motion compensation algorithm via simulation of a moving stenotic vessel." p. 86682B.

[13] T. Lossau, H. Nickisch, T. Wissel, R. Bippus, H. Schmitt, M. Morlock, and M. J. M. i. a. Grass, “Motion artifact recognition and quantification in coronary CT angiography using convolutional neural networks,” vol. 52, pp. 68-79, 2019.

Rapid and flexible high-resolution scanning enabled by cycloidal computed tomography and convolutional neural network (CNN) based data recovery

Daniël M. Pelt, Charlotte Maughan-Jones, Oriol Roche i Morgó, Alessandro Olivo, Charlotte Hagen

***Abstract*—We have combined a recently developed imaging concept ("cycloidal computed tomography") with convolutional neural network (CNN) based data recovery. The imaging concept is enabled by exploiting, in synergy, the benefits of probing the sample with a structured x-ray beam and applying a cycloidal acquisition scheme by which the sample is simultaneously rotated and laterally translated. The beam structuring provides a means of increasing the in-slice spatial resolution in tomographic images irrespective of the blur imposed by the x-ray source and detector, while the "roto-translation" sampling allows for rapid scanning. Data recovery based on the recently proposed Mixed-Scale Dense (MSD) CNN architecture enables an efficient reconstruction of high-quality, high-resolution images despite the fact that cycloidal computed tomography data are highly incomplete. In the following, we review the basic principles underpinning cycloidal computed tomography, introduce the CNN based data recovery method and discuss the benefit of combining both.**

***Index Terms*—computed tomography, micro-CT, convolutional neural networks, machine learning**

I. Introduction

X-RAY computed tomography (CT) plays a crucial role in various areas spanning from medicine and biomedical research to security, non-destructive testing of materials and study of archeological artefacts. A strength of this imaging modality is the broad range of spatial resolution levels that can be accessed – hospital scanners can typically resolve hundreds of µm, compact micro-CT machines can access resolutions of a few to tens of µm, and more specialized implementations like nano-CT enable resolutions on the sub-µm scale. A scanner's spatial resolution is, on the most basic level, defined by the x-ray source and detector, which introduce a combined blur to the images that defines (besides the signal-to-noise ratio) what can and cannot be detected[1]. While some x-ray sources (e.g. those with a variable focal spot) and detectors (e.g. CCD cameras with variable optics) allow controlling the blur to some extent[2], allowing to increase resolution within a limited range, generally this is a restrictive and inflexible constraint.

In an effort to overcome this, we have developed a (micro-CT) scanner that features a mask with alternating absorbing and transmitting septa in the x-ray beam path. Provided that the mask fulfills certain design criteria (see Section II), its presence introduces spatial frequencies higher than the cut-off normally imposed by the combined source and detector blur into the image formation process[3,4]. Therefore, the mask provides the ability of high-resolution imaging even with x-ray sources and detectors that would not normally allow this due their large focal spot and/or pixel size, respectively. Although the mask creates an under-sampling problem as some areas of the sample are shielded from radiation, this does not compromise the ability of reconstructing high-resolution images, provided that a dedicated cycloidal ("roto-translation") acquisition scheme is applied in combination with a suitable method for data recovery. The synergy between the beam structuring, "roto-translation" sampling and data recovery has led to a unique imaging concept ("cycloidal computed tomography") that enables rapid and highly flexible high-resolution CT scanning[5]. Here, we report on utilizing a machine learning technique (the Mixed-Scale Dense CNN architecture[6]) for data recovery, which outperforms previously applied "naive" approaches like bivariate interpolation.

II. Combining a structured X-ray beam setup with cycloidal sampling

In our scanner, the x-ray beam is structured into an array of long, thin beamlets (a few µm to tens of µm laterally, and extending uniformly in the y-direction) by means of a mask positioned immediately upstream of the sample (Fig. 1a). The mask period (p) matches the effective pixel size (s), i.e. p = s/m, where m is the magnification between the mask and detector (for simplicity, the distance between mask and

This work was supported through the ATTRACT initiative that is part of European Union's Horizon 2020 research and innovation programme under grant agreement No 777222. D. M. Pelt is supported by The Netherlands Organisation for Scientific Research (NWO), project number 016.Veni.192.235. A. O. is supported by the Royal Academy of Engineering (RAEng) under the Chair in Emerging Technologies scheme. C. K. Hagen is supported by the RAEng under the Research Fellowship scheme.

D. M. Pelt is with the Centrum Wiskunde & Informatica (CWI), 1090 GB Amsterdam, The Netherlands.

C. Maughan-Jones, O. Roche i Morgò, M. Endrizzi, A. Olivo and C. K. Hagen are with the Department of Medical Physics and Biomedical Engineering, University College London, London WC1E 6BT, United Kingdom. C. K. Hagen is the corresponding author (email: charlotte.hagen.10@ucl.ac.uk).

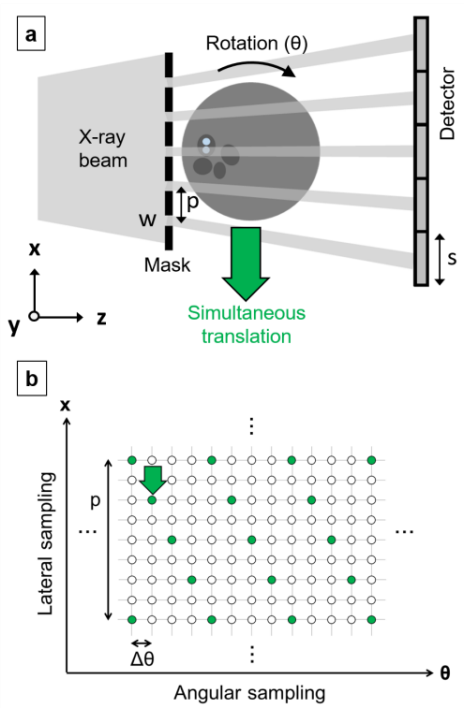


Figure 1. **a)** Schematic of the micro-CT scanner layout (seen from top), based on structuring the incoming x-ray beam into an array of narrow beamlets. Note the schematic is not to scale; in a real system the mask apertures are on the order of a few to tens of µm, the detector pixel size is around 50-100 µm, and the distance between mask and detector is between 15 cm and 1 m. **b)** Sinogram sampling grid realized through cycloidal ("roto-translation") sampling (see text). The grid is shown for a single mask period and a subset of angular increments. The filled circles indicate the available data points, whereas the empty circles indicate the entries which are to be recovered through an appropriate data recovery method.

sample, which is typically < 5% of the mask-detector distance, is considered negligible). It is important that the mask fulfills two design criteria; (1) its apertures (w) must be smaller than both the effective (i.e. scaled to the mask plane) projected focal spot of the x-ray source and the effective detector pixel size; (2) the mask period (p) must be sufficiently large to keep the beamlets physically separated. While the first criterion leads to the presence of additional spatial frequencies in the image formation process, the second criterion ensures that these frequencies remain accessible (by not being blurred). To a first approximation, spatial frequencies up to the inverse of the aperture width (1/w) are now present and can contribute to the in-slice spatial resolution of an image, provided that diffraction effects and detector cross-talk are negligible. It should be noted that the second criterion does not impose a constraint on the detector pixel size, as one can divert from the p = s/m relationship and employ a "line skipping" mask, in which p is an integer multiple (>1) of the effective pixel size such that only every second (or third etc.) detector pixel column receives one beamlet. This is also an effective strategy when working with detectors with non-negligible cross-talk between pixels; by skipping one or more pixel columns, any signal diffusion due to cross-talk can be minimized.

Generally, the use of beamlets leads to an under-sampled sinogram, i.e. the lateral sampling step defined by the mask period (p) is too large to adequately capture the newly gained spatial frequencies. To manage this problem efficiently, we have applied a cycloidal scanning scheme by which the sample is translated laterally (along x) simultaneously with being rotated around the vertical (y-) axis. This results in a "roto-translation" motion, where, if the translation is uni-directional, each feature in the sample follows a cycloidal trajectory. Although "roto-translation" sampling does not solve the under-sampling problem as such (as no additional data are acquired), it transforms a highly unbalanced sampling grid, which would correspond to rotation-only sampling, into a balanced grid where the available data are more evenly distributed (Fig. 1b). A balanced sinogram sampling grid is advantageous, as it provides a much better basis for restoring the missing data via a mathematical recovery method, to an extent that high-resolution images can be reconstructed.

We would briefly like to point out that there is another solution to the under-sampling problem that is, however, much less efficient than the cycloidal approach. In particular, the sample can be scanned laterally through the setup in multiple steps at each rotation angle; this leads to the actual acquisition of the missing entries in the sinogram. This procedure, which is often termed dithering, is incompatible with rapid scanning, as – although the overall exposure time can be re-distributed among the scanning steps – dead times are required for the necessary repeated stop-starting of the sample rotation and translation motors. This is not required in a cycloidal acquisition; in fact, the latter enables continuous scans (sometimes termed "fly-scans"), which do not suffer from any dead times, and in which the total scan time is solely determined by the exposure time. In the following, images obtained via dithering will be considered the "gold-standard" to which cycloidally sampled data are compared.

III. Convolutional neural network (CNN) reconstruction

To improve the quality of cycloidal computed tomography images, we have replaced "naive" data recovery based on bivariate interpolation by a machine learning approach. Specifically, we have trained and applied the recently proposed Mixed-Scale Dense (MSD) CNN architecture[6] to recover those sinogram entries that are not acquired during cycloidal acquisitions. In brief, the MSD CNN architecture differs from popular existing CNNs by using dilated convolutions instead of scaling operations to capture image features at different scales, and by connecting all network layers with each other. As a result, MSD networks typically require fewer intermediate images and learned parameters to achieve accurate results compared with existing CNNs, and are therefore well-suited for efficiently processing large datasets and accurately learning from relatively few training images. It has recently been demonstrated that the MSD CNN can improve the quality of images reconstructed from incomplete "conventional" CT measurements[7] (acquired without any beam structuring or cycloidal scanning). While previously the CNN was applied as a post-processing tool in the image domain, we have adapted it to and applied it in the sinogram domain. This has led to the recovery of complete sinograms (with as many entries as present in dithered

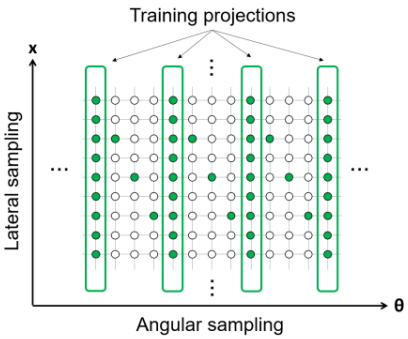


Figure 2. Sinogram sampling grid (shown for one mask period) corresponding to a cycloidal ("roto-translation") scan with an interleaved acquisition of training data. The angular intervals at which training data are acquired are variable.

sinograms), containing all the data required for the reconstruction of high-resolution tomographic images.

A particular challenge lay in the task of generating suitable training data for the CNN. We had aimed for a general approach that does not rely on scanning many samples of the same type and using these data for the training. Instead, we have developed a strategy by which training data are acquired as part of the scan of each individual sample. Specifically, the acquisition of a number of dithered projections, which would become the training data, was integrated into a cycloidal acquisition. The angles at which the training projections are acquired should be distributed evenly across the full angular range. The sinogram sampling grid that corresponds to this scanning sequence is shown in Fig. 2. For the most part, the grid is the same as the one shown on Fig. 1b, but for some angles all sinogram entries are now available.

IV. Results

To demonstrate the performance of CNN-based cycloidal computed tomography, we have acquired experimental data of a chicken bone (fixed in formalin) that was placed in a plastic container of approximately 9 mm diameter. To stabilize the chicken bone in the container, it was surrounded by agarose. The experimental setup comprised a MicroMax-007 HF x-ray tube (Rigaku, Japan) with a rotating molybdenum anode, operated at 40 kV and 25 mA, resulting in a horizontal focal spot of approximately 70 µm (full width half maximum). The detector was the Pixirad-2 photon counter with a pixel size of 62 µm. The mask (Creatv Microtec, USA) had a 79 µm period and 10 µm apertures. The mask and detector were positioned at 1.6 m and 2.53 m from the source, respectively. With these distances, the mask period covered two detector pixel columns when projected to the detector ("line-skipping"), which is equivalent to using a detector with twice as large pixels horizontally. A single, dithered scan of the sample was carried out, and the data were subsequently sub-sampled to mimic cycloidal acquisitions. During the scan, the sample was translated laterally in eight steps of 10 µm at each rotation angle, as this ensured that all sinogram entries were captured. Data were acquired as the sample was rotated in steps of 0.2 degrees over 180 degrees, corresponding to the acquisition of 900 x 8 = 7200 frames in total. The exposure time per frame was 2s.

A dithered image was reconstructed from all acquired data. Since dithering provides complete sinograms, no data recovery was required, and reconstruction was performed with filtered back projection (FBP). To generate cycloidally sampled data, only a single dithering step was considered, but this was a different step (corresponding to a different lateral sample offset) at each rotation angle, chosen in such a way that an interlaced sampling pattern like the one shown in Fig. 1b was created. However, at every 10^{th} angle, the cycloidal sub-sampling was interrupted and all dithering steps were included; the data from these angles were used as training projections (90 projections in total). Specifically, a 100-layer MSD network was trained with bicubic interpolated cycloidal sinograms as network input, and sinograms including only the dithered projections as training target. Out of 300 available sinograms (obtained for 300 cross-sections of the scanned sample), 270 were used for training the network, while the remaining 30 were used as a validation set to monitor performance during training. After training the CNN on the dithered projections, the CNN was applied to the remaining, cycloidally sampled sinogram areas, so as to restore the missing entries in these areas. Once the sinogram had been completed, image reconstruction was performed with FBP. For comparison, we have also applied bicubic interpolation to the same sinogram, and reconstructed an image via FBP. In both cases (CNN and bicubic interpolation), the 90 dithered training projections had been used in the tomographic reconstruction.

The results are shown in Fig. 3. Panel 3a shows the dithered image, panel 3b shows the cycloidal image reconstructed with bicubic interpolation, and panel 3c shows the cycloidal image reconstructed with CNN-based data recovery. Both of the cycloidal images were reconstructed from 22% of the data that contributed to the dithered image (this can calculated be from the fraction of training projections and the fact that cycloidal sampling produces an eighth of the frames produced by dithering). It can be seen that the spatial resolution in both of the cycloidal images is comparable to that in the dithered image, but that the CNN based recovery (panel c) leads to a better overall image quality than bicubic interpolation (panel b). To quantitatively analyse the performance of bicubic interpolation and CNN based data recovery, we have

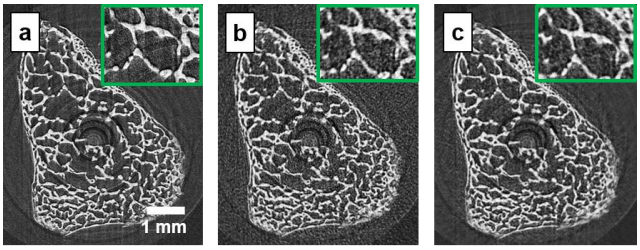


Figure 3. Tomographic images of a chicken bone: **a)** dithered image, **b)** cycloidally sampled image processed with bicubic interpolation, **c)** cycloidally sampled image processed using CNN based data recovery. The images shown in b) and c) were reconstructed from 22% of the data that contributed to the dithered image shown in a).

calculated the peak signal-to-noise ratio (PSNR) between the respective cycloidal image and the dithered image; the results are:

PSNR (bicubic interpolation) = 28.9

PSNR (CNN based recovery) = 30.9,

confirming the visual observation on image quality.

V. Conclusion

Cycloidal computed tomography offers an opportunity for flexible and efficient high-resolution scanning. The flexibility stems from the beam structuring, as it provides access to higher resolutions than those "dictated" by the source-detector blur. This implies that high-resolution imaging becomes possible also with "low-resolution equipment", such as x-ray sources with relatively broad focal spots and detectors with relatively large pixels, and that fast switching between resolution levels can be achieved simply by changing the mask inside the scanner. The flexibility of the concept also lies in another aspect: while the images shown in this document show attenuation contrast, cycloidal computed tomography can also be implemented in phase contrast mode. This can be achieved by including an array of beam stops, positioned immediately upstream of the detector, into the scanner. The beam stop array transforms the scanner into an edge illumination x-ray phase contrast imaging device[8], which has been demonstrated to produce high-quality phase contrast images for a broad range of samples and applications[9-11]. Phase contrast imaging has proven to be especially beneficial when a sample exhibits weak x-ray attenuation, as phase effects can be relatively stronger than attenuation ones for such samples[12].

The efficiency of cycloidal computed tomography stems from the dedicated sampling scheme ("roto-translation"); this removes the need for excessive, sub-pixel lateral scanning of the sample per rotation angle (i.e. dithering, an extremely cumbersome way to acquire data). A key advantage is that "roto-translation" sampling is compatible with continuous acquisitions ("fly-scans"). In this work, we have explored the benefit of applying the Mixed-Scale Dense (MSD) CNN architecture to the cycloidally sampled sinogram. We have shown that high-quality, high-resolution images can be obtained despite the fact that the sinogram is highly incomplete. Specifically, we have shown that the CNN outperforms bicubic interpolation. In future studies, the CNN should also be compared against an iterative reconstruction approach to fully appreciate its performance.

A strategy for generating training data has been proposed; this method is very general, and does not rely on previously acquired images of similar samples. In practical terms, when scanning in step-and-shoot mode, training projections can be acquired by applying dithering at a number of rotation angles. When scanning continuously ("fly-scan"), the training projections would have to be acquired before or after the scan. Although this leads to an increase in scan time, this can be relatively small (e.g. training projections were acquired at 10% of all angles in order to generate Fig. 3c).

In summary, we believe that, in combination, cycloidal computed tomography and CNN based data recovery bear great potential to increase the flexibility of CT machines in terms of spatial resolution and overall image quality, while also allowing to perform scans in a rapid and efficient manner.

References

[1] T. Buzug, "Computed Tomography; From Photon Statistics to Modern Cone-Beam CT," Berlin, Heidelberg: Springer, 2008

[2] J. Grimes, X. Duan, L. Yu, A. Halaweish, N. Haag, S. Leng, and C. McCollough, "The influence of focal spot blooming on high-contrast spatial resolution in CT imaging", Med. Phys. 42, 6011-6020, 2015

[3] P. Diemoz, F. A. Vittoria, and A. Olivo, "Spatial resolution of edge illumination x-ray phase contrast imaging", Opt. Express 22, 15514-15529

[4] C. K. Hagen, F. A. Vittoria, M. Endrizzi, and A. Olivo, "Theoretical framework for spatial resolution in edge illumination x-ray tomography", Phys. Rev. Appl. 10, 054050, 2018

[5] C. K. Hagen, F. A. Vittoria, M. Endrizzi, and A. Olivo, "Cycloidal Computed Tomography", UK Patent Application 1820362.0 (Filed Dec 13, 2018)

[6] D. M. Pelt and J. A. Sethian, "A mixed-scale dense convolutional neural network for image analysis", Proc. Natl. Acad. Sci. U.S.A. 115, 254-259, 2018

[7] D. M. Pelt, K. J. Batenburg, and J. A. Sethian, "Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks", J. Imaging 4, 128, 2018

[8] Zamir A, C. K. Hagen, P. C. Diemoz, M. Endrizzi, F. A. Vittoria, Y. Chen, M. A. Anastasio, and A. Olivo, "Recent advances in edge illumination x-ray phase contrast computed tomography", J. Med. Imaging 4, 040901, 2017

[9] C. K. Hagen, P. Maghsoudlou, G. Totonelli, P. C. Diemoz, M. Endrizzi, L. Rigon, R. H. Menk, F. Arfelli, D. Dreossi, E. Brun, P. Coan, A. Bravin, P. De Coppi, and A. Olivo, "High contrast microstructural visualization of natural acellular matrices by means of phase-based x-ray tomography", Sci. Rep. 5, 18156, 2015

[10] L. Massimi, C. K. Hagen, M. Endrizzi, P. R. T. Munro, G. Havariyoun, P. M. S. Hawker, B. Smit, A. Astolfo, O. J. Larkin, R. M. Waltham, Z. Shah, S. W. Duffy, R. L. Nelan, A. Peel, T. Suaris, J. L. Jones, I. G. Haig, and A. Olivo, "Laboratory-based x-ray phase contrast CT technology for clinical intra-operative specimen imaging", Proc. SPIE Vol. 10948, 2019

[11] D. Shoukroun, L. Massimi, F. Iacoviello, M. Endrizzi, D. Bate, A. Olivo, and P. Fromme, "Enhanced composite plate impact damage detection and chaacterisation using x-ray refraction and scattering contrast combined with ultrasonic imaging", Composites Part B 181, 107579, 2020

[12] A. Momose, "Development toward high-resolution x-ray phase imaging", Microcopy 66, 155-166, 2017

Projection-angle-dependent image intensifier distortion correction in high-speed tomography

Joaquim G. Sanctorum, Sam Van Wassenbergh, Van Nguyen, Jan De Beenhouwer, Jan Sijbers, and Joris J.J. Dirckx

***Abstract*—The geometric distortion pattern induced by electron deflection in x-ray image intensifiers is strongly influenced by the surrounding magnetic field. Therefore, variations in the magnetic field as sensed by the image intensifier causes the distortion pattern to alter accordingly, as is the case in C-arm imaging. In this work, we present a method to correct the image intensifier distortion, based on digital image correlation, and we show how this method can be efficiently used to correct for the time-varying distortion induced by an inhomogeneous distribution of ferromagnetic components in the sample rotation stage. Our experiments show that any projection recorded under an arbitrary angle can be corrected with a mean residual error below 0.48 pixels, using only three projection images of a custom reference mask sampled with a 120° angular interval to estimate the distortion behavior throughout a full rotation of the stage. As the method relies on widely available DIC software, it is very accessible and easy to use.**

***Index Terms*—image intensifier, computed tomography, distortion correction, digital image correlation, orientation-dependent**

I. Introduction

GEOMETRIC distortion is a well-known property of x-ray imaging systems utilizing image intensifiers, which is the result of two main contributions. The first contribution is the x-rays to be detected on a curved input phosphor to be subsequently projected onto a flat surface, introducing a radial stretching of the image. Secondly, the photo-electrons created and accelerated in the XRII are deflected under the influence of the surrounding magnetic field, which is mainly the Earth's magnetic field, causing a sigmoidal or "S-shaped" distortion. As the number of stereoscopic XRII-based imaging facilities is rising exponentially [1] and interest is rising in the use of XRII-based systems for low-cost tomography [2], [3], there is a need for fast and reliable distortion correction methods.

In general, an image transformation is applied to the deformed images to remove the distortion. However, the transformation to be applied is a priori unknown and during the past two decades, considerable efforts have been made to develop methods that determine a suitable correction transformation from a deformed reference input image, which is a square lattice pattern of dots in the majority of the existing correction methods. Due to the image distortion, the position of the markers will deviate from their ground truth position in the recorded images. These deviating positions are extracted from the images and related to their original positions to constitute (polynomial) dewarping functions, which are subsequently used to correct the deformed image through image resampling. Recently, a new method was developed in our lab, which is based on an entirely different principle, namely digital image correlation (DIC) [4]. This method has already been proven to be very accurate, robust and flexible.

In a typical C-arm set-up, the x-ray source and XRII rotate around the sample (e.g. a patient lying on a bench). During the rotation of the C-arm, the orientation of the XRII relative to the Earth's surface (and thus its magnetic field) is altered, causing the image distortion to be dependent on the projection angle. To solve this, various methods have been developed to estimate the distortion pattern throughout the trajectory of the XRII [5]–[7]. However, in a set-up where the XRII remains stationary, angle-dependent distortions may also arise due to an inhomogeneous distribution of ferromagnetic components in the rotation stage. In this work, we present an application of our previously developed DIC-based correction method to address such angle-dependent deformations using only a limited number of projection images for a full revolution of the stage.

II. Materials and Methods

A. Image acquisition

The radiographs are recorded using our 3DY²MOX (3D DYnamic MOrphology using X-rays) imaging system, of which the characteristics are described in [8], and can be found on the website of the project [9]. The recorded images consist of 2048 pixels × 2048 pixels covering a field of view of 292 mm × 292 mm. The maximum field of view was chosen to maximize the effect of distortion.

B. DIC-based distortion estimation

As mentioned, we used a DIC-based method to estimate the distortion pattern in the images. DIC is an optical full-field non-contact measurement technique to analyze deformations and strain behavior on the surface of objects [10]. The deformation analysis is enabled by applying a random speckle pattern to the object's surface and subsequently record a series of images of the object in its original and deformed state, yielding an image

J. G. Sanctorum and J.J.J. Dirckx are members of the Biophysics and Biomedical physics research group (BIMEF) of the University of Antwerp, Belgium. S. Van Wassenbergh is part of the Functional Morphology research group (FunMorph) of the University of Antwerp, Belgium. V. Nguyen, J. De Beenhouwer, and J. Sijbers are with the imec VisionLab research group of the University of Antwerp, Belgium. This research was funded by grant from the special research fund of the University of Antwerp (BOF-GOA 2016 33927).

series of the deformed random pattern. Through DIC analysis, the deformation of the pattern, which is assumed to be identical to the deformation of the surface, is characterized, establishing a relation between the surface coordinates of the original state and the deformed states.

In the case of image distortion, a similar relation between the original pixel positions and the distorted ones can be acquired through DIC analysis. To translate this optical technique to x-ray imaging, a random pattern of circular holes was designed and laser-cut in a sheet of stainless steel (referred to as the DIC-mask). The image of the DIC-mask recorded with an XRII will be deformed, which can be used along with the original pattern design as input for DIC analysis to readily obtain the relation between original and distorted pixel positions, which can then be used to correct the deformed images. A full description of the designed random pattern, the correction method, and the accuracy analysis can be found in [4].

C. Angle dependent correction method

Generally, image distortion can be formulated as a positional change of undistorted pixel locations, denoted as x_u and y_u:

$$\begin{aligned} x_d(x_u, y_u) &= x_u + \delta_x(x_u, y_u), \\ y_d(x_u, y_u) &= y_u + \delta_y(x_u, y_u), \end{aligned} \tag{1}$$

where the subscript d refers to the distorted position, and δ_x and δ_y are the unknown distortion surfaces, which depend on the undistorted pixel positions. Once the distortion surfaces are known, the gray values of the corrected image can be retrieved by evaluating the distorted image in the distorted coordinates:

$$I_c(x_u, y_u) = I_d\big(x_d(x_u, y_u), y_d(x_u, y_u)\big), \tag{2}$$

where I_d and I_c denote the distorted and corrected images, respectively. Generally, the distorted coordinate values x_d and y_d are non-integer, requiring interpolation of the gray values to correct the image. As described in [4], DIC analysis directly yields the distorted coordinate surfaces on a resolution of choice. In this work, these surfaces were retrieved on a 64 pixels × 64 pixels resolution for our 2048 pixels × 2048 pixels images. To obtain the intermediate values of the distorted coordinate surfaces needed to correct the images, 2D linear interpolation was used.

Once the distortion pattern is characterized by the distorted coordinate surfaces, any set of images recorded with the given set-up can be corrected for distortion. However, this statement is only true if the distortion is time-independent, which is not the case when the XRII is exposed to time-varying magnetic fields. Since the rotation of the stage is a periodic movement, the variation of the magnetic field (and thus the distortion pattern) is expected to be periodic as well. By recording a set of images of the laser-cut DIC-mask during the rotation of the stage and by analyzing the behavior of the distorted coordinate values of the pixels separately, we noticed the distortion behavior is indeed periodic for a certain image pixel. Moreover, as shown in Fig. 1, it resembles a sine function covering one period during one revolution of the stage. Based on these observations we can model the behavior of the distorted coordinate values using sinusoidal fitting:

$$\begin{aligned} x_d(\theta) &= x_0 + a_x \sin(\theta + \phi_x) \\ y_d(\theta) &= y_0 + a_y \sin\big(\theta + \phi_y\big) \end{aligned} \tag{3}$$

In this expression, θ is the projection angle. The fitting parameters x_0, y_0, a_x, a_y, ϕ_x, and ϕ_y are estimated using non-linear least-squares fitting. Due to the sinusoidal behavior of the distorted coordinates in both image directions, the distorted pixels move in elliptical trajectories over de image during the rotation of the stage. The C-arm distortion behavior as shown by Fahrig *et al*. [5] and Gutiérrez *et al.* [7] is very similar to the behavior observed in this work.

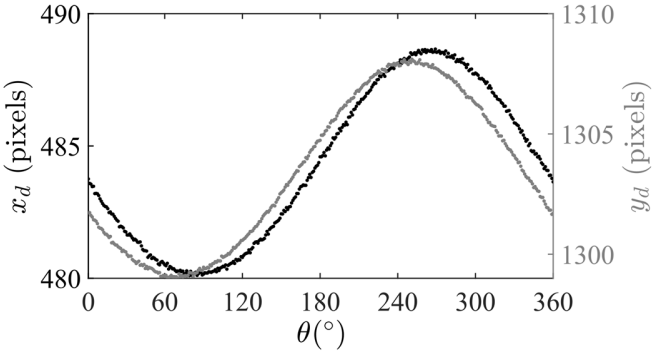


Fig. 1: Behavior of a distorted pixel position (x_d, y_d) as a function of the projection angle θ during a single revolution of the stage.

D. Experiments

For our experiments, the center of the rotation stage and the x-ray source were positioned at 264 mm and 1041 mm from the XRII, respectively. During a single revolution of the stage, we recorded 360 images using a 1° angular interval with the DIC-mask attached to the XRII, which are subsequently used in the DIC analysis to calculate the distorted coordinate surfaces. Next, we recorded 360 images of a square lattice pattern (11 × 11 grid of steel beads with a diameter of 2 mm and a center-to-center spacing of 25 mm) using the same projection angles. Then, the calculated distorted coordinate surfaces are used in two different ways. First, the surfaces are directly inserted in Eq. (2) to correct each distorted radiograph using the corresponding true distortion surfaces, which will serve as a reference. Then, a number of frames, referred to as N_{est}, are selected from this series to estimate the dynamic distortion behavior for intermediate angles using Eq. (3). These estimated distortion surfaces are subsequently used to correct the radiographs using Eq. (2).

By extracting the centroid positions of the dots after correction, and then calculating the Euclidean distance between the corrected centroid positions and their ideal positions, an estimation of the correction error can be made, as described in [4]. The standard deviation of the calculated Euclidean distances will serve as the error on the calculated means. In our experiment, we will investigate the influence of the number of projections used to estimate the distortion behavior over a single revolution of the stage on the mean and maximum error (Euclidean distance) in the corrected square lattice pattern images, denoted as R_{mean} and R_{max}, respectively. The results will be compared to those obtained by using the true corresponding distortion surfaces obtained through DIC

analysis. This way, we will deduce the minimum number of frames required to estimate the dynamic distortion behavior.

III. Results

In Fig. 2, we demonstrate the effect of the distortion correction on an image of the square lattice dot pattern. Before the correction, the distortion is clearly visible, which is shown in Fig. 2(a), whereas Fig 2(b) shows the corrected radiograph.

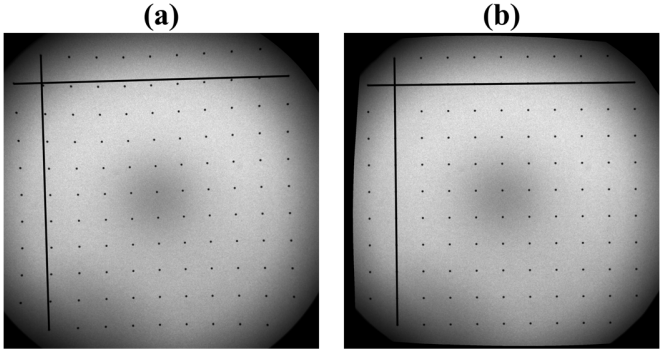


Fig. 2: (a) Original, distorted radiograph of the square lattice dot pattern. (b) Corrected radiograph of the square lattice pattern. The straight black lines drawn on the panels clearly show that the dots lie on straight lines after the correction, which corresponds to the ground truth.

A. Non-dynamic correction

For the non-dynamic correction, we used the calculated distorted coordinate surfaces of only the first image of the DIC-mask sequence to correct the distortion for the entire series of the square lattice pattern during a single revolution of the stage. Fig. 3 shows R_{mean} and R_{max} as a function of the projection angle. For images near the starting angle, the mean error is below 0.5 pixels and the maximum error is between 1 and 2 pixels. However, due to the temporal varying distortion, the mean error steeply rises slightly above 3 pixels and the maximum error rises above 10 pixels. This clearly illustrates the failure of a static correction approach.

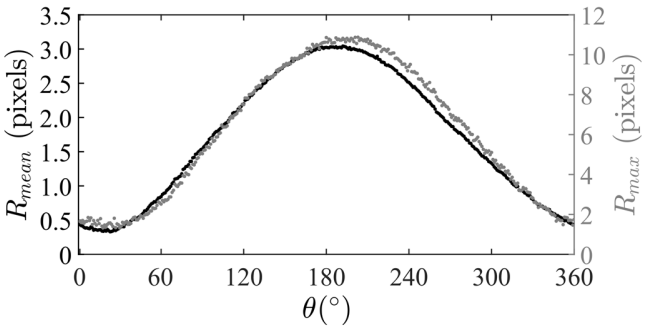


Fig. 3: Mean and maximum error (R_{mean} and R_{max} respectively) as a function of the projection angle θ for a single revolution of the stage using a static correction approach.

B. Non-dynamic correction

Next, we took into account the varying behavior of the distortion pattern by estimating the distorted coordinate values using the sinusoidal fits as defined in Eq. (3). Since three fitting parameters are present in each fit, the theoretical minimum value of N_{est} would be three. We have considered different values of $N_{est} \in [3,360]$ and selected that number of frames with equiangular intervals ($\Delta\theta$) to estimate the distortion behavior over a full revolution of the stage. Next, for each N_{est}, we corrected the full set of radiographs of the square lattice dot pattern and calculated R_{mean} and R_{max} in each frame, which are shown in Fig. 4 as a function of the projection angle θ. Regardless of N_{est}, any radiograph recorded under an arbitrary angle can be corrected with a mean residual error below 0.48 pixels and a maximum residual error below 1.66 pixels.

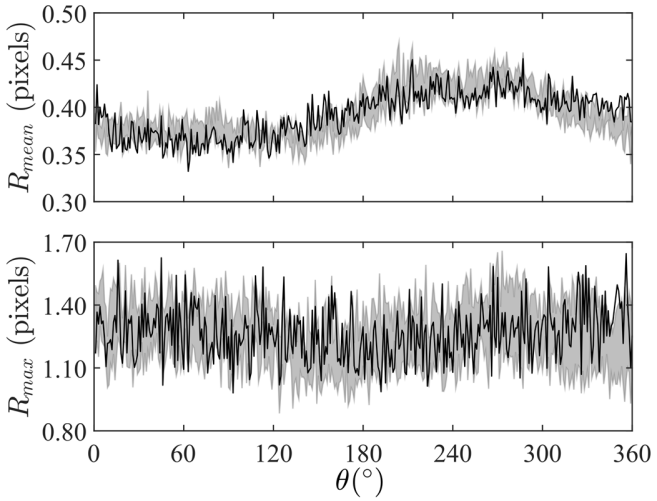


Fig. 4: The mean and maximum residual error (R_{mean} and R_{max}) as a function of the projection angle θ using a dynamic correction approach. The gray zones depict the range wherein the errors lie obtained using the different values of N_{est}, whereas the black line denotes the errors obtained using the true corresponding distortion surfaces obtained through DIC analysis.

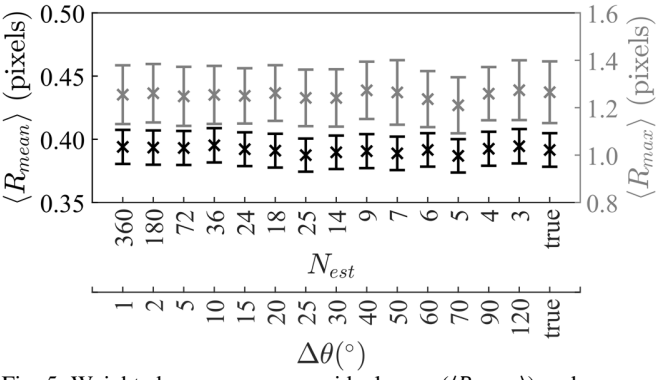


Fig. 5: Weighted average mean residual error ($\langle R_{mean} \rangle$) and average maximum error ($\langle R_{max} \rangle$) over a single revolution of the stage as a function of N_{est} or the corresponding angular interval $\Delta\theta$. The average errors obtained using the true surfaces are shown on the right.

For each value of N_{est}, we calculated the weighted average mean error and average maximum error over the 360 corrected frames, denoted as $\langle R_{mean} \rangle$ and $\langle R_{max} \rangle$, respectively. The results are shown in Fig. 5 as a function of N_{est} (or the sampling interval $\Delta\theta$). Fig. 5 shows that, regardless of N_{est}, $\langle R_{mean} \rangle$ lies below 0.40 pixels, whereas $\langle R_{max} \rangle$ lies between 1.2 pixels and 1.4 pixels. Moreover, for each N_{est}, the average values are very

similar and an ANOVA test showed that none of the datasets have significantly different averages.

IV. Discussion

This work aimed at presenting a projection-angle-dependent distortion correction method for XRII-based imaging. Since there is a large variety of used equipment and imaging conditions, it is appropriate to express our results in units of distance instead of pixels to compare our work to previously developed methods. Fig. 4 shows that any radiograph recorded under an arbitrary angle can be corrected with a mean residual error below 0.48 pixels, which corresponds to 0.068 mm as our images consist of 2048 pixels × 2048 pixels covering a field of view of 292 mm × 292 mm (the diameter of the XRII is 14"). Fig. 5 shows that the weighted average mean residual error is around 0.39 pixels, regardless of the sampling interval, which corresponds to 0.056 mm. As the ANOVA test showed no significant difference between the datasets, three images sampled with an angular interval of 120° are sufficient to accurately estimate the full dynamic behavior during a single revolution of the stage.

One of the earliest dynamic correction methods was introduced by Fahrig *et al.* [5]., who obtained mean residual errors of 0.039 mm for a field of view of 246 mm × 246 mm by sampling recordings of a square lattice pattern reference mask using an angular interval of 2°. To estimate the distortion for intermediate angles, linear interpolation was used. This result seems more accurate than ours, but the field of view is considerably smaller, implying less strong distortion. Liu *et al.* [6] used a similar approach, but instead of linear interpolation between the data points, they used polynomial fitting, reaching an accuracy of 0.048 mm using a sampling interval of 5° to 10° for a 225 mm × 225 mm field of view. Gutiérrez *et al.* [7] used a sampling interval of 30° to obtain an accuracy of 0.142 mm for a 330 mm × 330 mm field of view. This was done by first smoothing the trajectories of the distorted pixel coordinates before applying polynomial fitting. Compared to these results, our correction method blends in, as the average accuracy, as well as the size of the field of view, fall in between. Our method stands out from the rest due to the sampling interval of 120° and the need for only three images to estimate the full 360° behavior of the varying distortion.

As recording radiographs and subsequent distortion correction are the first steps in the CT-chain, an inaccurate correction method can result in the accumulation of errors in further processing steps, such as system calibration and tomographic reconstruction, which will be the topic of our future research. Therefore, users of XRII-based tomographic systems are advised to dynamically correct for the distortion, as they might be unaware of the presence of ferromagnetic components in rotating parts of their set-up.

V. Conclusion

We have presented an application of a DIC-based distortion correction method for XRII-based tomography purposes. Our experiments show that, based on only three projection images sampled using a 120° angular interval, the full 360° behavior of the distortion could be corrected with a mean residual error smaller than 0.48 pixels for each frame. An ANOVA test showed that this large sampling interval could be used without accuracy loss. As DIC software is widely available, the method is very accessible and easy to use.

Acknowledgements

The authors would like to acknowledge William Deblauwe and Wim Huyge for the construction of the rotation stage. This research is funded by a grant from the special research fund of the University of Antwerp (BOF-GOA 2016 33927).

References

[1] D. Orenstein, "Brown's influential biomechanics X-ray technology grows ever more powerful." https://phys.org/news/2016-12-brown-influential-biomechanics-x-ray-technology.html (accessed Mar. 17, 2020).

[2] C. de Molina, J. Pascau, M. Desco, and M. Abella, "Calibration of a C-arm X-Ray System for Its Use in Tomography," in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, Cham, 2014, pp. 245–248, doi: 10.1007/978-3-319-00846-2_61.

[3] M. Abella, C. de Molina, N. Ballesteros, A. García-Santos, Á, Martínez, "Enabling tomography with low-cost C-arm systems," *PLOS ONE*, vol. 13, no. 9, p. e0203817, Sep. 2018, doi: 10.1371/journal.pone.0203817.

[4] J. G. Sanctorum, S. Van Wassenbergh, P. Aerts, and J. J. J. Dirckx, "Technical Note: Correction of geometric x-ray image intensifier distortion based on digital image correlation," *Med. Phys.*, vol. 47, no. 2, pp. 597–603, 2020, doi: 10.1002/mp.13944.

[5] R. Fahrig, M. Moreau, and D. W. Holdsworth, "Three-dimensional computed tomographic reconstruction using a C-arm mounted XRII: Correction of image intensifier distortion," *Med. Phys.*, vol. 24, no. 7, pp. 1097–1106, Jul. 1997, doi: 10.1118/1.598013.

[6] R. R. Liu, S. Rudin, and D. R. Bednarek, "Super-global distortion correction for a rotational C-arm x-ray image intensifier," *Med. Phys.*, vol. 26, no. 9, pp. 1802–1810, 1999, doi: 10.1118/1.598684.

[7] L. F. Gutiérrez, C. Ozturk, E. R. McVeigh, and R. J. Lederman, "A practical global distortion correction method for an image intensifier based x-ray fluoroscopy system," *Med. Phys.*, vol. 35, no. 3, pp. 997–1007, Feb. 2008, doi: 10.1118/1.2839099.

[8] J. G. Sanctorum, D. Adriaens, J.J.J. Dirckx, J. Sijbers, C. Van Ginneken, P. Aerts, S. Van Wassenbergh, "Methods for characterization and optimisation of measuring performance of stereoscopic x-ray systems with image intensifiers," *Meas. Sci. Technol.*, 2019, doi: 10.1088/1361-6501/ab23e7.

[9] "Technical information - 3-Dimensional DYnamic MOrphology using X-rays (3D2YMOX) - University of Antwerp." https://www.uantwerpen.be/en/research-groups/3d2ymox/technical-information/ (accessed Jan. 06, 2020).

[10] M. A. Sutton, J.-J. Orteu, and H. Schreier, *Image Correlation for Shape, Motion and Deformation Measurements*. Boston, MA: Springer US, 2009.

Dual-energy in industrial CT – how simple math can reveal structures hidden by artefacts

Jochen Butzer, Annika Döring, Sven Gondrom-Linke, Markus Harbecke

***Abstract*— Imaging multi-material parts can be a challenging task in industrial CT, especially when beam-hardening artefacts shadow information. However, a simple and easy-to-use combination of two low-noise datasets from scans at different filter settings can reveal structures in those areas. This dual-energy approach has been evaluated for imaging a sensor and a connector where a significant reduction of artefacts can be achieved.**

***Index Terms*—Computed tomography, Dual-energy, Industrial CT, Image processing, Image fusion, Artefact reduction**

I. Introduction

MULTI-ENERGY imaging to optimize the quality of a CT slice is well-known [1] and used in certain areas of medical applications [2] and research [3]. However, dual- or multi-energy techniques are not widely used in industrial CT applications where failures within parts or geometry deviations are being investigated. A possible reason for this could be that the implementations are often proprietary [4] or require too extensive calculations or simulations [5] for an easy-to-use implementation.

The potential of dual-energy applied to industrial CT is high, especially since many of the samples are composed of more than one material. Imaging of such multi-material samples can be challenging if their attenuation differs strongly. In CT slices of e.g. connectors, which contain at the same time high-absorbing materials such as metal contacts and low-absorbing materials such as polymer overmolds, beam-hardening artefacts often worsen the image, since in certain directions of the CT acquisition a high thickness of metal needs to be penetrated. Those artefacts darken the area between high-absorbing structures thereby obstructing the evaluation.

II. A simple implementation

We have used a simple dual-energy implementation that can be carried out without the need for proprietary software. The principal idea behind this implementation is that different x-ray spectra influence the appearance of beam-hardening artefacts in a different way than they influence the representation of the part. By combining two scans at different filter settings, structures can be made visible again thus enabling an undisturbed view of areas formerly shadowed by artefacts.

A. Creation of the dual-energy dataset

A variation of the x-ray spectrum can be achieved in two ways. The first approach is to alter the acceleration voltage of the microfocus x-ray tube. However, the position of the focal spot, with a size of down to a few microns, can drift when changing the voltage. This would result in two different object representations, which should be avoided in order to perform the dual-energy calculation on the images of the object in the same orientation and magnification. Thus, the second approach of using different filters to alter the spectrum has proven beneficial. We performed a simulation of the influence of different filters on the x-ray spectrum shown in Fig. 1.

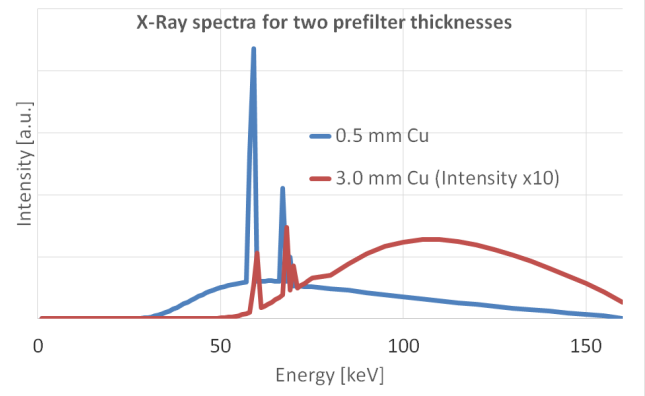


Fig. 1. Schematic representation of spectra for different thicknesses of a copper prefilter. The unfiltered spectrum was simulated for an x-ray tube operated at 160 kV using SpekCal [6]. Data from XCom [7] was used to determine the influence of the prefilters. In the spectrum with a thicker prefilter, the Bremsstrahlung dominates the contributing energies (note that the intensity of this plot is magnified by a factor of 10). In the spectrum with only 0.5 mm copper prefilter, the characteristic radiation is more pronounced.

The filters should be selected in a way that they change the appearance of the beam-hardening artefacts. Since dose is typically not the limiting factor in industrial CT imaging, each scan is performed with the aim of generating a top quality dataset. A top quality is especially indicated by a low noise level, which can be achieved by sufficient angular steps, by averaging projections and making use of the full dynamic range of the detector [8]. By reconstructing the CT slices in float values, the information is best preserved. This is necessary to perform calculations on the images in the next step.

Jochen Butzer is with Robert Bosch GmbH, 71272 Renningen, Germany (e-mail: jochen.butzer@de.bosch.com)
Annika Döring is with Robert Bosch GmbH, 71272 Renningen, Germany (e-mail: annika.doering@de.bosch.com)
Sven Gondrom-Linke is with Volume Graphics GmbH, 69115 Heidelberg, Germany (e-mail: gondrom@volumegraphics.com)
Markus Harbecke is with diondo GmbH, 45525 Hattingen, Germany (e-mail: markus.harbecke@diondo.com)

B. *Combining the two scans*

The two resulting volumes each consist of a stack of CT slice images of the identical object. A combination of the slice image I_L from the low-energy scan and the corresponding slice image I_H from the high-energy scan results in the new dual-energy image I_D. The calculation is simply done by multiplying image I_L with a weighting factor m and subtracting it from the image I_H as in (1).

$$I_D = I_H - m\, I_L \tag{1}$$

III. An Application Example

A. *Scan and dual-energy combination*

As a first example, we have investigated a sensor where an air inclusion, henceforth referred to as void structure, is located between two metal contacts. Beam-hardening artefacts appear between the contacts, partly shadowing the void and hindering its evaluation, as can be seen in a zoomed slice of the low-energy scan shown in Fig. 2a. This image was acquired using a diondo d5 CT-system with a microfocus x-ray tube operated at 200 kV and 75 µA and a prefilter of 0.5 mm copper and 3 mm aluminum. The voxel size of this scan was 15 µm, the scan time was 1.25 hours. The scan setup is shown in Fig. 3.

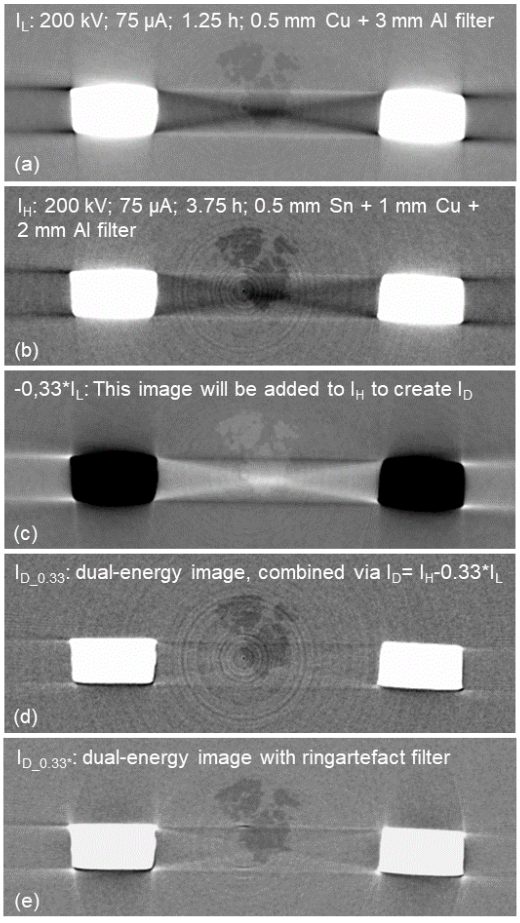


Fig. 2. Between two contacts of a sensor, beam-hardening artefacts hinder the assessment of voids (a). Their occurrence is less pronounced in the high-energy scan (b), but still not removed completely. After a successful image, combination (d), the artefact is heavily suppressed allowing for an evaluation of inner structures in this area. Software filtering can reduce ring artefacts (e).

A second scan was performed at higher energies achieved by changing the prefilter to a combination of 0.5 mm tin, 1 mm copper and 2 mm aluminum. The scan time was increased from 1.25 h to 3.75 h to counterbalance the reduced intensity. The resulting image is shown in Fig. 2b.

Comparing the high-energy image I_H with the low-energy image I_L, the appearance of the artefact becomes less pronounced. At the same time, the contrast of the void structure increases significantly in I_H.

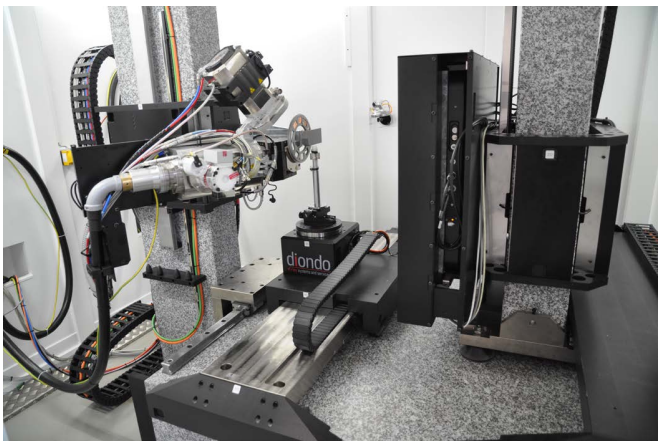


Fig. 3. Setup of our CT system diondo d5. A filter-changer is mounted on the holder of the directional microfocus x-ray tube. However, changing the filter manually between the scans is also possible making this method hardware-independent. The sample is positioned on a rotation stage placed between the microfocus x-ray tube and the a-Si detector.

To perform the dual-energy calculation, the image I_L is now multiplied by the weighting factor m = 0.33. The image of -0.33*I_L is shown in Fig. 2c. It contains scaled information about the artefact and residual information from the sample. Adding this image to I_H results in the dual-energy image I_D shown in Fig. 2d. The void structure can be successfully recovered while the disturbing beam-hardening artefact has been removed almost completely. This combination uses the fact that the artefacts appear less pronounced in the image with the thicker copper filter. Hence, the image with the thinner filter can be subtracted from it removing the artefact almost completely while preserving the structures of the object.

Since the overall structures of the part are identical in both scans, this information can be preserved and is not worsened by applying the combination. For this to hold, the noise has to be low in both datasets, since it is enhanced in the combined image I_D. Such a negative effect can be seen in Fig. 2d. The noise level increases and the ring artefacts appear more pronounced than in the original images.

Applying a ring artefact filter on the sinograms by amplifying underperforming pixels as in [10], the ring artefacts and noise can be reduced while the structures are mostly preserved. This is shown in Fig. 2e, where remainders of the beam-hardening artefacts reappear at the edges.

B. *Selecting an optimal weighting factor*

The procedure of generating image combination as in (1) for different values of *m* has been implemented as a macro in ImageJ [9]. In this batch operation, the calculations are performed in a selected range. We started with a coarse sampling followed by a finer stepping. The result of such a coarse sampling is shown in Fig. 4. An appropriate choice of *m* can be made by selecting the best resulting image. This is typically achieved just before the contrast of the void structure starts to weaken.

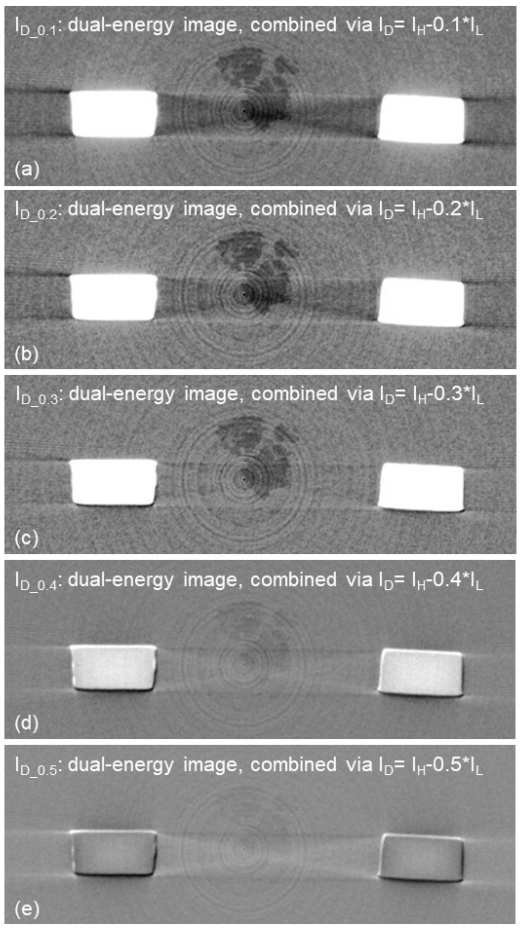


Fig. 4. Dual-energy images are created for different values of the weighting factor *m*. Increasing *m* makes the dark artefacts disappear and then reappear as bright structures. The contrast of the void structure is reduced with increasing *m*. An optimal selection of *m* shows enough contrast to evaluate the void with as little contribution from the artefact as possible. For an optimal visualization, the images have been normalized to the surrounding material.

It can be seen that the beam-hardening artefacts disappear before the contrast of the structures of the sample is reduced significantly. This small difference in the appearance of beam-hardening artefacts and structures of the sample at different energies enables the successful implementation of this method. Due to the tiny difference, the images need to be reconstructed in float values to be used in the calculation. The procedure of finding *m* is carried out on a slice where the artefacts appear most prominent. A dual-energy volume can be generated by applying the calculation presented in (1) for all slices. The value of *m* is kept constant for the whole volume.

C. *Choosing the best filter combination*

To improve the sample throughput, an automated filter-changer was installed enabling successive scans with different filters overnight. Furthermore, benchmark tests of filter settings for new material combinations can now be scanned automatically. A sketch of the filter-changer is shown in Fig. 5.

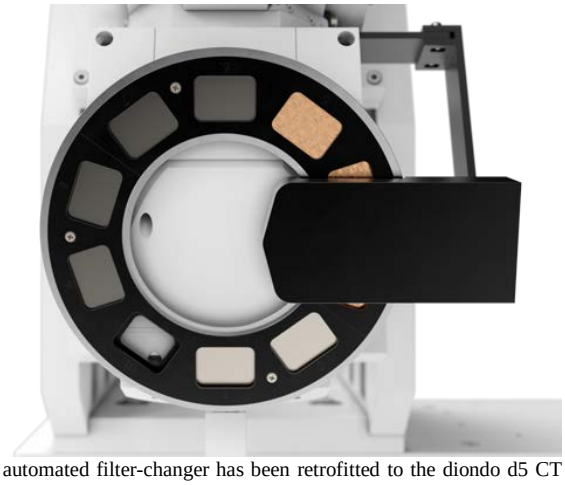


Fig. 5. An automated filter-changer has been retrofitted to the diondo d5 CT system allowing for scans at different filter settings without user-interaction. Complete dual-energy scans can now be performed overnight.

To investigate the criteria for an optimal choice of the prefilter combination, we have performed successive scans of the same sensor as in Fig. 2 and 4 with different prefilter sets. For all possible combinations of filters, the process of determining the best factor *m* as described before has been carried out with the aim of eliminating the artefacts between the two metal contacts as good as possible. The sets and resulting values of *m* are shown in Table 1.

TABLE 1: OPTIMAL CHOICES OF THE WEIGHTING FACTOR *m* FOR A COMBINATION OF DIFFERENT FILTER SETS

Filters [mm]			Set	#1	#2	#3	#4	#5
Sn	Cu	Al						
0	0	4	#1					
0	0.5	3	#2	0.54				
0	1	3	#3	0.40	0.64			
0	2	4	#4	0.25	0.45	0.55		
0.5	1	2	#5	0.18	0.33	0.42	0.72	
1	1	2	#6	0.13	0.18	0.25	0.30	0.40

The weighting factor *m* decreases with thicker filter differences, since the appearance of the artefact differs more between both scans. If the filter of the high-energy scan and the low-energy scan are too similar, the dual-energy combination cannot be carried out successfully. These cases are indicated by light gray values in Table 1. Either, the artefacts remain disturbing or the contrast of the void is reduced significantly. This holds for filter combinations at lower energies as well as for filter combinations at higher energies. The reason is that both, information and artefact are now subtracted from the image in approximately the same amount.

IV. Results

We have applied this dual-energy approach to various samples and could prove the success by generating images almost free from artefacts in most cases. The approach presented here is a great improvement to the single scan data available without a dual-energy combination, even at thick prefilter settings. The best results were achieved in small samples containing a line of metallic objects such as wires. The filter for the low-energy scan can be chosen to the standard filter settings for this sample. For the high-energy scan, an addition of 3 mm copper filter has proven beneficial.

An example of such a sample where the dual-energy method has clearly achieved a huge impact is the scan of a connector shown in Fig. 6. The scans have been performed with an x-ray tube operated at 215 kV and 0.11 mA. The filter was altered from 0.5 mm copper to 3.5 mm copper between the scans. To achieve a top quality in both scans, the acquisition time was adjusted resulting in 1 hour and 4 hours scan time respectively. For a dual-energy combination, $m = 0.28$ was chosen.

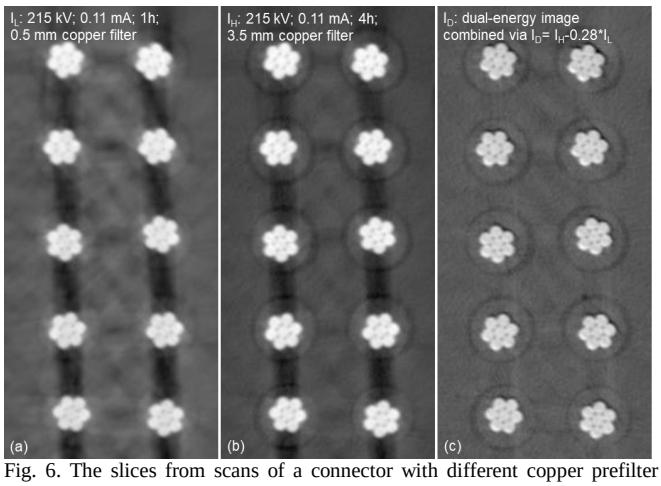


Fig. 6. The slices from scans of a connector with different copper prefilter thicknesses show beam-hardening artefacts in-line with the wires. The dual-energy image I_D (c) is a combination of the high-energy image I_H (b) and the low-energy image I_L (a) where those artefacts are reduced. X-ray voltage and current are kept constant while the exposure time has been adjusted resulting in a longer scan time for the high-energy scan. Areas between the metal wires are visible in this zoomed image only after the combination of image I_L and I_H. Small details like the air surrounding the wire shielding can be assessed now.

The shielding of the wire between the metal can neither be evaluated in the low-energy image I_L shown in Fig. 6a nor in the high-energy image I_H shown in Fig. 6b. However, this information becomes clearly visible in the combination of both images resulting in the dual-energy image I_D shown in Fig. 6c. Formerly dark areas unveil structures surrounding the wire shielding which have been partly shadowed by beam-hardening artefacts in the images I_L and I_H. The beam-hardening artefacts are completely suppressed. Almost no additional noise has been introduced. Now a full evaluation of the wire shielding is possible.

V. Outlook

The factor *m* for merging the two volumes is currently determined by visual inspection of a batch created image stack. An automated suggestion of this factor is desirable. We plan to investigate such an implementation in the future.

In our work, we have shown the great impact of dual-energy for CT scans of a connector and a sensor. This method will certainly be applied to many more samples where an improvement of the image quality is to be expected. To prepare for a routine use of this method, we plan to investigate the optimal choice of filter combinations for different materials with a focus of minimizing the required scan time.

References

[1] R. Alvarez *et al.*, „Energy-selective reconstructions in X-ray computerized tomography", Phys. Med. Biol. 21, Vol 5, p 733-744, 1976

[2] J. R: Grajo *et al.*, "Dual energy CT in practice: Basic principles and applications", Appl. Radiol. 45 Vol 7, p 6-12, 2016

[3] J. Butzer, "MARS-CT, Biomedical Spectral X-ray Imaging with Medipix", Thesis, University of Canterbury, Christchurch, New Zealand, 2009 available *https://ir.canterbury.ac.nz/handle/10092/3863*

[4] I. Bauscher *et al.*, "Korrektur von Artefakten in CT-Daten (Streustrahlen und Multimaterialien)" in *DGZfP-Jahrestagung 2018*, Leipzig, Germany, 2018, available: *https://www.ndt.net/article/dgzfp2018/papers/Mi.2.C.2.pdf*

[5] C. Heinzl *et al.*, „Artefaktreduktion mittels Dual Viewing für Röntgencomputertomographie von Multimaterialbauteilen" presented at the DGZfP Jahrestagung 2010, available at *https://jt2010.dgzfp.de/Portals/jt2010/BB/mi1b2.pdf*

[6] G. Poludniowski *et al.*, "SpekCalc: a program to calculate photon spectra from tungsten anode x-ray tubes" Phys Med Biol. 54 Vol 19, p. 433-438, 2009

[7] M. Berger *et al.*, "XCOM: Photon Cross Section Database (version 1.5)." Available at: *http://physics.nist.gov/xcom.* National Institute of Standards and Technology, Gaithersburg, MD, 2010

[8] Deutsche Gesellschaft für Zerstörungsfreie Prüfung (DGZfP), Merkblatt D6: „Anforderungen und Rahmenbedingungen für den Einsatz der Röntgencomputertomographie in der Industrie", 2019

[9] C. Schneider *et al.*, "NIH Image to ImageJ: 25 years of image analysis", Nature methods 9 Vol. 7, p. 671-675, 2012

[10] D. Niederloehner, "Untersuchungen zur Energiewichtung in der medizinischen Röntgenbildgebung mit dem Medipix2-Detektor", Thesis, Friedrich-Alexander University, Erlangen-Nürnberg, 2006

Dual domain learning basis material decomposition for dual-spectral CT via attention mechanism

Genwei Ma, Xing Zhao*, Shusen Zhao

***Abstract*—Theoretically two basis materials with different linear attenuation coefficients can be accurately reconstructed using dual-spectral CT (DSCT) technique. However, the reconstruction of dual-spectral CT is a nonlinear problem and existing scatting and noisy, which can not give the analytical solution easily, and make much challenge to solve this reconstruction problem by using numerical method or optimization programs. Moreover, the nonlinearity lead to the large noisy signals of basis material images by using traditional calibration method. In this paper, considering the advantage of neural network-based method in solving the nonlinearity problem and artifact suppression, we propose a learning-based method for dual domain learning basis material decomposition method using attention mechanism. The method includes two strategy: projection domain decomposition and image domain refinement. Furthermore, our learning-based method requires neither information of the spectra nor the linear attenuation coefficients. Numerical simulation and real data experiment demonstrate that our method can achieve promising gains in terms of basis materials decomposition and monochromatic images, and reduce hardening artifacts, effectively.**

***Index Terms*—dual-spectral CT, basis material decomposition, dual domain learning, attention mechanism**

I. Introduction

In x-ray dual spectral computed tomography (DSCT), an object is scanned with two different x-ray spectra. The collected polychromatic projections are utilized to perform energy- and material-selective reconstructions [1]. Compared with the conventional x-ray computed tomography with a single x-ray source, DSCT not only can distinguish basis material more better, but also allow monochromatic imaging and basis material decomposition [2]. Currently, dual-spectral x-ray data collect methods including kVp-switching [3], dual-source scanning [4], dual-layer detection [5] and simplistic two-pass scanning.

Several decomposition methods for dual-spectrum CT were developed over the years. In summary, its include three classes of reconstruction methods: image-domain methods, projection-domain methods and iterative methods. Image-domain decomposition methods first reconstruct images from the dual-spectra sinograms using CT reconstruction algorithms, such as filtered back projection (FBP), then decompose basis material image from two reconstruction images [6]. This type of image-domain decomposition makes substantial first-order approximations of the true images, resulting in quantitatively inaccurate results and cannot remove the beam hardening artifacts well. Alternatively, projection domain decomposition methods by solving a nonlinear integral equation obtain line integral of basis material images from the polychromatic projections. Generally, dealing with the nonlinear integral equation is approximated by high order function passing using some available information. Such as Empirical dual energy calibration (EDEC) for cone-beam computed tomography [7]. The difficulty of projection based methods are to obtain the parameters of high order functions and select calibration phantom. Several iterative methods incorporate an accurate physical model to reconstruct images directly from dual energy measurements [8]. These methods improve the decomposed image quality, effectively. The great computation cost and slow convergence speed significantly reduces the practicality of the algorithm. The ill-posedness of the decomposition process makes it sensitive to noise, leading to reduced image quality such as peak signal to noise ratio (PSNR) [9].

Recently, the deep learning technique or neural network algorithm has become more and more popular in many fields [10] [11]. In DSCT field, some researchers attempt achieving DSCT imaging using a single-energy CT data [12] or generating pseudo DSCT imaging using deep learning-based framework [13]. However, it is a concern when exist these phenomenon of same CT value but not same material and same material but not same CT value. Here, taking the advantage of neural network and combine with DSCT theory, a learning-based method for dual domain basis material decomposition is proposed. Specifically, The method includes two strategy. Firstly, in projection domain, using training phantoms input artificial neural network (ANN), which aim at realizing decomposition. Since DSCT is a nonlinear problem and the neural network has self-learning ability, which can learn and adapt to unknown or uncertain systems and optimize systems parameters through optimizer. Secondly, the structure of basis material images, which are reconstructed after first strategy, is not taken into consideration leading low-quality basis material images and artifacts. Hence, it is necessary attaching CNN combined attention mechanism for artifacts suppression and improving images quality by using basis material images combined with high- and low-energy images, which reconstruct by raw projection data.

The remainder of this paper is organized as follows. In section II, the DSCT system are introduced. In section III, we introduce ours neural network frames. In the sections IV,V, numerical simulation experiments are carried out to verify the dual domain learning method. Finally, conclusions are drawn in section VI.

G. Ma, X. Zhao, S. Zhao are with school of Mathematical Sciences, Capital Normal University, Beijing, 100048, China, and with Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, 100048, China. e-mail: (MaGenwei@126.com, zhaoxing_1999@126.com).

X. Zhao serves as the correspondence author (e-mail: zhaoxing_1999@126.com).

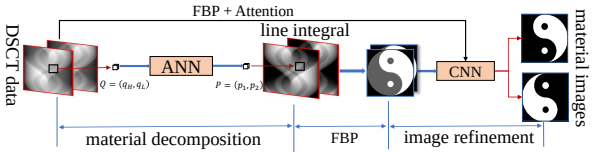


Fig. 1. illustration of proposed method architecture. Material decomposition by ANN and image refinement by attention mechanism's CNN.

II. Mathematical Model of DSCT

The polychromatic projection along an x-ray path L can be expressed as [7]:

$$q_i = -\ln \int w(E) e^{-\int_L \mu(x,E)dl} dE + S_j, i = 1, 2 \quad (1)$$

where $\mu(x, E)$ is the linear attenuation coefficient of the scanned object at point x with energy E, $w(E)$ is the normalized effective spectrum and S_j represent the scattered radiation.

In DSCT, the linear attenuation coefficient $\mu(x, E)$ is generally decomposed as a linear combination of some predefined basis functions, which are functions of energy E, i.e.,

$$\begin{aligned} \int_L \mu(x,E)dl &= \phi(E) \int_L \varphi_1(x)dl + \theta(E) \int_L \varphi_2(x)dl \\ &:= \phi(E)p_1(x) + \phi(E)p_2(x) \end{aligned} \quad (2)$$

where $\varphi_i(x)$ are the corresponding coefficients, which are functions of spatial position x and $\phi(E)$ and $\theta(E)$ are functions of energy E. There are two physical explanations for this equation: photoelectric/Compton effect based decomposition and basis material decomposition. For the first explanation, $\phi(E) = E^{-3}$ and $\theta(E) = f_{KN}(E)$ (Klein-Nishina function), where $\varphi_i(x)$ represents two effects distribution. For basis material decomposition, $\phi(E)$ and $\theta(E)$ represent the mass attenuation coefficients of the basis materials, $\varphi_1(x)$ and $\varphi_2(x)$ represent the density images of the two basis materials [8]. And p_i are the line integral of basis material. Simply, (1) and (2) can be write as:

$$\begin{cases} q_L &= q_L(p_1, p_2) \\ q_H &= q_H(p_1, p_2) \end{cases} \quad (3)$$

In general, solving this nonlinear inverse problem use Taylor expansion, polynomial approximation or rational fraction approximation in Radon domain or image domain. Such as EDEC method, which can be express as follow:

$$p_i = \sum_{n=0}^{N-1} c_{in} b_n(q_H, q_L) \quad (i = 1, 2) \quad (4)$$

where basis function use the polynomials $b_n(q_H, q_L) = q_H^k q_L^l$ with $k = 0, \cdots, K$ and $l = 0, \cdots, L$. The total number of basis functions is $N = (K + 1)(L + 1)$. The total number of coefficients in each coefficient vector c_i is N for each material.

III. Method

In this section, we introduce the whole process of our method which include two strategy. Material decomposition and image refinement. Fig.(1) give a illustration of proposed method architecture. In projection domain, The DSCT rawdata are used as the inputs of artificial neural network (ANN) and obtained the line integral of basis materials. Then, the reconstruction images from DSCT rawdata and line integral of basis materials is used as the inputs of CNN and obtained the basis material images.

A. ANN for material decomposition

As mentioned in Section II, the main difficulty of DSCT problem is to find the inverse

$$\begin{cases} p_1 &= q_L^{-1}(q_L, q_H) \\ p_2 &= q_H^{-1}(q_L, q_H) \end{cases} \quad (5)$$

where q_L^{-1}, q_H^{-1} are some yet unknown decomposition function. Thinking that any function can be expressed approximately by piecewise linear function and ANN also be proved to be approximate to piecewise linear function, So we define

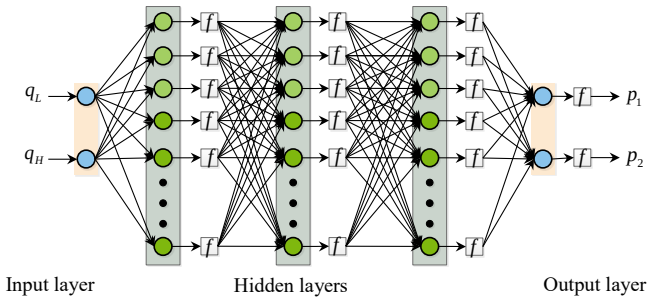


Fig. 2. ANN architecture for material decomposition.

$$P = f_n(\cdots(W_2 f_1(W_1 Q + b_1)\cdots) + b_n) \quad (6)$$

where $P = (p_1, p_2)^T$, $Q = (q_L, q_H)^T$, $f_i(i = 1, 2, \cdots, n)$ are the activiation function, here we set the Rectified Linear Units(ReLU) functions $f_i = \max(0, x)$ as the activiation function. $W_i, b_i(i = 1, 2, \cdots, n)$ are the unknown coefficients. In training processing, the unknown coefficients $W_i, b_i(i = 1, 2, \cdots, n)$ can be confirmed by neural network. Test objects will be decomposed by confirmed coefficients in test processing. According to (6) can set the structure of network, which are shown in Fig.(2).

B. Attention mechanism's CNN for image refinement

Considering the ill-posedness of the decomposition precess, which it noticeably sensitive to noise, Attention mechanism's CNN for decomposition image refinement is necessary, which consider the low peak signal to noise of decomposition precessing and the structure of reconstruction from DSCT raw data. Here we define each label to be a pair ground truth basis material images corresponding to the reconstruction images of the input images. Difference between our CNN architecture and typical CNN is that there is neither batch normalization layer nor pooling layer.

Our CNN consists of 7 layers, including 3 convolutional layers, 3 Rectified Linear Unit (ReLU) layers and one last loss layer. The convolutional layer has 32 filters with kernel

size 3×3 and no padding. The output of convolutional neural network can be formulated as:

$$\varphi_i = f(w_3 \otimes f(w_2 \otimes f(w_1 \otimes (R^T p_1, R^T p_2, R^T q_H, R^T q_L)))) \quad (7)$$

where R^T denotes the Radon inverse transform, f denotes activation function of Relu, $\otimes$ denotes the convolution operation. For the loss function, we use l_2 loss functions. During the training process, the parameters of our network are updated by Adam.

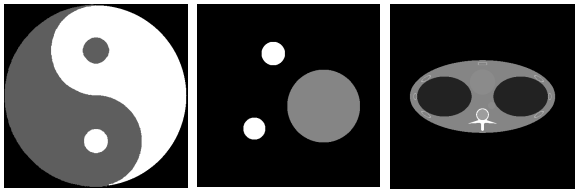


Fig. 3. Training phantoms: Yin-Yang phantom and composed phantom and test object: forbild thorax phantom.

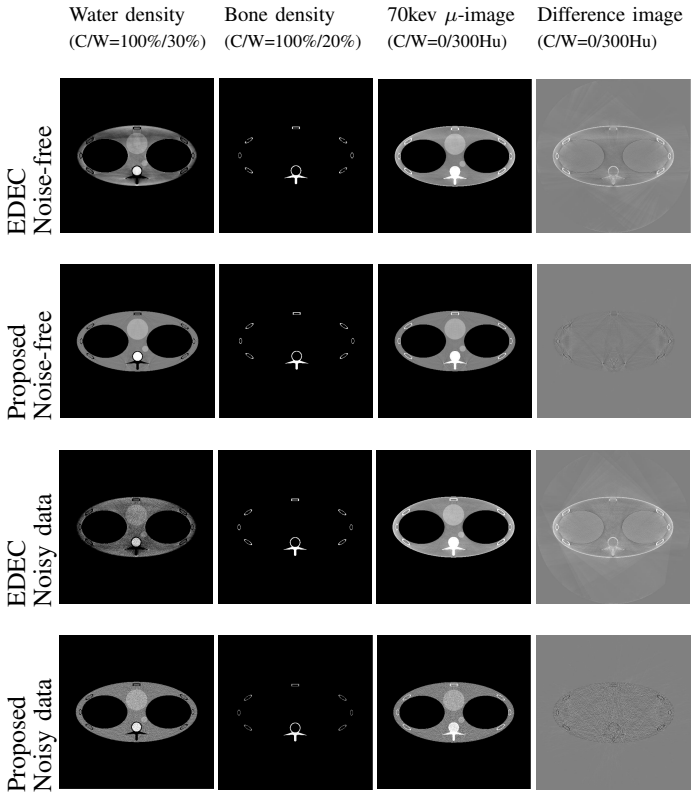


Fig. 4. Numerical simulation results from noise-free projections data and noisy projections data with EDEC and proposed method.

IV. Numerical simulation experiment

A. Datasets generation and Simulation environment

In numerical simulation experiments, we use two training phantoms to generate the datasets. Fig.(3) are two phantoms. One is Yin-Yang phantom and the size is $45cm \times 45cm$. The other phantom is composed of water basis material (big circle) and bone basis materials(two small circles), and the diameter of water basis material is 18cm, where the bone basis materials are 5.8cm and 5.4cm, respectively. We use composed phantom to denote this phantom. The two phantoms are filled with water density of $1.0g/cm^3$ and bone density of $1.92g/cm^3$. The two phantoms are utilized to generate DSCT data and line integral of basis materials which are training databases and label databases, respectively.The test phantom is a slice of the 3D forbild thorax phantom and the size is $40cm \times 40cm$, as shown in Fig.(3). For more detailed information, including the material and density of each part of the phantom, refer to its website (http://www.imp.uni-erlangen.de/phantoms/thorax/thorax.htm).

The x-ray spectrum are simulated by SpectrumGUI software (https://sourceforge.net/projects/spectrumgui/). The simulated normalized energy spectra are under the condition of a tube voltage at 80kVp without filter, and a tube voltage at 140kVp with a 1.0mm Cu filters, respectively. And linear attenuation coefficient obtained from NIST (https://physics.nist.gov/PhysRefData/XrayMassCoef/tab3.html).

The parameter of simulation experiments for fan-beam CT are set as follows. The distance from x-ray source to rotation center is 800mm, the distance from source to detector is 1200mm, totally 800 detector cells with cell size of 0.775mm. The phantoms are 512×512. Using the above high and low energy spectrum, a high- and low- spectra projection data from 0° to 360° with an interval of 1° are generated. We also simulate poisson noisy data corresponding to emission flux of 10^5 photons per measurement. The proposed neural network is implemented using the pytorch framework in Python3 environment under Windows 10. All experiments were performed using a GPU NVIDIA Quadro K2200 graphics card with 4GB video memory and an Inter(R) Xeon(R) E5-2620 v3 2.4GHz processor with 16GB RAM.

TABLE I
PSNR, MAD and SSIM of numerical simulation results of Fig.(4)

		PSNR	MAD	SSIM
70keV μ-image	EDEC without noise	38.6141	8.06460	0.95281
	Proposed without noise	47.4708	2.39689	0.98566
	EDEC with noisy	35.9762	11.8132	0.92934
	Proposed with noisy	32.6184	16.1719	0.94034
water density	EDEC without noise	31.2036	0.01349	0.31920
	Proposed without noise	46.5609	0.00218	0.99043
	EDEC with noisy	28.7922	0.02895	0.23759
	Proposed with noisy	31.6133	0.01571	0.93443
bone density	EDEC without noise	35.5786	0.01572	0.56070
	Proposed without noise	54.5126	0.00046	0.99847
	EDEC with noisy	32.6692	0.03178	0.56407
	Proposed with noisy	48.2349	0.00099	0.99844

B. Results of Simulation Experiments

Under this setting, we test the noise-free case and the noisy case by using the EDEC and the proposed method,respectively (see Fig.(4)). Moreover, three common image quality evaluations are employed, i.e., peak signal-to-noise ratio (PSNR), mean absolute deviation (MAD) and structural similarity (SSIM). The corresponding results are shown in Table I. All the experimental results demonstrate the merits of noise tolerance for the proposed method.

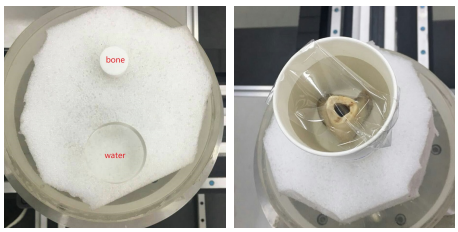


Fig. 5. The scanned phantom: training phantom and test object.

V. Real Experimental Results

In the real experiment, the scanned phantoms including training phantom and test object are shown in Fig.(5). The training phantom use composed phantom, which the bone basis material is bone-equivalent plastic with the diameter 1.7cm and water basis material is also water-equivalent plastic with the diameter 3.7cm. The test object is a sheep bone, which is immersed in water. The parameters of the scanning configuration are set as follows: An x-ray source is operated at tube voltage of 90 kV and 140 kV for low- and high-energy spectra scan, and the tube current is 5mA and 3mA, respectively. The detector is a linear detector, which has 3710 detector units with 0.083mm unit pitch. The SOD is 610.0mm and the SDD is 1225.0mm. 1440 projections are scanned per rotation. The dual-spectrum CT rawdata are shown in Fig.(5), which are clipped some region and displayed the main region.

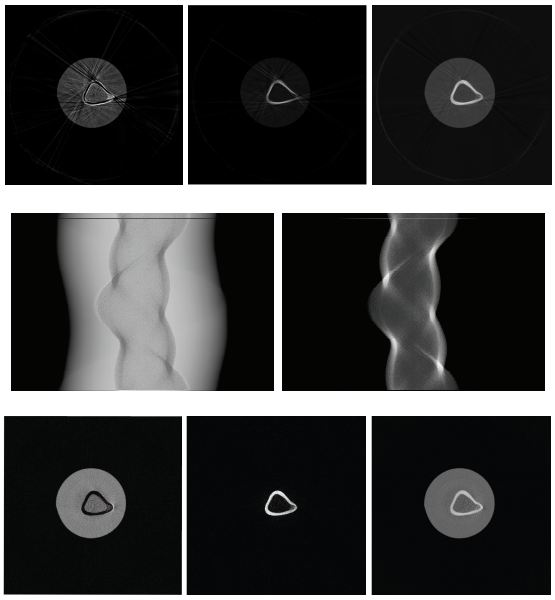


Fig. 6. Real experiments by using the EDEC (the first row), material decomposition by ANN (the second raw) and final results of proposed method (the third row).

The our methods is greater than the EDEC method. There are some reasons. Firstly, we use the tube current is too low, which could make a loud noise. However, the quality of CT images would be degraded due to excessive x-ray quanta fluctuation and system inherent electronic noise background in the acquired data, if there is no adequate noise treatment in processing the data for image reconstruction. Secondly, the EDEC method should a calibration phantom that must provide path length variations through basis material, and combinations of path lengths through both basis materials, and our path length of water basis material is less than the test object. But path lengths of our method don't have to include it strictly, you just have to approximate it.

VI. Conclusion

In this paper, we use the ANN achieving material decomposition and image refinement by CNN for solving the problem of DSCT reconstruction. Numerical experiments verify that the method can be effectively introduced into the decomposition process. The main advantage of this method is that only need a few phantoms which are simply and without requirement in structure to test object. The size of training objects and test objects should be at the same order of magnitude, but other than that no assumptions on its structure are made.

Our method is a simple and effective way to decompose dual spectral CT. And the method is not only applicable to dual spectral CT but also applicable to multiple spectral CT and multiple basis material decomposition case via adding the neuron of the input or output layer. Furthermore, we can add some preprocessing or post-processing like adding some regularization or using some available information.

References

[1] R. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerized tomography," *Physics in medicine and biology*, vol. 21, pp. 733–44, 10 1976.

[2] L. Yu, S. Leng, and C. McCollough, "Dual-energy ctbased monochromatic imaging," *American Journal of Roentgenology*, vol. 199, pp. S9–S15, 11 2012.

[3] D. Xu, D. Langan, X. Wu, J. Pack, T. Benson, J. Tkaczky, and A. Schmitz, "Dual energy ct via fast kvp switching spectrum estimation," *Proceedings of SPIE - The International Society for Optical Engineering*, 02 2009.

[4] M. Petersilka, H. Bruder, B. Krauss, K. Stierstorfer, and T. Flohr, "Technical principles of dual source ct," *European journal of radiology*, vol. 68, pp. 362–8, 11 2008.

[5] N. Allec, S. Abbaszadeh, and K. Karim, "Single-layer and dual-layer contrast-enhanced mammography using amorphous selenium flat panel detectors," *Physics in medicine and biology*, vol. 56, pp. 5903–23, 08 2011.

[6] T. Szczykutowicz and G.-H. Chen, "Dual energy ct using slow kvp switching acquisition and prior image constrained compressed sensing," *Physics in medicine and biology*, vol. 55, pp. 6411–29, 10 2010.

[7] M. Kachelrieβ, T. Berkus, P. Stenner, and W. Kalender, "Empirical dual energy calibration (edec) for cone-beam computed tomography," vol. 4, 12 2006, pp. 2546 – 2550.

[8] Y. Zhao, X. Zhao, and P. Zhang, "An extended algebraic reconstruction technique (e-art) for dual spectral ct," *IEEE transactions on medical imaging*, vol. 34, 11 2014.

[9] Q. Wang, Y. Zhu, and H. Yu, "Locally linear constraint based optimization model for material decomposition," *Physics in Medicine and Biology*, vol. 62, 09 2017.

[10] G. Wang, "A perspective on deep imaging," *IEEE Access*, 09 2016.

[11] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, 04 2014.

[12] W. Zhao, T. Lv, P. Gao, L. Shen, X. Dai, K. Cheng, M. Jia, Y. Chen, and L. Xing, "A deep learning approach for dual-energy CT imaging using a single-energy CT data," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 11072, 2019, pp. 355 – 359.

[13] Y. Liao, Y. Wang, S. Li, J. He, D. Zeng, Z. Bian, and J. Ma, "Pseudo dual energy ct imaging using deep learning-based framework: basic material estimation," in *Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573, 2018, pp. 1190 – 1194.

Improving the Accuracy and Dose Efficiency of Photon-Counting Spectral CT via Source Spectrum Shaping

Yan Ren, Wenting Long, Huiqiao Xie and Xiangyang Tang[1]

***Abstract*—Recognizing its great potential in material decomposition and virtual monochromatic imaging/analysis, which is the basis for better imaging performance in extensive clinical and preclinical applications, tremendous effort has been being devoted to the research and development of photon-counting spectral CT. In this study, we investigate the shaping of x-ray source spectrum, one of the two ways to optimize spectral channelization in photon-counting spectral CT, and its effect on the performance of material decomposition and virtual monochromatic imaging/analysis. The preliminary data show that significantly better imaging performance and lower x-ray dose can be achieved simultaneously in photon-counting spectral CT via adequate shaping (by additional filiation) of x-ray source spectrum. Hence, it is believed that, in addition to binning of x-ray source spectrum via energy thresholding, the shaping of x-ray source spectrum via additional filtration is another important requirement on system design and implementation of photon-counting spectral CT.**

***Index Terms*— photon counting, spectral CT, photon-counting spectral CT, material decomposition, virtual monochromatic imaging, virtual monochromatic analysis.**

I. Introduction

With increasing value added by energy-integration spectral (dual energy) CT in the clinic, the research and development in photon-counting spectral CT is gaining the momentum [1, 2]. In general, the optimization of spectral channelization, including energy binning and spectral shaping, is essential to the imaging performance of spectral CT. One advantage of photon-counting spectral CT is its capability of spectral channelization via energy binning, rather than the triple scans at different tube peak voltage associated with adequate spectrum shaping (by metal foil additional to x-ray tube's intrinsic filtration [3]) in its energy-integration counterpart. It has been studied and reported in the literature that, in the scenarios where contrast agent (e.g., iodine and gadolinium based) is used for imaging, energy binning, one of the two approaches to optimize spectral channelization, plays an important role in determination of imaging performance in photon-counting spectral CT [4].

As we know, the photons emitting from an x-ray tube used in clinical CT are polychromatic. The vast majority of low energy photons are preferably attenuated along their journey passing through the object to be imaged, which may lead to beam hardening or even photon starvation in data acquisition. This is an issue in the energy-integration (dual energy) spectral CT and may substantially degrade the performance of material decomposition and thus virtual monochromatic imaging/analysis. As demonstrated later in this work, it may become an even larger issue that degrades the performance of spectral imaging in photon-counting spectral CT, since, with increasing number of spectral channels, fewer x-ray photons are allocated in the lowest energy bin. Hence, in this study, we investigate x-ray source spectrum shaping, the other of the two approaches to optimize spectral channelization in photon-counting CT, and its effect on the performance of material decomposition and virtual monochromatic imaging/analysis, aimed at finding the guidelines on system architecture, imaging chain design and performance optimization for advanced clinical applications.

II. Theory & Modeling

A photon-counting detector collects the signal in a number of energy bins at the same time by setting thresholds in energy:

$$I_k(L) = \int_{E=0}^{\mathrm{kVp}} D_k(E)\, N_0(E) \exp\left(-\int_L \mu(E,r)dl\right) dE \quad (1)$$

where L is the path of x-ray beam, $N_0(E)$ is the normalized spectrum in photon counts, and $D_k(E)$ denotes the integrated

[1]All authors are with Imaging and Medical Physics, Department of Radiology and Imaging Sciences, Emory University School of Medicine, 201 Dowman Driv, Atlanta, GA 30322 USA (xiangyang.tang@emory.edu)

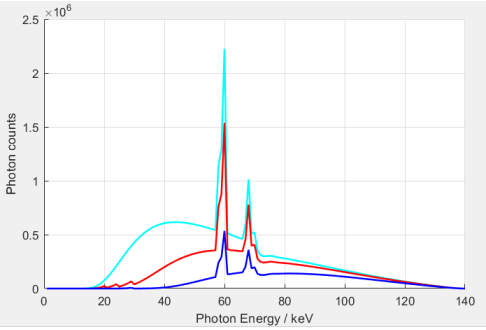


Fig. 1 Original x-ray source spectrum (cyan) and modified versions by Filtration-I (red) and Filtration-II (blue).

effects of detector's efficiency and spectral response corresponding to the k-th energy bin. In this work, we have determined that $k = 1, 2, 3$, i.e., a total of three energy bins are configured in the data acquisition for spectral CT.

On the assumption that the distribution of mass attenuation of a material $\mu(E, r)$ can be decomposed into a linear combination of three basis materials:

$$\mu(E, r) = \sum_{i=1}^{3} a_i(r)\mu_i(E) \tag{2}$$

equation (1) can be rewritten as

$$I_k(L) = \int_{E=0}^{\mathrm{kVp}} D_k(E)\, N_0(E) \exp\left(-\sum_{i=1}^{3} A_i(L)\mu_i(E)\right) dE \tag{3}$$

where $A_i(L) = \int_L a_i(r)dl$ is the line integral of linear combination coefficient $a_i(r)$. In a way similar to the direct approximation method detailed in the literature [5], we model the inverse relation by 3rd-order polynomials to solve the nonlinear system and obtain A_i:

$$\mathrm{A_i} = \alpha_{\mathrm{i},0} + \sum_{\mathrm{k}=1}^{3}\sum_{\mathrm{m}=1}^{3} \alpha_{\mathrm{i,k,m}} \mathrm{I_k^m} \tag{5}$$

Coefficients $\alpha_{\mathrm{i},0}$ and $\alpha_{\mathrm{i,k,m}}$ can be estimated via calibration [5]. Then, the images $a_i(r)$ corresponding to decomposed material can be obtained from $A_i(L)$ using the conventional filtered back-projection [6] or iterative image reconstruction algorithms [7].

III. METHODS & EXPERIMENTS

The investigation is carried out using an in-house simulation software kit of photon-counting spectral CT, in which the system parameters are chosen to mimic the geometry of a clinical MDCT for diagnostic imaging. The photon-counting detector is a two-dimensional curved array. The photon-counting spectral CT is assumed to work at 140 kVp, 250 mA and 1 rot/s gantry rotation speed.

In shaping the source spectrum to remove the x-ray photons at low energy from entering the object to be imaged, two three-layered additional filtrations (Filtration-I and Filtration-II) are designed. Filtration-I is comprised of Mo (0.02 mm), Pd (0.02 mm) and Sn (0.03 mm), to remove the photons at energy lower than 30 keV, while Filtration-II is of Al (2.0 mm), Cu (0.2 mm) and Sn (0.2 mm), to remove the photons at energy lower than 40 keV. The detail of source spectrum shaping is displayed in Fig. 1.

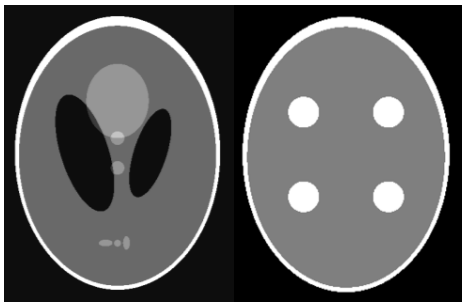


Fig. 2. The original Shepp-Logan phantom (left) and the humanoid head phantom modified from it with four (cortical) bony rods inserted (right).

Energy binning are set by thresholding at different energy levels for data acquisition under different additional filtration (spectrum shaping) such that the photon counts in each energy bin are roughly equal prior to their entering into the object to be imaged. The energy thresholds are set to be [1 51], [52 68] and [69 140] keV for the case of original x-ray spectrum (case0), [1 59], [60 76] and [77 140] keV for Filtration-I (case1), and [1 68], [69 90], [91 140] keV for Filtration-II (case2). For simplicity, the detector's efficiency is assumed ideal (i.e., 100% and no variation over energy).

A humanoid head phantom modified from the Shepp-Logan [8] phantom (Fig. 2) with four (cortical) bony rods in the brain parenchyma is designed to illustrate the adverse effect caused by beam hardening and photon starvation on the accuracy of photon-counting spectral CT. Moreover, the phantom is used to demonstrate the improvement in the performance of photon-counting spectral imaging achieved via adequate source spectrum shaping. The mass attenuation coefficients of the materials in the phantom and their variation over energy are set by consulting authoritative publications, e.g., NIST [9].

IV. RESULTS

In this section, we present and compare the preliminary results corresponding to the three cases defined above.

A. Polychromatic CT images acquired at each individual energy bin

The polychromatic CT images corresponding to each energy bin of the three cases are presented in Fig. 3. Notably, beam hardening artifacts exist in almost all the polychromatic CT images and are much stronger in the ones corresponding to the lowest energy bin. It is also observed that, with increasing number of low energy photons being removed in the lowest

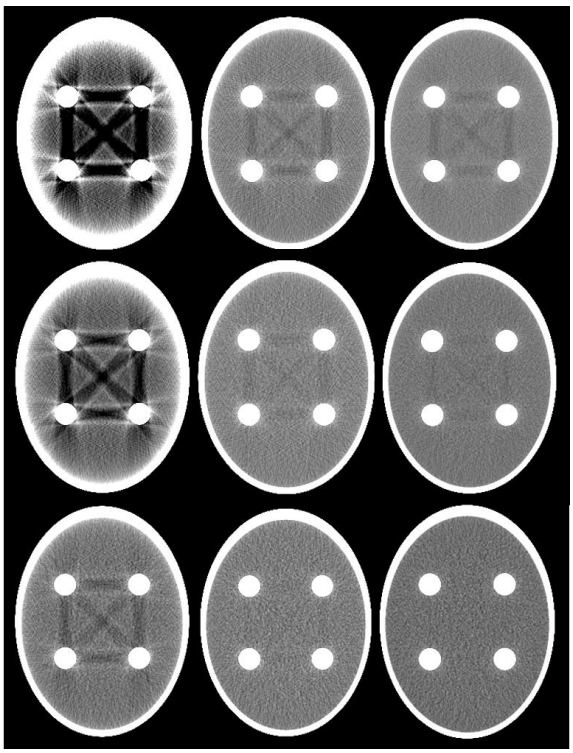


Fig. 3. Polychromatic CT images corresponding to each energy bin (left to right): case0 (top), case1 (middle) and case2 (bottom) (display width: [0 80] HU).

energy bin in case1 and case2, the artifacts caused by beam hardening and photon starvation are significantly reduced.

B. CT images corresponding to decomposed materials

The CT images corresponding to each basis material over the three cases are presented in Fig. 4. Notably, severe artifacts exist in the images corresponding to each decomposed material, due to beam hardening and/or photon starvation (1st and 2nd row). Fortunately, the artifacts (inaccuracy) in those material images can be removed by adequate shaping of x-ray source spectrum and energy binning, as shown in the images at the bottom row of Fig. 4.

C. Virtual monochromatic imaging and analysis

Once material decomposition is carried out, virtual monochromatic image can be formed using equation (3), and then virtual monochromatic analysis can be carried out by varying the energy levels in equation (3). Presented in Fig. 5 are the virtual monochromatic CT images at 40, 70 and 100 *keV*, respectively (10 *mm* slice thickness).

It is observed that, in all the three cases, the virtual monochromatic images at lower energy levels suffer from residual beam hardening artifacts, in addition to higher noise. With more low energy photons being removed from the x-ray beam via adequate additional filtration (source spectrum shaping), the imaging performance (accuracy) of spectral imaging improves dramatically. Particularly, the virtual

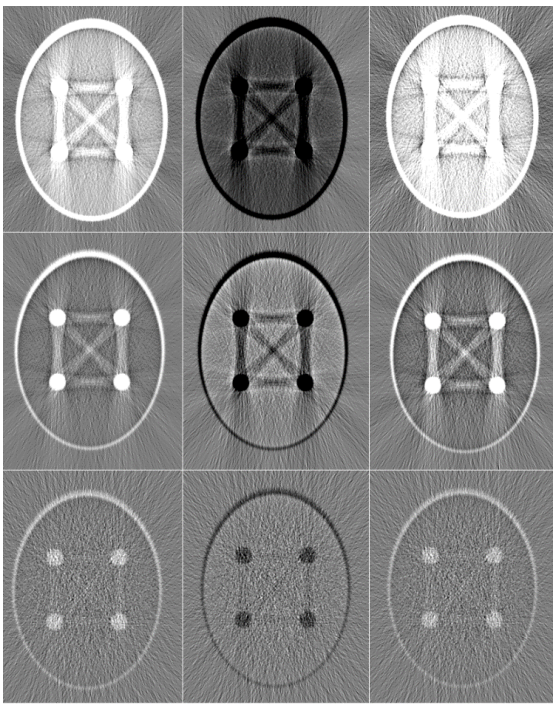


Fig. 4. CT images corresponding to the basis materials (left to right: Teflon, PMMA and LDPE) in case0 (top), case1 (middle) and case2 (bottom) (display consistent with Fig. 3).

monochromatic images corresponding to case2 at 70 and 100 *keV* becomes nearly artifacts free, showing the efficacy of adequate x-ray source spectrum shaping.

V. DISCUSSIONS AND CONCLUSION

In this work, we investigated the shaping of x-ray source spectrum via additional filtration and its effect on the performance of spectral CT imaging, with an emphasis on material decomposition and virtual monochromatic imaging/analysis. As we know, in theory, there should be no artifacts (inaccuracy) left in the images corresponding to the basis materials, and in turn, there should be no artifacts (inaccuracy) in virtual monochromatic imaging/analysis. Unfortunately, however, severe beam hardening or even photon starvation can occur during the data acquisition in the energy bin at lowest energy, leading to severe artifacts (inaccuracy) in material decomposition and then virtual monochromatic imaging/analysis.

Through shaping the x-ray source spectrum via additional filtration, the performance of spectral imaging in photon-counting spectral CT can be improved substantially. The preliminary data obtained thus far show that the removal of low energy photons from the lowest energy bin induces a favorable effect on the accuracy of material decomposition (Fig. 4) and virtual monochromatic imaging/analysis (Fig. 5), which is desirable in the practice for photon-counting spectral CT to meet the challenges in advanced clinical applications. To the best of our knowledge, such an observation on the improvement of accuracy in material decomposition and

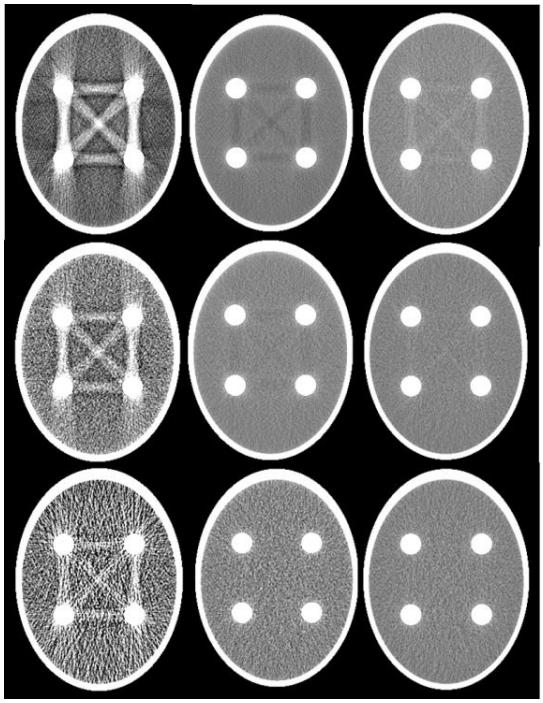


Fig. 5. Virtual monochromatic images at 40 (left), 70 (middle) and 100 (right) keV (top: case0; middle: case1; bottom: case3; display window: [-40 40] HU).

virtual monochromatic imaging/analysis in photon-counting spectral CT reported here is new and of relevance.

Another major benefit of additional filtration for x-ray source spectrum shaping in photon-counting spectral CT is the reduction of x-ray dose, especially the skin dose, to the patients who undertakes the CT scan. This is due to the fact that the vast majority of the low energy x-ray photons have no chance to penetrate the body for image formation are preferably removed from the beam. Notably, such a benefit in dose saving concurs with the improvement in imaging performance, which may promote the clinical utility of photon-counting spectral CT in practice, especially in the scenarios, e.g., pediatric application, wherein the subject undertaking the scan is extremely sensitive to radiation dose.

It should be noted that the energy binning in this study is implemented based on the criteria that the number of x-ray photons emitting from the source are equally allocated over all energy bins. Another design criteria can be that the number of x-ray photons reaching the detector should distribute roughly even over all the energy bins. Apparently, the latter criteria makes more sense from the perspective of system and imaging performance design, but it brings in more factors to take into account in a task-driven manner.

Prior to ending this paper, we would like to indicate that the shaping of x-ray source spectrum via additional filtration, one of the two approaches to optimize the spectral channelization in photon-counting spectral CT, is feasible in practice from the perspectives of system design and implementation of photon-counting spectral CT. We believe that an optimal combination of x-ray source spectrum shaping and energy binning should become a general and standard (necessary) practice in photon-counting spectral CT for advanced clinical and preclinical applications. The data presented in this work is preliminary and more in-depth investigation is under the way in our laboratory. We will promptly update the community with more results obtained from our investigation in the near future.

REFERENCES

[1] C. H. McCollough, and S. Leng, L. Yu, and J. G. Fletcher, "Dual- and Multi-Energy CT: Principles, Technical Approaches, and Clinical Applications," *Radiology*, vol. 276, no. 3, pp. 637-653, Aug. 2015

[2] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Med. Phys.*, vol. 40, no. 10, pp. 100901, Oct. 2013

[3] Primak AN, Ramirez JC, Liu X., Yu L. and McCollough CH, "Improved dual-energy material discrimination for dual-source CT by means of additional spectral filtration," *Med. Phys.*, vol. 36, no. 4, pp. 1359-1369, Mar. 2009

[4] Roessl E. and Proksa R 2007. "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," *Phys. Med. Biol.*, vol. 52, pp. 4679-4695, Jul. 2007

[5] K. Chuang and H. K. Huang, "Comparison of four dual energy image decomposition methods," *Phys. Med. Biol.* vol. 33, no. 4, pp. 455-466, Apr. 1988

[6] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm,", *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612-619, Jun. 1984

[7] L. L. Geyer, U. J. Schoepf, F. G. Meinel, J. W. Nance, G. Bastarrika, J. A. Leipsic, N. S. Paul, M. Rengo, A. Laghi and C. N. De Cecco, "State of the Art: Iterative CT Reconstruction Techniques," *Radiology*, vol. 276, no. 2, pp. 339-257, Jul. 2015

[8] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. 21, no. 3, pp. 21–43, Jun. 1974

[9] J. H. Hubbell and S. M. Seltzer, "Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients (version 1.4)," National Institute of Standards and Technology, Gaithersburg, MD, 2004, Accessed on: Feb. 28, 2019, [Online] Available: http://physics.nist.gov/xaamdi

Phase-contrast x-ray tomography of neuronal tissue with laboratory sources at sub-micron resolution

Marina Eckermann, Mareike Töpperwien, Anna-Lena Robisch, Franziska van der Meer, Christine Stadelmann and Tim Salditt

***Abstract*—Recently, progress has been achieved in implementing phase-contrast tomography of soft biological tissues at laboratory sources. This opens up opportunities for three-dimensional (3d) histology based on x-ray computed tomography (μ- and nanoCT) in direct vicinity of hospitals and biomedical research institutions. Combining advanced x-ray generation and detection techniques with phase-reconstruction algorithms, 3d histology can be obtained even of unstained tissue from the central nervous system, as shown for example for biopsies and autopsies of human cerebellum. Depending on the setup, i.e. source, detector, and geometric parameters, laboratory-based tomography can be implemented at very different sizes and length scales.**

Here, we investigate to which extent 3d histology of neuronal tissue can exploit the cone-beam geometry at high magnification M using a nanofocus x-ray source (Excillum) with $300nm$ minimal spot size, combined with a single-photon counting camera. Tightly approaching the source spot with the biopsy punch, we achieve high $M \approx 10^1$-10^2, high flux density and exploit the superior efficiency of this detector technology.

Different nanotube configurations such as spot size and flux, M as well as exposure time, Fresnel number and coherence are selected and evaluated in view of resolution, field of view, and phase contrast requirements.

Data show that the information content for the cyto-architecture is enhanced by the phase effect. Comparison of results to those obtained at a microfocus rotating-anode x-ray tomography setup with a high resolution detector, i.e. at low-M geometry, reveals similar to slightly superior data quality for the nanotube setup. Besides its compactness, reduced power consumption by a factor of 10^3 and shorter scan duration, the particular advantage of the nanotube setup also lies in its suitability with pixel detector technology, enabling an increased range of opportunities for applications in laboratory phase-contrast x-ray tomography. Further progress is thus likely to bring 3d virtual histology to the performance in scan time and throughput required for clinical practice in neuropathology.

***Index Terms*—Phase contrast tomography, Nanofocus x-ray source, Cone-beam geometry, Neuronal tissue, Human cerebellum**

I. Introduction

PHASE-contrast x-ray tomography offers a unique potential to realize 3d virtual histology, with sub-cellular resolution, and for 3d volumes which are inaccessible by more established techniques. While volume and throughput in conventional histology are limited by slicing and staining, volume penetration for both light- and electron-based microscopy techniques is unsuitable for larger tissue volumes. A substantial amount of propagation-based x-ray phase-contrast tomography studies with synchrotron radiation has demonstrated this potential for soft tissues from the central nervous system (CNS) [1], [2], [3], [4].

Towards a broader accessibility and use of the technique in a clinical setting, translation from synchrotron to laboratory sources is an important goal of ongoing technique and instrumentation development. Laboratory-based phase-contrast μCT has already been implemented exploiting different x-ray generation techniques: transmission microfocus sealed tubes [5], liquid metal-jet anodes [6], [7], as well as microfocus rotating-anode sources [8]. Nanofocus transmission x-ray tubes (nanotubes) offer another opportunity to implement phase-contrast tomography, with possible advantages in view of increased spatial coherence length. While high resolution tomography of different stained tissues has been demonstrated in [9], a particular challenge is to reach sufficient image quality for unstained soft tissues with nanofocus laboratory radiation.

In this work, we investigate the suitability of a home-built setup installed at a nanotube (NanoTube N1 60kV, Excillum AB, Stockholm, Sweden) for 3d histology of unstained *post mortem* human brain tissue. As in our earlier study [3], we chose cerebellar tissue, as a reference structure for phase-contrast tomography data evaluation, due to its well-known anatomy and features covering various relevant length scales and e^--densities. Fig. 1 illustrates its basic anatomical features, as the *white matter* (WM) with *fiber tracts* (FT), the cell-dense *granular layer* (GL) with smaller *granular cells* (GC) and *Golgi cells* (GgC), and *Purkinje cells* (PC, PCL), whose long, planar and widely branched dendritic trees reach into the cell-light *molecular layer* (ML) with its *molecular cells* (MC) in order to place the tomographic results obtained on biopsy punches from paraffin-embedded tissue blocks with typical diameter of a mm into proper anatomical perspective.

II. NanoTube implementation

a) Experimental setup: We have designed a nanofocus x-ray setup for propagation-based x-ray phase-contrast tomography, depicted in Fig. 2. Its central component is the NanoTube N1 60 kV (Excillum AB) x-ray source, with two-dimensional

This abstract is a shortened version of: M. Eckermann et al., "Phase contrast x-ray tomography of neuronal tissue at laboratory sources with sub-micron resolution", Proc. SPIE 111120R, X-Ray Nanoimaging: Instruments and Methods IV, 111130S (2019) and M. Eckermann et al., "Phase contrast x-ray tomography of neuronal tissue at laboratory sources with sub-micron resolution", J. Medical Imaging, (submitted, 2019).

M. Eckermann, M. Töpperwien, A.-L. Robisch and T. Salditt are with the Institute for X-ray Physics, Göttingen University, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany. e-mail: marina.eckermann@phys.uni-goettingen.de.

F. van der Meer and C. Stadelmann are with the Institute for Neuropathology, University Medical Center Göttingen, Robert-Koch-Straße 40, 37099 Göttingen, Germany.

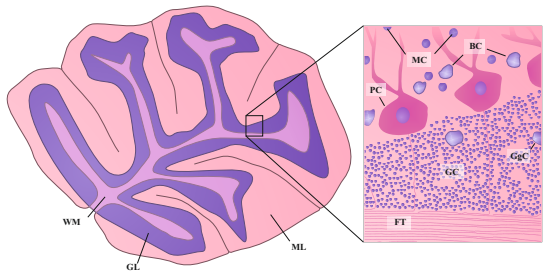


Fig. 1. Depiction of the cerebellum anatomy, giving a rather macroscopic perspective on the left and a zoom-in to the cellular scale on the right.

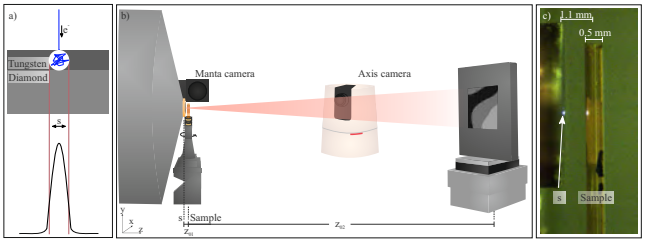


Fig. 2. Illustration of the nanotube setup: a) Schematic of the transmission target X-ray source. b) Recording geometry: the cone-shaped front end of the nanotube allows for small distances z_{01} between source s and the motorized (rot., transl.) sample. The detector is aligned at z_{02} from the source. The axis camera (Axis Communications AB) allows for visual inspection of the setup in general, while the manta camera ()Allied Vision Technologies GmbH) gives a microscopic control of the tight z_{01} environment, see exemplary view in c).

(2d) spatial resolution down to $150nm$. The source is operated at $60keV$, at a power of 0.2-1.2W, depending on the source spot sizes s. The NanoTube system was calibrated for three different s (full width at half maximum, FWHM): "big spot, high flux" at $\sim 1\mu m$, "middle spot, middle flux" at $\sim 0.5\mu m$ and "small spot, low flux" at $\sim 0.3\mu m$ (FWHM). Fig. 2a) shows the schematic of the transmission-anode target ($0.50\mu m$ W-film on a $100\mu m$-diamond layer). As shown in Fig. 2c), samples can be positioned in direct proximity to the target, allowing for source-to-sample distances down to $z_{01} \geq 100\mu m$. Due to the sub-μm spot size s, high geometrical magnification $M = z_{02}/z_{01} \gg 1$ can be realized without source blurring, making it possible to use direct (photon-counting) pixel detectors, available only with relatively large pixel size px. Accordingly, the effective pixel size in the sample plane is reduced to $px_{\text{eff}} = \frac{1}{M} \cdot px = \frac{z_{01}}{z_{02}} \cdot px$.[10] Phase-contrast implementation in free-space propagation-mode requires sufficient degree of coherence to allow for self-interference of the x-ray beam. NanoTube's small s assures the spatial coherence length ζ to be of the same order of magnitude as px_{eff}

$$\zeta = \frac{\lambda \cdot z_{01}}{s} \geq px_{\text{eff}} \ , \tag{1}$$

where λ denotes the x-ray wavelength. A single-photon counting Timepix Hexa H05-W0154 detector (XIE, Freiburg, Germany) was used, with a $500\mu m$ Si-sensor, $px = 55\mu m$ and 768 x 512 pixels (w x h), to record the 2d projections.

b) Data recording settings: For tomographic acquisition, x-ray projections were recorded at 1201 rotation angles, equally distributed over 192°, with sets of 25 flat field images before and after each scan. $z_{02} \approx 20cm$ was kept fixed, varying z_{01} and s. The detector was operated with a lower-limit cut-off energy of $4keV$. The acquisition time was adjusted depending on the spot size s to avoid overexposure. Most scans were split into four consecutive tomographic scans with $\frac{1}{4}$ exposure time each and then recombined. All scan parameters are tabulated in Tab. I. This acquisition scheme was found to reduce ring artifacts arising from a statistically varying response of the modules.

c) Phase retrieval: In contrast to absorption-based CT, phase-contrast is based on the real part of the complex refractive index $n = \delta + i\beta$, and is not directly accessible from the data but requires a reconstruction scheme. The data are recorded in the *direct contrast* or *edge enhancement regime* at Fresnel numbers

$$F = \frac{px_{\text{eff}}^2}{z_{\text{eff}}\lambda} = \frac{px^2}{z_{12}M\lambda} \approx 1, \tag{2}$$

where $px_{\text{eff}} = px/M$ for the effective pixel size, and $z_{\text{eff}} = z_{12}/M$ for the effective propagation distance have been used (Fresnel scaling theorem [10]). In this regime, even data from low coherence sources can be successfully reconstructed based on the *Bronnikov-aided correction* (BAC) scheme as

$$I(\vec{r}_\perp, z = 0) = \frac{I(\vec{r}_\perp, z)}{1 - \gamma\nabla_\perp^2 \tilde{\phi}(\vec{r}_\perp)} [11], \tag{3}$$

where the phase magnitude is estimated via $\tilde{\phi}(\vec{r}_\perp) = 2\pi F \cdot \mathcal{F}_\perp^{-1}\left[\frac{\mathcal{F}_\perp\left[\frac{I(\vec{r}_\perp,z)}{I_0} - 1\right]}{|\vec{k}_\perp|^2 + \alpha}\right]$, [12] with the assumption of small propagation distances z_{eff} and a purely phase shifting object, with intensity $I(\vec{r})$ and k being the wavenumber. The parameter α is introduced to regularize the singularity at zero spatial frequencies, and in practice is chosen such that edge enhancement is canceled. After, the phase-retrieved 2d projection data are processed for ring artifact mitigation according to and recombined to a 3d volume using the ASTRA-toolbox[13]. Visualization was done with Avizo (Thermo Fisher Scientific, Waltham MA).

A. X-ray spectrum and flux assessment

The spectrum of the X-ray source was assessed using an energy-resolving XR-100CdTe detector (Amptek, Bedford). The measured spectrum was corrected for air absorption and detector sensitivity, all plotted in Fig. 3. The blue curve represents the mean photon energy for the tomography data, with a mean photon energy of $E_{\text{w}} = 9.11keV$.
The photon flux is dependent of s. For $s = 0.95\mu m$, it was found to be in the order of $\sim 2 \cdot 10^{11}$ph/s in 2π-space, $\sim 4 \cdot 10^{10}$ph/s for $s = 0.51\mu m$ and $\sim 1 \cdot 10^{10}$ph/s for $s = 0.30\mu m$.

B. Phase-contrast tomography of unstained human brain tissue

In this work, paraffin-embedded human cerebellum was used to investigate suitability of the setup for 3d virtual

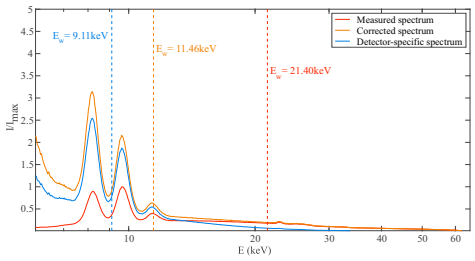


Fig. 3. Plot of the source spectrum: (red) measured data (detector output), (yellow) corrected for absorption in air, and (blue) the spectrum taking into account the energy dependence in the quantum efficiency of the detector used for the tomographic scans.

TABLE I

PARAMETERS AND QUALITY MEASURED FOR THE TOMOGRAPHIC SCANS AT FIXED F: (TOP) EXPERIMENTAL PARAMETERS, (BOTTOM) QUALITY MEASURES. EACH FIRST TO ROWS REFER TO II-B1, ROWS 3-6 TO II-B2 AND EACH LAST ROW PRESENTS THE DATA QUALITY FROM THE ROTATING ANODE SCAN. EXPOSURE TIMES ALSO INDICATE WHETHER THE SCAN WAS PERFORMED AS A SINGLE OR MULTIPLE ONES. SCAN TIME REFERS TO THE FULL SCAN. *Fourier-shell correlation* (FSC)[14] IS BASED ON VOLUMES OF 400^3 VOXELS. THE SIGNAL-TO-NOISE RATION (SNR) WAS CALCULATED AS $(\mu_{FT} - \mu_{BG})/\sigma_{BG}$.

s (μm)	F	z_{01} (mm)	px_{eff} (μm)	Exp. time (s)	Scan time (h)	Dose (kGy)
0.95	1.97	3.5	0.959	$1 \cdot 3.5 = 3.5$	2.0	10.6
0.51	1.96	3.5	1.067	$4 \cdot 5 = 20$	8.0	14.3
0.95	1.97	3.5	0.959	$4 \cdot 3.5 = 14$	8.0	40.5
0.51	1.40	2.5	0.687	$4 \cdot 5 = 20$	8.0	26.0
0.30	0.60	1.1	0.300	$4 \cdot 20 = 80$	32.5	189.2
70.0	1.24	494.0	1.070	$1 \cdot 50 = 50$	16.7	

s (μm)	F	FSC (μm)	SNR GCL	SNR PC nucleus	SNR PC body
0.95	1.97	0.12	5.46	5.49	0.74
0.51	1.96	0.10	6.39	9.25	2.74
0.95	1.97	1.00	6.47	5.93	0.86
0.51	1.40	1.13	6.18	6.91	1.24
0.30	0.60	1.02	3.82	3.81	0.83
70.0	1.24	1.82	5.36	6.80	2.55

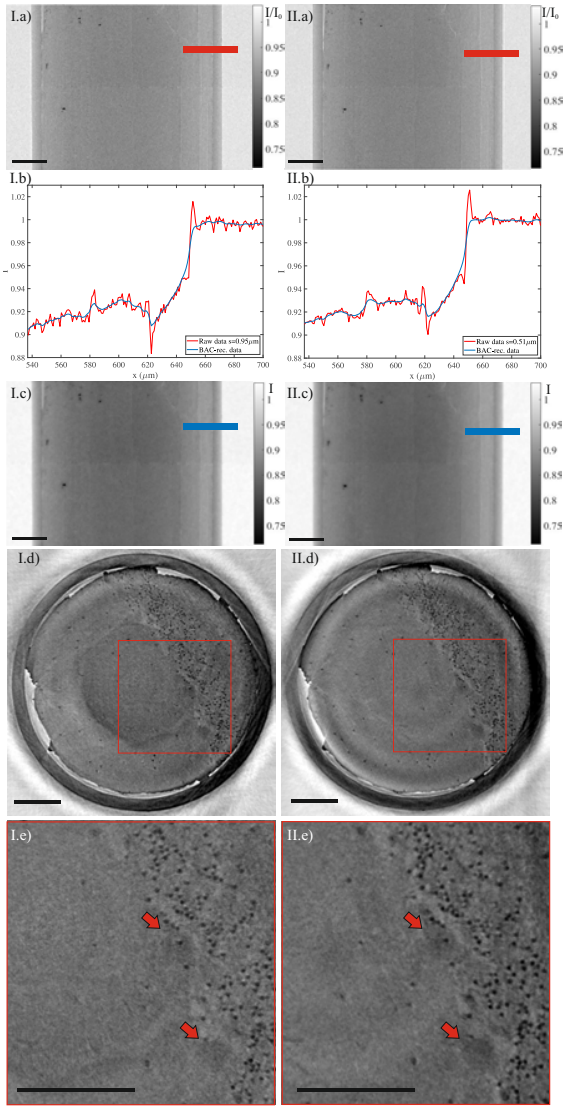


Fig. 4. Data from tomography scans at fixed Fresnel number of 2.0, at (I) $s = 0.95\mu m$ ($\zeta = 0.501\mu$m) and (II) 0.51μm ($\zeta = 0.934\mu$m). (a) Respective flat-field corrected projections and (c) BAC-reconstructed projections. Colored bars indicate the position of the profiles shown in (b). The same virtual slice (d) in the xz-plane with a zoom-in in (e). Red arrows mark PCs. Data from quantitative analysis are summarized in table I. Scale bars: 100μm.

histology and neuropathology. From the tissue block, a 0.5mm biopsy punch was extracted and transferred into a 0.5mm polyimide tube, as illustrated in Fig. 2.

1) Variation of s/ζ at constant F : We have first investigated the influence of the lateral coherence length ζ on the image quality, by setting two different spot sizes (I) $s = 0.95\mu m$ and (II) $s = 0.51\mu m$, respectively, while keeping $F \approx 2.0$ constant, and the dose also approximately to the same level as given in Tab. I. Hence the larger source size resulted in a significantly reduced total scan time. Fig. 4 illustrates data reconstruction steps, from projection to virtual slices through the reconstructed volume. The projections in (a) reveal slight edge enhancement in both cases, especially for the polyimide-air-interface, as plotted in (b), being increased at $s = 0.501\mu m$, providing a higher degree of coherence. Virtual slices through the same position in the xz-plane and xy-plane are shown in (d), respectively, with a corresponding zoom shown in (e). Based on visual inspection, both data sets appear to be of very similar quality, with case (I) delivering superior feature contrast. Different cerebellum-specific regions as outlined in Fig. 1 are clearly identifiable: as for the zoom-ins (I.e) and (II.e), the cell-dense GL on the right and the ML on the left are separated by the sparsely-distributed, bold PCL-cells (red arrows).

2) Variations of F at constant source spot size s : Next we investigated the influence of F on tomographic image quality as being another crucial parameter in phase contrast formation, controlled by z_{01}, while keeping z_{02} fixed. px_{eff} (Eq. 2) and ζ (Eq. 1) vary accordingly. Scans were recorded at $s = 0.30\mu m$, $0.51\mu m$ and $0.95\mu m$. Parameters for selected scans are detailed in Tab. I. Note the 3d resolution at $s = 0.95\mu m$ being close to the voxel size. Fig. 5 shows the virtual slices along the xy-plane through the reconstruction volume, in similar positions. The SNR also increased for GCs and PC nuclei, and slightly for PC bodies. As an overall trend, reducing F (via z_{01}) results in increased resolution, which is expected

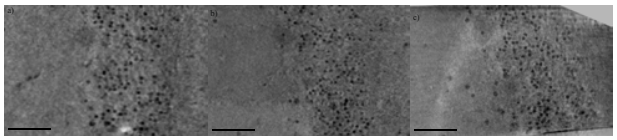


Fig. 5. Virtual xy-slices through the tomographic reconstruction, at similar positions in the biopsy punch, for (a) spot size of 0.95μm, (b) 0.51μm and (c) 0.30μm and Fresnel numbers of (from (a) to (c)) 1.97, 1.40 and 0.60. Note the rise in geometrical magnification from (a) to (c). The respective analysis parameters are given in table I. Scale bars: 50μm.

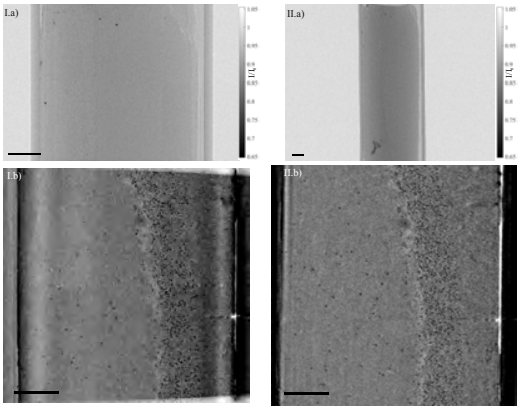


Fig. 6. Comparison of (I) nanotube tomography with (II) microfocus rotating-anode results. Settings were chosen such that $\text{px}_{\text{eff}} \approx 1.07\mu$m same for both data sets. (a) Flat-field corrected projections. Virtual slices through the same position in the reconstructed sample volume are shown in (b) the xy-plane. Scale bars: 100μm.

based on higher dose (smaller px_{eff}). At the same time, despite rise in dose, the SNRs are lower, indicating that the increase in coherence is more important, in particular for intrinsically low-contrasted features such as PC bodies.

3) Comparison with data from microfocus laboratory setup: To put the results into perspective with earlier laboratory implementations of 3d virtual histology, Fig. 6 and Tab. I present a comparison to a reconstruction obtained at a microfocus rotating-anode x-ray source-based setup[8]. For the comparison, the nanotube dataset recorded at $s = 0.51\mu m$ and $F = 1.96$ was selected, serving a similar px_{eff} and coherence. The nanotube setup achieves slightly higher resolution and increased SNRs within only half of the scan time, which however comes at the cost of reduced FOV. (b) shows virtual slices through the xy-plane to judge image quality by visual inspection.

III. Conclusion

In summary, we have successfully demonstrated phase-contrast tomography of unstained neuronal tissue, using a home-built laboratory nanotube-based setup and a photon counting pixel detector. Sufficient image quality for the detection of neurons and hence representation of the cyto-architecture was achieved. In particular, FSC analysis indicated a resolution of $0.90\mu m$ for a source setting of $s = 0.30\mu m$. The phase retrieval and reconstruction scheme presented here mixes amplitude and phase information, but is very suited to visualize the small electron density differences in unstained tissue and hence the cyto-architecture, for example of neuronal tissue. The information content for the cyto-architecture is enhanced by the phase effect, as evidenced by the variation of F. Compared to a microfocus rotating-anode setup, data quality and information content appear to be on a similar level. However, the particular advantage of the nanotube setup (apart from its more compact size and much reduced power consumption) is in its scalability with respect to progress in pixel detector technology. While already performing on the same level for a relatively thin silicon sensor, future replacement by $1mm$ thick Si sensors or even $GaAs$ sensors, along with an increase in the detection panels may result in significantly reduced scan times. We also note that new pixel detector technologies allow for registration of counting with sub-pixel registration and for photon energy determination, opening up an entirely new opportunity for laboratory phase contrast tomography. Further progress is hence likely to bring virtual 3d histology to the performance in scan time and throughput required for clinical practice in neuropathology, combined with automated evaluation of cyto-architecture as in [3]. Importantly, the demonstrated laboratory-based image quality would already be sufficient for this task.

References

[1] M. Bartels, M. Krenkel, P. Cloetens, W. Möbius, and T. Salditt, "Myelinated mouse nerves studied by x-ray phase contrast zoom tomography," *J. Struct. Biol.*, vol. 192, 2015.

[2] M. Töpperwien, A. Markus, F. Alves, and T. Salditt, "Contrast enhancement for visualizing neuronal cytoarchitecture by propagation-based x-ray phase-contrast tomography," *NeuroImage*, vol. 199, 2019.

[3] M. Töpperwien, F. Van der Meer, C. Stadelmann, and T. Salditt, "Three-dimensional virtual histology of human cerebellum by X-ray phase-contrast tomography," *PNAS*, vol. 115, no. 27, 2018.

[4] A. Khimchenko, H. Deyhle, G. Schulz, G. Schweighauser, J. Hench, N. Chicherova, C. Bikis, S. E. Hieber, and B. Müller, "Extending two-dimensional histology into the third dimension through conventional micro computed tomography," *NeuroImage*, vol. 139, 2016.

[5] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, "Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources," *Nat. Phys.*, vol. 2, no. 4, 2006.

[6] M. Bartels, V. H. Hernandez, M. Krenkel, T. Moser, and T. Salditt, "Phase contrast tomography of the mouse cochlea at microfocus x-ray sources," *Appl. Phys. Lett.*, vol. 103, no. 8, 2013.

[7] M. Krenkel, M. Töpperwien, M. Bartels, P. Lingor, D. Schild, and T. Salditt, "X-ray phase contrast tomography from whole organ down to single cells," *Proc. SPIE*, vol. 9212, 2014.

[8] M. Reichardt, J. Frohn, M. Toepperwien, J.-D. Nicolas, A. Markus, F. Alves, and T. Salditt, "Nanoscale holographic tomography of heart tissue with x-ray waveguide optics," *Proc. SPIE 10391*, 2017.

[9] M. Müller, I. de Sena Oliveira, S. Allner, S. Ferstl, P. Bidola, K. Mechlem, A. Fehringer, L. Hehn, M. Dierolf, K. Achterhold, B. Gleich, J. U. Hammel, H. Jahn, G. Mayer, and F. Pfeiffer, "Myoanatomy of the velvet worm leg revealed by laboratory-based nanofocus X-ray source tomography," *PNAS*, 2017.

[10] D. M. Paganin, *Coherent X-Ray Optics*. N.Y.: Oxford University, 2006.

[11] Y. D. Witte, M. Boone, J. Vlassenbroeck, M. Dierick, and L. V. Hoorebeke, "Bronnikov-aided correction for x-ray computed tomography," *J. Opt. Soc. Am. A*, vol. 26, no. 4, 2009.

[12] A. Groso, R. Abela, and M. Stampanoni, "Implementation of a fast method for high resolution phase contrast tomography," *Opt. Express*, vol. 14, no. 18, 2006.

[13] W. J. Palenstijn, K. J. Batenburg, and J. Sijbers, "The astra tomography toolbox," in *13th International Conference on Computational and Mathematical Methods in Science and Engineering. CMMSE*, 2013.

[14] G. Harauz and M. van Heel, "Exact filters for general geometry three dimensional reconstruction," in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conf*, vol. 73, 1986.

Dose reconstruction in dental CBCT

Jinwoo Kim and Ho Kyung Kim

***Abstract*—A numerical algorithm for the estimation of the patient-specific absorbed dose distributions in cone-beam computed tomography is introduced. The developed dose estimation algorithm is based on the ray-tracing techniques and requires the reconstructed voxel data in values of linear attenuation coefficients and the scanning protocol. The algorithm can be divided into two parts. First, the algorithm calculates the absorbed dose due to the primary photons by considering beam attenuation along the beam path between the source and each reconstructed voxel in conjunction with the solid angle subtended by given voxel. Then, this primary dose distribution becomes the source for the dose distribution due to the scattered photons. For the pre-calculated primary dose value in a given voxel, the scatter dose values to all the other voxels are similarly calculated as the primary dose. The developed algorithm shows a good agreement with the Monte Carlo (MC) simulation for the anthropomorphic phantom. The accuracy of the analytical method is investigated by comparing results with the MC estimates and the acceleration strategy using deep-learning techniques is also discussed.**

I. Introduction

SINCE its introduction in 1998[1], dedicated dental cone-beam computed tomography (CBCT) examinations are being greatly increased for not only therapeutic but also cosmetic purposes[2], [3]. As a result, there is an increasing call for radiation dose tracking from medical examinations and procedures and patient-specific dose management has become a great concern[4], [5] . In this regard, it is highly desirable that the radiation dose and potential cancer risk associated with each CT scan of each patient be reported and documented to manage the patient dose in CT examinations[6]. However, the average dose measurement with a standard-size cylindrical phantom is the only method for the practical dose estimation in CT examinations. This standard CT dose index[7] (CTDI) method is aimed at quality control of the CT scanners and is not suitable for patient-specific dose estimations[8], [9].

One of the other alternatives to estimate patient-specific dose distributions is the Monte Carlo (MC) method, but it requires extremely high computational cost, although they can provide accurate estimates[10], [11], [12]. The other methods for estimating patient-specific dose generally employ several approximations due to the complexity of photon transport in heterogeneous media. For example, the methods based on MC pre-calculated dose-spread or electron-spread kernels which employ the convolution/superposition technique approximately account for the inhomogeneity of the human body[13], [14], [15]. The deterministic absorbed dose estimation method which numerically solves the Boltzmann transport equation using discrete ordinates methods[16] has also been investigated by Norris et al.[17], but they only considered the standard CTDI phantom, which cannot account for inhomogeneity. Inspired by this investigation, the authors have developed a numerical approach to estimate the patient-specific dose distributions in the CBCT. The algorithm is fundamentally based on the ray-tracing and deep-learning technique and the dose distributions are estimated in a two-step process: the absorbed dose due to the primary photons first and then that due to the scattered photons, as demonstrated in figure 1.

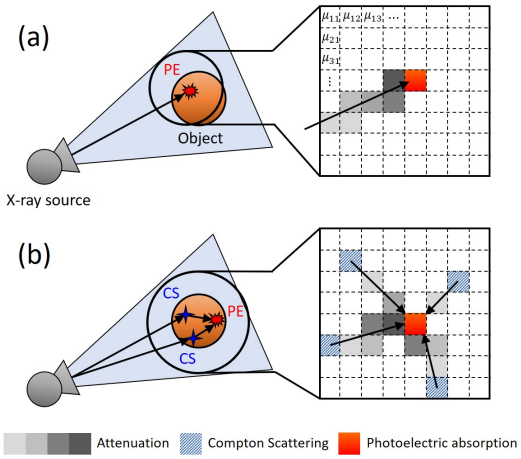


Figure 1. Schematic illustrations of the algorithm for the CT-dose map computation: (a) the primary and (b) the scatter dose calculations.

II. Ray-tracing approach

The patient dose is calculated in a two-step process. First, the algorithm numerically calculates the radiation dose due to the primary x-ray photons. The primary x-ray photon can be defined as any x-ray photon which is produced from the x-ray source and the first photoelectric absorption of that photons leads to the primary dose. Once the primary dose distribution is calculated, it becomes the source causing the additional dose deposition due to the scattered x-ray photons. The scatter dose refers to the dose due to the radiation that has undergone at least one scattering event. The x-ray photon interactions considered in this study for dose calculations are briefly showed in figure 2.

As shown in figure 1(a), the primary dose is estimated accounting for the attenuation along the ray. The primary

This work was supported by the National Research Foundation of Korea (NRF) through the Korean government (MSIP) under Grant 2017M2A2A6A01071267. *H. K. Kim is the corresponding author for this article.

J. Kim is with the School of Mechanical Engineering, Pusan National University, Busan 46241, South Korea.

*H. K. Kim is with the School of Mechanical Engineering and the Center for Advanced Medical Engineering Research, Pusan National University, Busan 46241, South Korea (e-mail: hokyung@pusan.ac.kr).

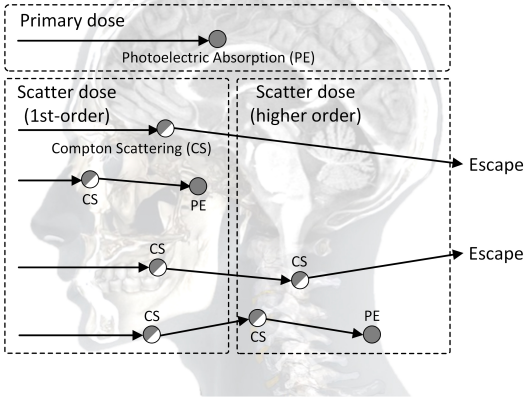


Figure 2. X-ray photon interactions considered for dose calculations.

absorbed energy in a voxel $\mathbf{r}$ is given by:

$$\varepsilon_{prim}(\mathbf{r}) = \int a(\mathbf{r})\phi\Omega(\mathbf{r};\mathbf{r}_0)e^{-\boldsymbol{\mu}^{\mathbf{T}}\mathbf{t}}(1-e^{-\mu(\mathbf{r})t(\mathbf{r})}) \frac{\mu_{pe}}{\mu}(\mathbf{r})E_0 d\beta. \quad (1)$$

The integral in (1) is over the source locations as the source rotates around the patient (i.e., $\mathbf{r}_0$ depends on the β). The terms $a(\mathbf{r})$ and ϕ in (1) refer to the projected area of the voxel $\mathbf{r}$ and the photon fluence at the source, respectively. The $\Omega(\mathbf{r};\mathbf{r}_0)$ term in (1) is the fractional area subtended by the emitted photon by voxel $\mathbf{r}$. The next terms $(e^{-\boldsymbol{\mu}^{\mathbf{T}}\mathbf{t}})$ and $\{1-e^{-\mu(\mathbf{r})t(\mathbf{r})}\}$ in (1) refer to the attenuation of voxels along the beam path $(\mathbf{r}_0 \rightarrow \mathbf{r})$ and the absorption of the passing x-ray beam in the voxel $\mathbf{r}$ itself, respectively. The latter reflects the fact that this voxel is absorbing some of the x-ray energy that is passing through it. The remaining terms $\mu_{pe}/\mu(\mathbf{r})$ and E_0 in (1) account the probability of photoelectric absorption and the incident x-ray energy, respectively.

For the pre-calculated primary dose value in a given voxel, the scatter dose values to all the other voxels are similarly calculated as the primary dose. Since the scattered x-ray photon is generally redirected from its original trajectory and undergoes subsequent interactions within the human organs, each time partially depositing its energy, the scatter dose can be divided into two parts: the doses by the scattering interaction itself and by the subsequent interactions of the corresponding scattered x-ray photons. First, the local deposition energy by the scattering interaction itself in a voxel $\mathbf{r}$ is given by:

$$\varepsilon^{l}_{scat}(\mathbf{r}) = \varepsilon_{prim}(\mathbf{r})\frac{\mu_{cs}}{\mu_{pe}}(\mathbf{r})f(E_0). \quad (2)$$

The $f(E_0)$ term in (2) considers the relative amount of the energy partially deposited by the scattering interaction from the incident x-ray energy E_0. Then, as shown in Fig. 1(b), the remote deposition energy by the photoelectric absorption of the scattered photons in a voxel $\mathbf{r}$ can be computed by accumulating the contribution of scattered photons given off from all the other voxels:

$$\varepsilon^{r}_{scat}(\mathbf{r}) = \int \varepsilon_{prim}(\mathbf{r}')\frac{\mu_{cs}}{\mu_{pe}}(\mathbf{r}')[1-f(E_0)]\Omega(\mathbf{r};\mathbf{r}') e^{-\boldsymbol{\mu}^{\mathbf{T}}\mathbf{t}}(1-e^{-\mu(\mathbf{r})t(\mathbf{r})})\frac{\mu_{pe}}{\mu}(\mathbf{r})d^3\mathbf{r}'. \quad (3)$$

The integral in (3) represents the fact that the scatter contribution from all surrounding voxels is considered to compute the scattered dose in the voxel $\mathbf{r}$. The product of the first three terms in (3), $\{\varepsilon_{prim}(\mathbf{r}')\frac{\mu_{cs}}{\mu_{pe}}(\mathbf{r}')[1-f(E_0)]\}$, describes the energy of the scattered photon given off from the voxel $\mathbf{r}'$. The other terms in (3) are similar to those for primary dose calculation in (1).

To achieve higher accuracy, the scatter contribution of higher-order scattering interactions $\varepsilon^{l}_{scat,i}(\mathbf{r})$ may also be calculated by repeating the calculation described in (2), replacing the $\varepsilon_{prim}(\mathbf{r})$ and $f(E_0)$ terms with the $\varepsilon^{r}_{scat,i-1}(\mathbf{r})$ and $f(E_{i-1})$ terms, respectively:

$$\varepsilon^{l}_{scat,i}(\mathbf{r}) = \varepsilon^{r}_{scat,i-1}(\mathbf{r})\frac{\mu_{cs}}{\mu_{pe}}(\mathbf{r})f(E_{i-1}). \quad (4)$$

Similarly, the absorbed energy by the photoelectric absorption of the corresponding scattered x-ray photons $\varepsilon^{r}_{scat,i}(\mathbf{r}_v)$ in the voxel $\mathbf{r}$ is given by:

$$\varepsilon^{r}_{scat,i}(\mathbf{r}) = \int \varepsilon^{r}_{scat,i-1}(\mathbf{r}')\frac{\mu_{cs}}{\mu_{pe}}(\mathbf{r}')(1-f(E_{i-1}))\Omega(\mathbf{r};\mathbf{r}') e^{-\boldsymbol{\mu}^{\mathbf{T}}\mathbf{t}}(1-e^{-\mu(\mathbf{r})t(\mathbf{r})})\frac{\mu_{pe}}{\mu}(\mathbf{r})d^3\mathbf{r}'. \quad (5)$$

The total absorbed dose in the voxel $\mathbf{r}$ is given by dividing the summation of the absorbed energies by the mass:

$$D_{total}(\mathbf{r}) = \frac{\varepsilon_{prim}(\mathbf{r}) + \sum_{i=1}^{N}\{\varepsilon^{l}_{scat,i}(\mathbf{r}) + \varepsilon^{r}_{scat,i}(\mathbf{r})\}}{\rho V(\mathbf{r})}, \quad (6)$$

where the terms N and $\rho V(\mathbf{r})$ in (6) refer to the number of the scatter dose calculations and the mass of the voxel $\mathbf{r}$, respectively.

Table I
SUMMARY OF THE DATA SETS USED FOR DEEP LEARNING. FOV DENOTES THE FIELD OF VIEW.

Number	FOV	Gender	kVp
1	Dentoalveolar	Male	100
2			120
3		Female	100
4			120
5	Maxillofacial	Male	100
6			120
7		Female	100
8			120

III. DEEP-LEARNING APPROACH

To calculate the scatter dose distribution with the ray-tracing based algorithm described in section II, about N^6 times of ray-tracing operations are required where N is the number of voxels of the patient model. It is a time-consuming task and not suitable for real-time dose management. Therefore, we used the deep-learning technique to accelerate the

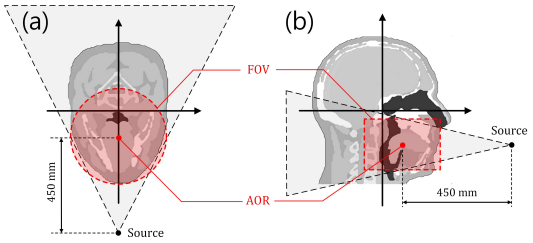


Figure 3. The CBCT geometry for the validation of the developed algorithm: (a) top view and (b) side view. FOV and AOR denote the field of view and the axis of rotation, respectively.

scatter dose calculations. The U-net[18] is employed as a network for calculating the scatter dose distributions from the primary dose distribution. We acquired eight different sets of absorbed dose distributions including primary and scatter dose distributions using commercial MC radiation transport simulation tool (MCNP version 5, RSICC, Oak Ridge, TN). A detailed description of the data sets is summarized in table I. The attenuation map and the corresponding primary dose distributions were inputted to the networks for training, and the corresponding scatter or total dose distributions were used as target data for indirect or direct applications, respectively. In the case of indirect application, we obtained total dose distributions adding the scatter and primary dose distributions.

Because of a limited number of data sets, the attenuation map and the MC pre-calculated dose distributions were split into small-sized sub-patches. The patches have a dimension of $32 \times 32 \times 32$ pixels and the total number of patches was $\sim 10^5$ considering data augmentation by rotating images as well as flipping images side to side. Six sets of absorbed dose distributions were used for training and the other sets were used for validation. For training, the batch size was chosen to be 2 and the Adam (adaptive moment estimation) optimizer was adopted for a stochastic optimization with specific parameters: the first and second moments of the gradients of 0.9 and 0.999, respectively, and the epsilon of 10^{-8}. The learning rate was initially set as 10^{-4}.

IV. Validation

The accuracy of the algorithm is investigated by comparing estimates with the Monte Carlo (MC) estimates. The head-and-neck part of the anthropomorphic phantom (XCAT version 2.0, Duke University, Durham, NC) is used to validate the developed algorithm. As shown in figure 3, the field of view (FOV) has a size of 130×80 mm and is located in the maxilla and mandible. The distance from the source to the axis of rotation (AOR) is 450 mm and the 100 kVp x-ray spectrum which is produced by a tungsten target from the TASMIP model [19] is used at every 1 degree while the source rotates entire 360 degrees.

V. Results

Figure 4 shows the interaction-wise absorbed dose spatial distributions in the XCAT phantom. As shown in figure 4(a) and (b), the dose distributions by the primary photons are

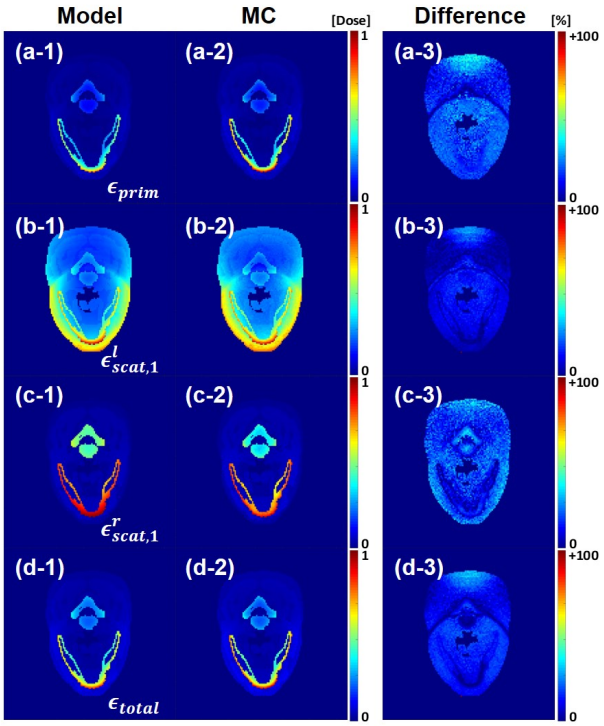


Figure 4. The spatial distributions of the absorbed dose in the anthropomorphic phantom. The rows from top to bottom represent the primary, local scatter, remote scatter, and the total doses, respectively, and the columns from left to right represent the results from developed algorithm, MC simulation, and their relative difference, respectively.

matched to the MC result with mean voxel error of 14.0%. The developed algorithm underestimates the remote dose deposition by the scattered photons with the mean voxel error of 15.9% and the total absorbed dose distribution shows mean voxel error of 13.0% compared to the MC results.

Figure 5 compares the total absorbed dose distributions obtained from the U-net, which were trained with MC estimates as targets. The U-net works for estimating scatter and total dose distributions from primary dose distributions, although those accuracies are not enough to be substantially comparable to the MC estimates. Figure 5 also compares the performances between indirect and direct applications of the U-net. The results obtained from the scatter-dose training are shown in the first two rows, and those from the total-dose training are shown in the third and fourth rows. Since the scatter-dose training network produces a scatter dose distribution, the total dose distribution is obtained by adding the scatter and primary dose distributions (i.e., input). As shown in figure 5, the indirect application shows a better performance in the region that directly irradiated by the x-ray beam. Reversely, the direct application shows better performance in the brain region, which is not included in the FOV.

VI. Summary and Future Work

The absorbed dose distributions have been obtained using analytic algorithms based on the ray-tracing technique for anthropomorphic phantoms, and the proposed algorithm showed good agreement with the MC method. Compared to the MC

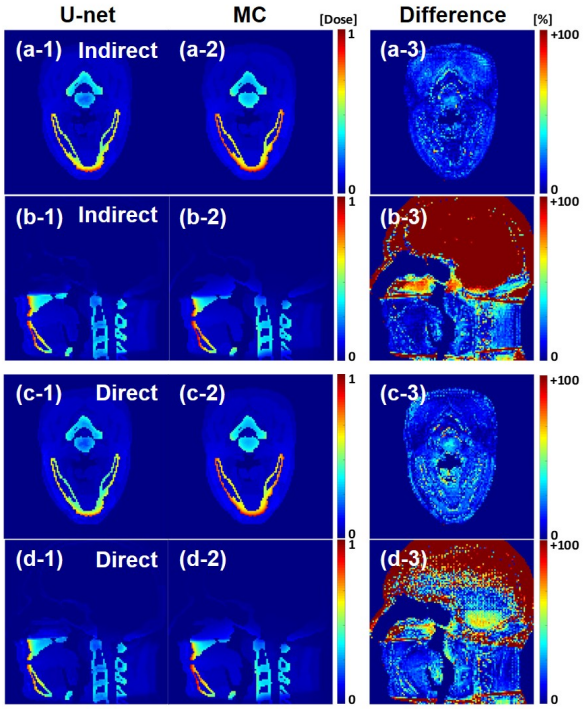


Figure 5. Comparison of absorbed dose distributions obtained from the U-net. The first two rows represent the axial and sagittal slices of the absorbed dose distributions obtained from the indirect application (scatter-dose training) respectively and the last two rows show those from the direct application (total-dose training). The columns from left to right represent the output from the U-net, target (MC estimate), and their relative difference, respectively.

estimates, the algorithm underestimates the absorbed doses with the mean voxel error of 7.8% but takes approximately 1 hour to compute for $(128 \times 128 \times 128)$ voxels and FOV considered in this study with 3.6-GHz single-core CPU and Nvidia Geforce GTX 1080 Ti graphics cards. The deep-learning-based approach for the acceleration of the scatter-dose calculation have also been introduced, and it takes only 30 seconds on the same hardware. However, compared to the MC estimates, the absorbed dose distributions obtained from the U-net show the large errors in the regions that are not directly exposed by the x-ray beam, as shown in figure 5. This large error may become from the fact the primary dose distributions do not have a value in the regions that x-rays are not directly exposed. The lack of training data may also degrade the accuracy of the resultant dose distributions although the U-net was designed for working with very few training data[18]. To improve the accuracy of the deep-learning-based scatter-dose estimation, our future work will be to obtain additional training data which include absorbed dose distributions for FOVs of various sizes and positions.

References

[1] P. Mozzo, C. Procacci, A. Tacconi, P. Tinazzi Martini, and I. A. Bergamo Andreis, "A new volumetric CT machine for dental imaging based on the cone-beam technique: preliminary results," *Eur. Radiol.*, vol. 8, no. 9, pp. 1558–1564, Nov 1998. [Online]. Available: https://doi.org/10.1007/s003300050586

[2] W.C. Scarfe, A.G. Farman and P. Sukovic, "Clinical applications of cone-beam computed tomography in dental practice," *J. Can. Dent. Assoc.*, vol. 72, no. 1, pp. 75–80, 2006.

[3] M. Ahmad, J. Jenny, and M. Downie, "Application of cone beam computed tomography in oral and maxillofacial surgery," *Aust. Dent. J.*, vol. 57, no. s1, pp. 82–94, 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1834-7819.2011.01661.x

[4] D. J. Brenner and E. J. Hall, "Computed tomography - an increasing source of radiation exposure," *New Eng. J. Med.*, vol. 357, no. 22, pp. 2277–2284, 2007. [Online]. Available: http://dx.doi.org/10.1056/NEJMra072149

[5] E. S. Amis, P. F. Butler, K. E. Applegate, S. B. Birnbaum, L. F. Brateman, J. M. Hevezi, F. A. Mettler, R. L. Morin, M. J. Pentecost, G. G. Smith, K. J. Strauss, and R. K. Zeman, "American college of radiology white paper on radiation dose in medicine," *J. Am. Coll. Radiol.*, vol. 4, no. 5, pp. 272–284, 2007.

[6] X. Li, E. Samei, W. P. Segars, G. M. Sturgeon, J. G. Colsher, G. Toncheva, T. T. Yoshizumi, and D. P. Frush, "Patient-specific radiation dose and cancer risk estimation in CT: Part II. Application to patients," *Med. Phys.*, vol. 38, no. 1, pp. 408–419, 2011. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3515864

[7] T. B. Shope, R. M. Gagne, and G. C. Johnson, "A method for describing the doses delivered by transmission x-ray computed tomography," *Med. Phys.*, vol. 8, no. 4, pp. 488–495, 1981. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.594995

[8] D.J. Brenner, "Is it time to retire the CTDI for CT quality assurance and dose optimization?" *Med. Phys.*, vol. 32, pp. 3225–3226, 2005.

[9] W. A. Kalender, "Dose in x-ray computed tomography," *Phys. Med. Biol.*, vol. 59, no. 3, p. R129, 2014.

[10] K. Perisinakis, A. Tzedakis, and J. Damilakis, "On the use of Monte Carlo-derived dosimetric data in the estimation of patient dose from CT examinations," *Med. Phys.*, vol. 35, no. 5, pp. 2018–2028, 2008. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.2896075

[11] P. Deak, M. van Straten, P. C. Shrimpton, M. Zankl, and W. A. Kalender, "Validation of a Monte Carlo tool for patient-specific dose simulations in multi-slice computed tomography," *Eur. Radiol.*, vol. 18, no. 4, pp. 759–772, Apr 2008. [Online]. Available: https://doi.org/10.1007/s00330-007-0815-7

[12] M. Myronakis, K. Perisinakis, A. Tzedakis, S. Gourtsoyianni, and J. Damilakis, "Evaluation of a patient-specific Monte Carlo software for CT dosimetry," *Radiat. Prot. Dosim.*, vol. 133, no. 4, pp. 248–255, 03 2009. [Online]. Available: https://doi.org/10.1093/rpd/ncp051

[13] P. J. Keall and P. W. Hoban, "Superposition dose calculation incorporating Monte Carlo generated electron track kernels," *Med. Phys.*, vol. 23, no. 4, pp. 479–485, 1996. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.597679

[14] B. De Man, M. Wu, P. FitzGerald, M. Kalra, and Z. Yin, "Dose reconstruction for real-time patient-specific dose estimation in CT," *Med. Phys.*, vol. 42, no. 5, pp. 2740–2751, 2015. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.4921066

[15] M. Wu, Z. Yin, and B. De Man, "Model-based dose reconstruction for CT dose estimation," *Med. Phys.*, vol. 44, no. 9, pp. e255–e263, 2017. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.12409

[16] S. Chandrasekhar, *Radiative Transfer*. Dover publications, Inc., 1950.

[17] E. T. Norris, X. Liu, and J. Hsieh, "Deterministic absorbed dose estimation in computed tomography using a discrete ordinates method," *Med. Phys.*, vol. 42, no. 7, pp. 4080–4087, 2015. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.4922000

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[19] J. M. Boone and J. A. Seibert, "An accurate method for computer-generating tungsten anode x-ray spectra from 30 to 140 kV," *Med. Phys.*, vol. 24, no. 11, pp. 1661–1670, 1997. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.597953

Fast and accurate photon-counting detector-based material decomposition via nearest neighborhood method

Jeonghyeon Nam and Okkyun Lee*

***Abstract*—A fast and accurate material decomposition is an important issue for the photon-counting detector (PCD)-based computed tomography (CT). The maximum likelihood (ML)-based method is the common choice for performing material decomposition due to the asymptotic optimal property; however, the solver for the method is usually iterative; thus, time-consuming due to the non-linear nature of the problem. In this paper, we propose the nearest neighborhood (NN)-based method, which is computationally efficient and accurate. The method is motivated by the mean squared error (MSE)-based estimator, which is guaranteed the asymptotic optimal property when the number of energy bins is equal to the number of basis materials. We validate the proposed method using various numerical simulations and show that the proposed method is faster than the maximum likelihood-based method while being as accurate.**

***Index Terms*—Photon-counting detector, nearest neighborhood algorithm, material decomposition, maximum likelihood**

I. Introduction

Photon-counting detector (PCD)-based computed tomography (CT) has potential benefits such as increased signal-to-noise ratio, identification of specific tissue, and K-edge imaging by utilizing the energy-dependent measurements from the PCD. It can also perform an accurate material decomposition [1], [2]; however, the measurements of photon-counting detector suffer from the spectral distortions by various causes such as K-escape photons and charge sharing [3].

Two groups of methods, a model-based method [4], [5] and a calibration-based method [6], [7], were developed over the years to address the distortions in the PCD while performing the material decomposition. The model-based method employs models of the system, such as the spectrum of incident x-ray and the spectral distortion in the PCD. The calibration-based method, on the other hand, adopts the calibration process to manage both bias and noise of the system. One of the common approaches to the model-based method is the maximum likelihood (ML)-based algorithm. The ML-based algorithm exploits the noise statistics of the measurements; thus, the algorithm is proven to be unbiased with minimum noise, unlike the heuristic calibration-based methods. However, solving the ML algorithm is an iterative process that can be time-consuming.

J.Nam and O.Lee are with the Department of Robotics Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, 42988, Republic of Korea.

*Corresponding author(Contact: oklee@dgist.ac.kr)

This work was supported by the DGIST Start-up Fund Program of the Ministry of Science and ICT(2020010044).

In this paper, we propose the nearest neighborhood (NN)-based material decomposition method that is as accurate as of the ML-based method, but computationally more efficient. Instead of finding the optimal solution for the cost function, the proposed method finds the nearest point to the measurement (i.e., the number of photons) in the pre-created data, so that the corresponding basis line-integrals of the nearest point are the optimal solution. The proposed NN-based algorithm mimics the mean squared error (MSE)-based estimator, motivated from the observation by Alvarez [7], [8]: The optimal solution from the MSE-based estimator is equivalent to those from the ML-based estimator, and the existence of the solution of the MSE-based one is guaranteed for the matched number of energy bins and the number of basis materials. In this study, we investigate the two-material decomposition with two energy bins.

The paper is organized as follows. Section II explains the background information and notations used in this paper. Section III describes the MSE-based method and the proposed method, the implementation issue of the method, and the setting for the numerical simulation. Section IV shows the results followed by the conclusion in Section V.

II. Background

A. Material Decomposition

The linear attenuation coefficient $\mu(\vec{r}, E)\,[\mathrm{cm}^{-1}]$ of an object can be modeled with a linear combination of two or more basis functions, which are the linear attenuation coefficients of basis materials in this paper. For basis materials, two materials, water and bone, are sufficient to describe the human body in the absence of contrast agents [2]

$$\mu(\vec{r}, E) = c_W(\vec{r})\,\mu_W(E) + c_B(\vec{r})\,\mu_B(E). \tag{1}$$

The vector $\vec{r}$ represents the position of the object, E is the x-ray energy, $c_W(\vec{r})$ and $c_B(\vec{r})$ are the unitless basis coefficient of water and bone, respectively, $\mu_W(E)$ and $\mu_B(E)\,[\mathrm{cm}^{-1}]$ are the linear attenuation coefficient of water and bone, respectively. Taking the line integral along the path of the x-ray gives the following equation:

$$\int \mu(\vec{r}, E)\,d\vec{r} = v_W \mu_W(E) + v_B \mu_B(E), \tag{2}$$

where $v_W = \int c_W(\vec{r})\,d\vec{r}$ and $v_B = \int c_B(\vec{r})\,d\vec{r}[\mathrm{cm}]$ are the thickness of water and bone along the given x-ray path, respectively.

B. Forward Model

The forward model of the PCD is as follows [4]:

$$\bar{y}_b\left(\vec{v}\right) = \int_0^\infty N_0 S_0\left(E\right) S_b\left(E\right) e^{-\left(\mu_W(E) v_W + \mu_B(E) v_B\right)} dE, \quad (3)$$

where $\vec{v} = \left[v_W, v_B\right]^T$ is the vector of the basis line-integrals, $\bar{y}_b$ is the average number of photons for the b^{th} energy bin, N_0 is the total number of incident photons when there is no object, $S_0\left(E\right)$ is the normalized x-ray incident spectrum, and $S_b\left(E\right)$ is the model for the spectral distortion in the PCD for the b^{th} energy bin. y_b denotes the measurement for the b^{th} energy bin, $\vec{y} = \left[y_1, y_2\right]^T$ is the vector of the PCD measurements, and $\vec{\bar{y}} = \left[\bar{y}_1, \bar{y}_2\right]^T$ is the vector of the mean number of photons.

C. Maximum Likelihood-based Method

The measurements of the photon-counting detector in each energy bin are assumed to follow Poisson distribution independently:

$$p\left(y_b; \bar{y}_b\left(\vec{v}\right)\right) = \frac{\bar{y}_b\left(\vec{v}\right)^{y_b} e^{-\bar{y}_b(\vec{v})}}{y_b!}, \quad (4)$$

which is the likelihood for the b^{th} energy bin. The conventional ML-based method estimates the basis line-integrals from the following time-consuming non-linear optimization problem:

$$\vec{v}_{ML} = \arg\min_{\vec{v}} L\left(\vec{v}\right) = \arg\min_{\vec{v}} \sum_{b=1}^{N_b} \bar{y}_b\left(\vec{v}\right) - y_b \ln \bar{y}_b\left(\vec{v}\right), \quad (5)$$

where $L\left(\vec{v}\right)$ is the negative log-likelihood for the entire energy bins.

III. Methods

In this section, we first describe the mean squared error-based method and then propose the nearest neighborhood-based method that is practically a non-iterative version for the MSE-based one.

A. The Mean Squared Error-based Method

The mean squared error-based method is motivated by Alvarez [7], which claims that $\vec{v}$ satisfying $\vec{y} = \vec{\bar{y}}\left(\vec{v}\right)$ is the solution for the ML. The claim implies that the solution of the maximum likelihood is equal to the optimal solution to the following:

$$\vec{v}_{MSE} = \arg\min_{\vec{v}} \left\| \vec{y} - \vec{\bar{y}}\left(\vec{v}\right) \right\|_2, \quad (6)$$

and $\vec{v}_{MSE}$ satisfying $\vec{y} = \vec{\bar{y}}\left(\vec{v}_{MSE}\right)$ exists when the number of energy bins is equal to the number of basis materials. We will refer (6) as the MSE-based method in this paper. Note that the MSE-based method also needs iterative solver; thus, computationally expensive as the ML-based algorithm.

B. The Proposed Nearest Neighborhood-based Method

The nearest neighborhood-based method is inspired by the MSE-based method. Minimizing the MSE in (6) is the same as finding $\vec{v}^\star$ that makes the shortest Euclidean distance from $\vec{y}$ to $\vec{\bar{y}}\left(\vec{v}^\star\right)$. From this idea, we propose the NN-based method which is computationally efficient.

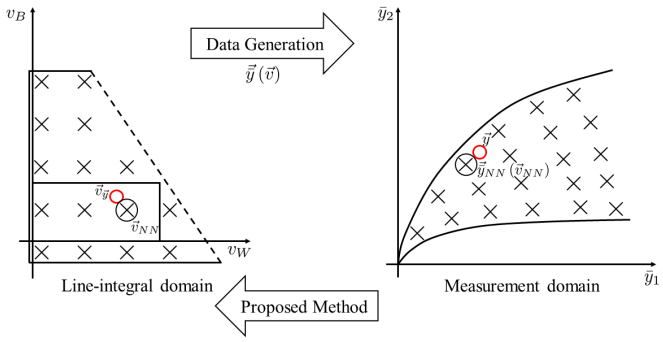


Fig. 1: Visual demonstration of the nearest neighborhood method

Figure 1 shows the visual demonstration of the NN-based method. In the basis line-integral domain, we create a regular grid, which is a set of evenly distributed combinations of basis line-integrals. From the regular grid, a measurement grid is built using (3). When the measurement from the photon-counting detector ($\vec{y}$, represented as the red circle in the measurement domain) is given, the algorithm searches the nearest point in the measurement grid ($\vec{\bar{y}}_{NN}\left(\vec{v}_{NN}\right)$, indicated as the black circle with x mark in the measurement domain). The corresponding basis line-integral of the nearest point, $\vec{v}_{NN}$, is the optimal solution of the NN-based method.

C. Practical Implementation

To generate the regular grid, we need to decide the ranges for each basis line-integral and the interval between each point. We set the ranges by assuming the minimum and maximum thickness of each material in the object, and adjust the ranges considering the noise of the measurements. In this paper, thicknesses of water and bone are expected from 0 to 35 and 0 to 5 [cm], respectively. The ranges are adjusted from -0.6 to 50.5 and -1.7 to 14.5 [cm], respectively, taking account of the noise in the estimates. The interval is heuristically set to 0.05 cm for each basis line-integral. After calculating the measurement grid, points whose L1-norm is smaller than the grid threshold are removed from data to reduce the size of the data.

$$grid\ threshold = 0.9 \sum_{i=1}^{N_b} \bar{y}_b\left(\left[35, 5\right]^T\right) \quad (7)$$

The grid threshold (7) is 90% of the photon counts in the assumed maximum thickness in each material, since we presumed any data lower than the grid threshold is unlikely to happen in our settings.

D. Simulation Settings

X-Ray: A 140kVp incident spectrum was used with $N_0 = 10^6$.

Object: To validate the proposed method, slab phantoms with different combinations of water (from $0cm$ to $35cm$) and bone($0cm$ and $4cm$), as well as an abdominal phantom were

used. The abdominal phantom was generated by the XCAT v2 program [9]. The phantom image has 512×512 pixels with a pixel width of 0.75 mm.

Photon-counting detector: In a two dimensional equi-distance fan-beam geometry, the distances from the center of rotation to the x-ray source and to the detectors were both set to 50 cm with 1,000 views per rotation. We used an array of 601 detectors (1.5 mm/channel) and two energy bins were set as 20 and 75 keV. For the distortion model in the photon-counting detector, physical effects such as charge sharing and K-escape fluorescence were considered [4], while a pulse pile-up effect was neglected since it may affect on the boundary of the objects [1]. The simulation was performed with ASTRA toolbox [10].

Algorithm: For the ML and MSE-based methods, we utilised *fminunc* function (quasi-Newton algorithm) in Optimization toolbox (MATLAB R2018a). To solve the NN-based method, we processed the measurement grid to a k-d tree with *KDTreeSearcher* function, and searched the nearest point by *knnsearch* function; both functions are in Statistics and Machine Learning toolbox (MATLAB R2018a).

Device: The runtime was measured using Intel® Core™ i7-9700K (3.60GHz) in a Windows 10 system.

IV. Numerical Results

To prove the assumption that there exists a $\vec{v}^{\star}$ satisfying $\vec{y} = \vec{\bar{y}}(\vec{v}^{\star})$, we used the following figure of merit:

$$score\left(\vec{\bar{y}}(\vec{v}_{MSE}), \vec{y}\right) = \frac{\left\|\vec{\bar{y}}(\vec{v}_{MSE}) - \vec{y}\right\|_1}{\|\vec{y}\|_1}. \quad (8)$$

Figure 2 shows the score of the slab geometry. The score is sufficiently low, around 10^{-5} (caused by mostly numerical computations), to convince that the assumption is true. Figure 3 illustrates the standard deviation and the bias of the slab geometry reconstruction results with different grid intervals. The 0.01 cm and 0.05 cm cases mostly satisfy the Cramér-Rao lower bound, except for the slight increase at low attenuation areas, and the bias for both grid intervals is relatively low. Both bias and noise with grid interval of 0.50 cm is highly unstable due to the sparsity of the data. In terms of the speed, the coarser the grid size is, the faster computation time it takes. For example, Table I demonstrates the average runtime, processing the full sinogram of the abdominal phantom, of the proposed method with various grid intervals. The runtime of

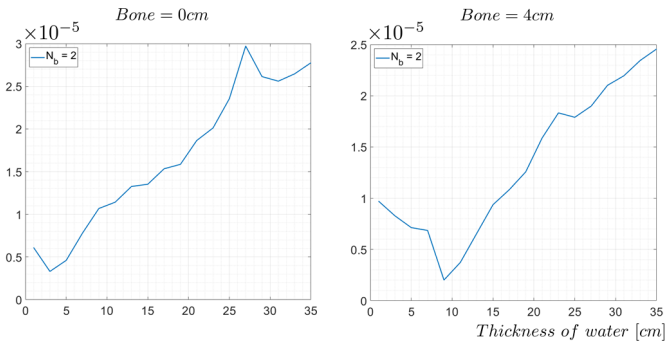


Fig. 2: Score (see (8)) of the slab geometry

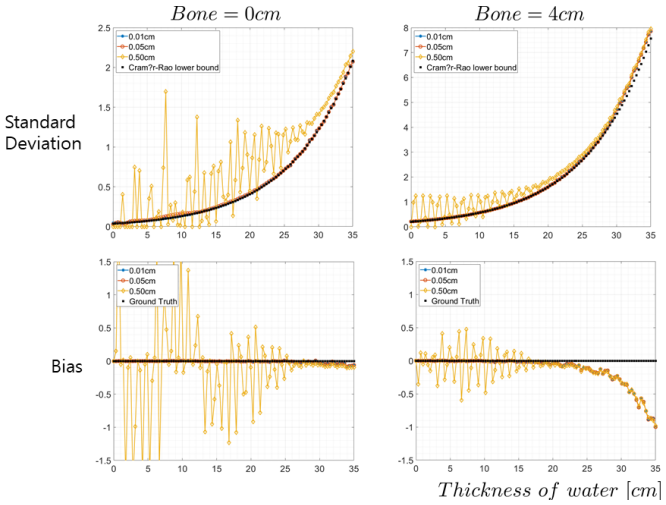


Fig. 3: Standard deviation and bias of the estimation for thickness of the water with different grid intervals

Interval [cm]	0.01	0.05	0.50
Time [ms]	158.3	56.6	42.1

TABLE I: Average (for 10,000 noise trials) runtime of the proposed method with various grid intervals for estimating the entire sinogram of the abdominal phantom.

the proposed method increases as the grid interval decreases. Thus, in this paper, we chose the 0.05 cm interval for the speed and accuracy.

The comparison of different estimators is shown in Figure 4. The three estimators have almost the same bias and noise, satisfying the Cramér-Rao lower bound. This demonstrates that the effectiveness of the proposed algorithm (between NN and MSE) as well as the theory behind the algorithm (between MSE and ML).

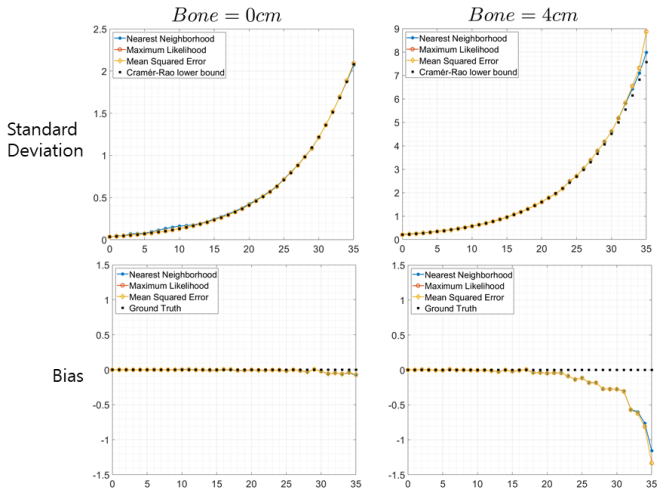


Fig. 4: Standard deviation and bias of the estimation for thickness of the water with different estimators

The estimated basis images (for a single noise trial) for the abdominal phantom using various methods and the difference

images are illustrated in Fig. 5. As presented in the difference images in Fig. 5, the ML and MSE-based estimators provide practically same results; however, the MSE and NN-based estimators have noticeable difference due to the discretization in the NN-based algorithm. Figure 6 shows the synthesized CT images from the basis images in Fig. 5. The difference image shows that results from both methods have minimal difference inside the phantom. However, the nearest neighborhood-based method only takes 0.0566 seconds in average to process an entire sinogram($601 \times 1,000$), while the ML-based and MSE-based methods took 1,995 and 2,372 seconds, respectively (10 noise trials for ML and MSE-based ones).

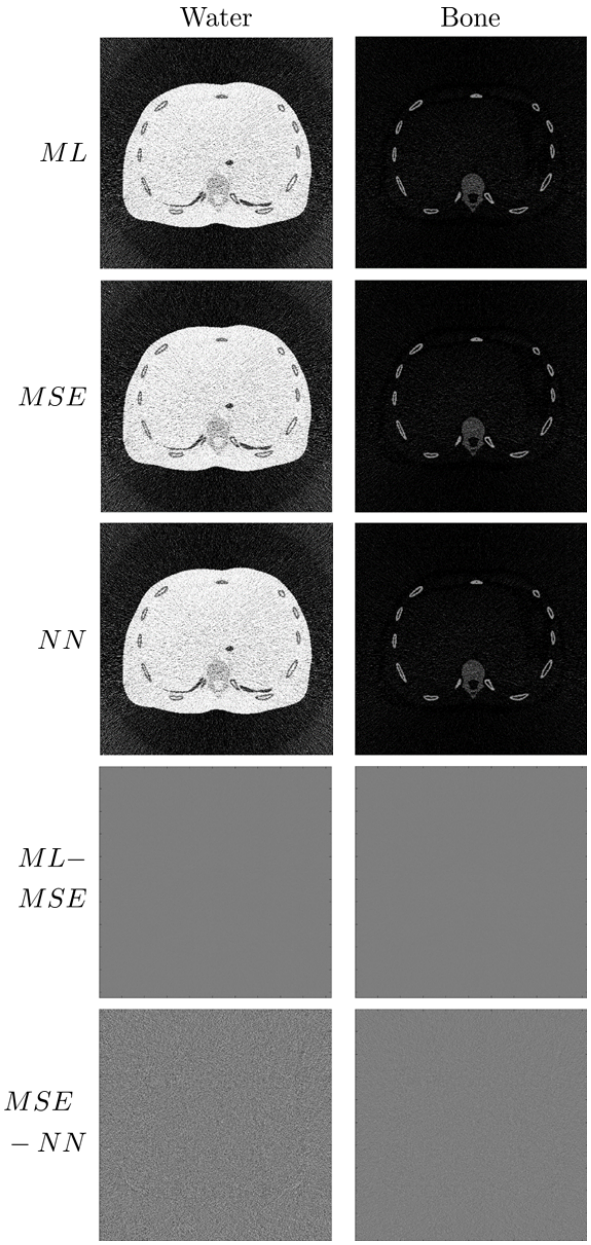


Fig. 5: Material decomposition images and difference images with different estimators (WW = 1, WL = 0.5 for basis images, WW = 0.2, WL = 0 for the difference images)

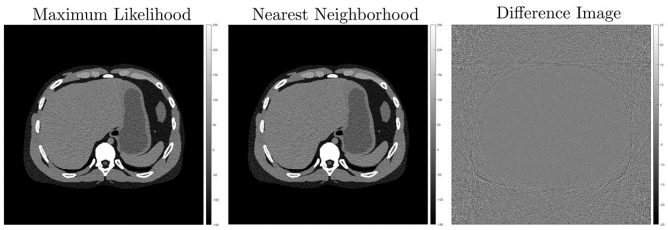


Fig. 6: Reconstructed CT images and difference image at 70 keV (CT images: WW = 400HU, WL = 50HU, Difference image: WW = 50HU, WL = 0HU)

V. Conclusion

In this paper, we proposed the computationally efficient nearest neighborhood-based method for material decomposition in photon-counting detector CT. The proposed method uses the x-ray spectrum and distortion model to formulate a k-d tree data so that the measurement can be rapidly searched and return the basis line-integrals. The algorithm is validated by numerical simulations, proving that the proposed method is faster than the maximum likelihood-based estimator, while maintaining the accuracy of the reconstruction when both the number of energy bins and the materials are two.

References

[1] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: single photon counting x-ray detectors in medical imaging," *Med. Phys.*, vol. 40, no. 10, pp. 100 901(1–19), October 2013.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerized tomography," *Phys. Med. Biol.*, vol. 21, no. 5, pp. 733–744, 1976.

[3] J. Cammin, J. Xu, W. C. Barber, J. S. Iwanczyk, N. E. Hartsough, and K. Taguchi, "A cascaded model of spectral distortions due to spectral response effects and pulse pileup effects in a photon-counting x-ray detector for CT," *Med. Phys.*, vol. 41, no. 4, pp. 041 905(1–15), April 2014.

[4] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler, A. Livne, and R. Proksa, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys. Med. Biol.*, vol. 53, pp. 4031–4047, 2008.

[5] O. Lee, S. Kappler, C. Polster, and K. Taguchi, "Estimation of basis line-integrals in a spectral distortion-modeled photon counting detector using low-order polynomial approximation of x-ray transmittance," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 560–573, February 2017.

[6] D. Wu, L. Zhang, X. Xu, and S. Wang, "A weighted polynomial based material decomposition method for spectral x-ray ct imaging," *Phys. Med. Biol.*, vol. 61, no. 10, pp. 3749–3783, May 2016.

[7] R. E. Alvarez, "Efficient, non-iterative estimator for imaging contrast agents with spectral x-ray detectors," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1138–1146, April 2016.

[8] ——, "Invertibility of the dual energy x-ray data transform," *Med. Phys.*, vol. 46, no. 1, pp. 93–103, January 2019.

[9] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys.*, vol. 37, no. 9, pp. 4902–4915, September 2010.

[10] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the ASTRA toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25 129–25 147, October 2016.

Segmentation-free statistical method for polyenergetic X-ray computed tomography with a calibration step

Cristóbal Martínez, Jeffrey A. Fessler, Manuel Desco and Mónica Abella

Abstract— **The polyenergetic nature of the spectra in X-ray sources can cause artifacts and non-quantitative values in the reconstructed image because of the beam-hardening effect.**

There are several strategies to correct the beam-hardening artifacts, but, unless the spectrum is known, the bone values remain not quantitative. We recently proposed a quantitative method that avoids the use of the spectrum by characterizing the beam-hardening effect by acquiring a phantom composed of soft tissue and bone. However, this method needs a bone segmentation in a preliminary reconstruction that can be difficult in low-dose acquisitions. This work solves the segmentation problem by incorporating the previous characterization of the beam hardening effect in a statistical iterative reconstruction method.

Evaluation using simulations showed a high reduction of beam-hardening artifacts in low dose studies when using the proposed method, while recovering the real density values.

Index Terms—**Beam-hardening, CT, artifacts, penalized-likelihood, streaks, polychromatic.**

I. Introduction

The beam hardening effect in computed tomography derives from the polychromatic nature of the radiation produced by X-ray tubes. Due to the energy dependence of mass attenuation coefficients, low energy photons are preferably absorbed, causing a shift of the mean energy of the X-ray beam to higher values. This effect leads to two main artifacts in the uncorrected images: cupping in homogeneous regions and streaks between dense areas in heterogeneous regions [1].

Several strategies exist in the literature to compensate for this effect. Physical filters are generally used to pre-harden the beam before reaching the sample, but this is not enough to remove the artifacts. Another method implemented in most commercial scanners is the water-linearization, based on a prior calibration with a water-equivalent phantom. The calibration step characterizes the relation between the attenuation values and the quantity of material traversed, that we will refer to as beam-hardening function. This method models the object as composed only of soft tissue and corrects only cupping artifact [2]. To correct also streaks, Nalcioglu et al. [3] applied a correction factor to the original projections by using the knowledge of the spectrum, the linear attenuation coefficients and the thickness of soft tissue and bone traversed estimated by segmenting a preliminary reconstruction. Joseph et al. [4] proposed a similar idea, modeling the corrected data with a second-order polynomial dependent on the bone traversed thickness, also needing a complete characterization of the spectrum to obtain the optimum parameters for this model. This need to know the spectra was avoided in [5, 6], based on a linear combination of basis images to correct streaks. That approach has two main limitations: the coefficients of this linear combination are obtained iteratively maximizing the flatness of the soft tissue areas, which could reduce the soft-tissue contrast, and it did not provide quantitative values in the bone.

To obtain quantitative values without knowing the spectrum, we recently proposed two methods extending the water-linearization to characterize the 2D beam-hardening function corresponding to soft tissue and bone [7, 8]. In [8], we used a phantom composed of soft tissue and bone in a calibration step. To avoid the calibration step, in [7] we characterized the function by using the information provided by the sample. Both methods have two disadvantages. First, the image values are the attenuation coefficients of the tissues, which depend on the acquisition energy. Second, they need the quantity of traversed bone, requiring a good bone segmentation, which may hinder their use in low-dose studies.

To deal with low-dose studies and eliminate the dependence of the attenuation values on the energy, Elbrakri et al. presented a statistical method that requires knowledge of the spectrum [9, 10]. That requirement was avoided in [11] with a simplified statistical algorithm that parameterizes the beam-hardening function following the model proposed by Joseph and Spital [4], but where the parameters of the model are found empirically. To avoid the issues related to this parameter optimization, we combined the works [7] and [11] in [12] to eliminate the beam-hardening artifacts. However, it was also

Manuscript received January 1, 2020. This work has been financed by the Ministry of Science, Innovation and Universities (Carlos III Health Institute, projects DTS17/00122; State Research Agency, project DPI2016-79075-R - AEI/FEDER, UE), financed jointly by Funds of the European Union (FEDER), "A way of making Europe". The CNIC is financed by the Ministry of Science, Innovation and Universities and the PRO-CNIC and it is a centre of excellence Severo Ochoa (SEV-2015-0505).

C. Martinez, M. Abella and M. Desco are with the Dpto. Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid and the Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España (e-mail: mabella@ing.uc3m.es, cmartinez@hggm.es).

J. A. Fessler is with Electrical Engineering and Computer Science department, The University of Michigan, Ann Arbor, MI 48109-2122, USA.

M. Abella and M. Desco are with the Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, España (e-mail: manuel.desco@cnic.es).

M. Desco is with the Centro de investigación en red salud mental (CIBERSAM), Madrid, España. (e-mail: desco@hggm.es).

necessary to know the bone density. This paper combines the previous works [8, 11] by calculating the beam-hardening function through a calibration step with a phantom composed of soft tissue and bone.

II. Materials and Methods

A. Forward model

This section briefly reviews the forward model in [11, 12] for clarity. The measurements are modeled as independently distributed Poisson random variables corrupted with extra background counts [13]:

$$Y_i \sim Poisson\left\{\overline{Y_i}\right\},\ i = 1, \dots, N \tag{1}$$

with:

$$\overline{Y_i} = \int I_i(\varepsilon) e^{-\int_{L_i} \mu(\varepsilon) dl} d\varepsilon + r_i \tag{2}$$

where the integral in the exponent follows the trajectory of the line L_i, $\mu(\varepsilon)$ is the attenuation coefficient at each energy ε, $I_i(\varepsilon)$ is the incident intensity and the term r_i accounts for mean scatter and background signals for the *i*th ray.

We model the attenuation coefficient in (2) at pixel *j* as:

$$\mu_j(\varepsilon) = \sum_{k=1}^{K} mac_k(\varepsilon) f_k^j \rho_j \tag{3}$$

where ρ is the density, mac_k is the mass attenuation coefficient of the material *k*, and f_k^j is a unitless fraction that describes the contribution of the material *k* to attenuation in pixel *j*. In this work, we assume that the object contains only soft tissue (*ST*) and bone (*B*). We define the contribution of each tissue type to the line integral along the *i*th ray as line density thickness:

$$t_{ST}(\rho) = \sum_{j=1}^{p} a_{ij} f_{ST}^j(\rho_j) \rho_j \tag{4}$$

$$t_B(\rho) = \sum_{j=1}^{p} a_{ij} f_B^j(\rho_j) \rho_j \tag{5}$$

where a_{ij} are the elements of the system matrix. Here we allow the unitless fraction (f_k^j) to be between 0 and 1, i.e., the pixels contain mixtures of tissues [11]. Eq. (2) for the expected value of the measured data along path *i* becomes:

$$\overline{Y_i}(\rho) = \int I_i(\varepsilon) e^{-\int_{L_i} \mu(\varepsilon) dl} d\varepsilon + r_i = \\ = I_i e^{-F\left(t_{ST}^i(\rho), t_B^i(\rho)\right)} + r_i \tag{6}$$

where

$$I_i \equiv \int I_i(\varepsilon) d\varepsilon \tag{7}$$

and the beam-hardening function, *F*, is:

$$F(t_{ST}, t_B) = -\log \int \frac{I(\varepsilon)}{I} e^{-mac_{ST}(\varepsilon) t_{ST} - mac_B(\varepsilon) t_B}\ d\varepsilon, \tag{8}$$

dropping the dependence on ray *i* for simplicity.

B. Beam-hardening function

We obtain the beam-hardening function $F(t_{ST}, t_B)$ simulating a digital calibration phantom made of soft tissue and bone, as described in [8]. The phantom has two semicircles of each material to maximize the combination of traversed soft tissue and bone. Both tissues are segmented in a preliminary reconstruction with Filtered Back Projection (FBP) using binary thresholding and the resulting bone and soft-tissue masks are projected to obtain the traversed thicknesses. Fig. 1 shows the workflow of the calibration step.

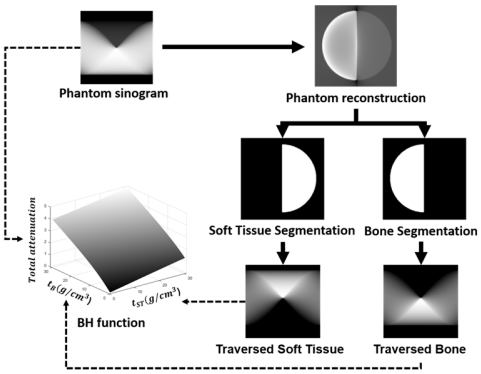


Fig. 1. Workflow of the calibration step.

C. Cost function

The negative log-likelihood for independent Poisson measurement is:

$$L(\rho) = -\sum_{i=1}^{N} h_i\left(F\left(t_{ST}(\rho), t_B(\rho)\right)\right) \tag{9}$$

where

$$h_i(d) = -Y_i \log(I_i e^{-d} + r_i) + I_i e^{-d} + r_i \tag{10}$$

Since minimizing $L(\rho)$ is generally an ill-posed problem, regularization is included by adding a penalty term to control how much the object ρ departs from our assumptions about image properties. In this work, we use a 3D roughness penalty function with the convex edge-preserving Huber potential. The resulting penalized cost function is:

$$\Phi(\rho) = L(\rho) + \beta R(\rho) \tag{11}$$

where β is a scalar that controls the tradeoff between the data-fit and penalty terms.

D. Algorithm

We derive an iterative algorithm based on separable quadratic surrogates using the principles of optimization transfer [14], resulting in the following update:

$$\rho^{n+1} = \rho^n - D^{-1} \nabla \Phi(\rho^n) \tag{12}$$

where *D* is a diagonal matrix that influences the rate of convergence. We originally designed *D* to ensure that the algorithm monotonically decreases the cost function. As in [14], in practice we choose the elements of *D* approximately by using the precomputed curvature:

$$d_j = \left(mac_{ST}^2(\varepsilon_{eff}) + mac_B^2(\varepsilon_{eff})\right) \sum_{i=1}^{N} a_{ij} \sum_j\ a_{ij}\, Y_i \tag{13}$$

where the effective $mac(\varepsilon_{eff})$ values for each tissue are approximated using the derivative of beam-hardening function at (0,0).

III. Evaluation

Preliminary evaluation was based on a simulation of a 2D phantom of soft tissue (1.06 g/ cm^3) with ten cortical bone inserts having density 1.92 g/ cm^3 and two inserts of adipose tissue with density equal to 0.9 g/ cm^3 (Fig 2).

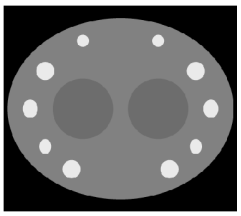


Fig. 2. Test phantom with different cortical bone inserts and two adipose tissue inserts inside of a soft-tissue ellipse.

The polyenergetic X-ray datasets were generated using MIRT (https://github.com/JeffFessler/mirt) with a 45 kVp spectrum and 0.1 mm aluminum filtration, commonly used in preclinical studies. We chose 10^5 and 10^6 number of counts per ray to simulate low-SNR and high-SNR scenarios respectively. We obtained 60 projections and 180 projections of 512×512 pixels within a 180-degree span for the low-SNR and high-SNR cases respectively. These datasets were reconstructed with FBP, with FBP corrected by the simple calibration method (FBP+sCM) proposed in [7], by a monochromatic statistical algorithm [15] and the proposed polyenergetic statistical algorithm. Root mean square error (RMSE) with respect to the true density phantom of the FBP, the FBP+sCM, the monoenergetic and the proposed methods is calculated in three ROIs (whole phantom, soft tissue and bone).

IV. Results

Fig 3. and Fig. 4, show the results for the high and low SNR scenarios respectively. Results of FBP+sCM show a good compensation of beam-hardening artifacts but no correction of streaks in the low-SNR scenario. The monoenergetic algorithm reduces the streaks due to low-sampling in the low-SNR case, but it is not able to eliminate completely the beam-hardening artifacts. The proposed method is able to correct both the beam-hardening artifacts and the streaks associated to low-sampling in both scenarios.

Table I shows the RMSE for the different scenarios of high and low SNR and the different corrections, where the proposed method has the lowest value in all the cases.

TABLE I
RMSE FOR THE EVALUATION PHANTOM

	RMSE	FBP	FBP+sCM	Monochromatic	Proposed Method
High SNR	Total[a]	0.64	1.20	0.68	0.06
	Bone[b]	2.5	4.90	2.69	0.09
	ST[c]	0.21	0.16	0.20	0.06
Low SNR	Total[a]	0.67	1.14	0.68	0.06
	Bone[b]	2.52	4.53	2.69	0.09
	ST[c]	0.28	0.32	0.19	0.06

Units of density g/cm^3.
[a]Average true density of 1.08 g/cm^3.
[b]Average true density of 1.92 g/cm^3.
[c]Average true density of 1.02 g/cm^3

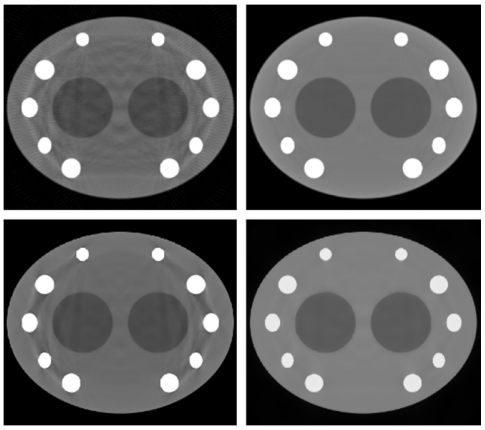


Fig. 3. Results for the 180 projections and high-SNR datasets using FBP (top-left), FBP + sCM (top-right), monoenergetic statistical algorithm (bottom-left) and proposed method (bottom right).

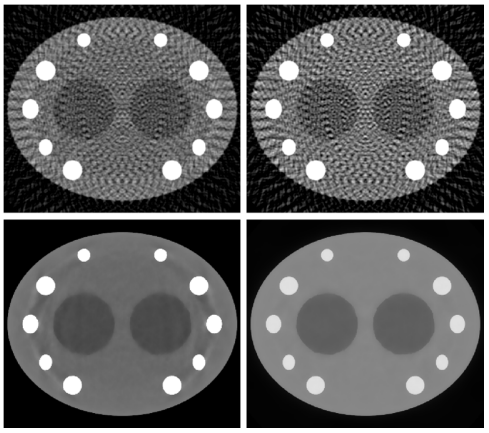


Fig. 4. Results for the 60 projections and low-SNR datasets using FBP (top-left), FBP + sCM (top-right), monochromatic statistical algorithm (bottom-left) and proposed method (bottom right).

V. Conclusions

We have presented a new statistical reconstruction algorithm that includes beam-hardening correction based on the modeling of the polychromatic effect with a simple phantom composed of soft tissue and bone, avoiding the need of spectrum knowledge or tuning of parameters.

Results on simulated data show reduced dark bands associated with beam-hardening and reduced streaks due to low number of projections, while recovering density values.

Future work includes evaluation on real data. To this end, we need to search for soft tissue and bone equivalent materials to build a realistic calibration phantom, since here we considered an ideal one not possible to manufacture.

References

[1] J. F. Barrett and N. Keat, "Artifacts in CT: recognition and avoidance," *Radiographics,* vol. 24, pp. 1679-91, Nov-Dec 2004.

[2] R. A. Brooks and G. Di Chiro, "Beam hardening in x-ray reconstructive tomography," *Phys Med Biol,* vol. 21, pp. 390-8, May 1976.

[3] O. Nalcioglu and R. Lou, "Post-reconstruction method for beam hardening in computerised tomography," *Physics in medicine and biology,* vol. 24, p. 330, 1979.

[4] P. M. Joseph and R. D. Spital, "A method for correcting bone induced artifacts in computed tomography scanners," *J Comput Assist Tomogr,* vol. 2, pp. 100-8, Jan 1978.

[5] Y. Kyriakou, E. Meyer, D. Prell, and M. Kachelrieß, "Empirical beam hardening correction (EBHC) for CT," *Med Phys,* vol. 37, pp. 5179-5187, 2010.

[6] S. Schuller, S. Sawall, K. Stannigel, M. Hulsbusch, J. Ulrici, E. Hell, *et al.*, "Segmentation-free empirical beam hardening correction for CT," *Med Phys,* vol. 42, pp. 794-803, Feb 2015.

[7] C. Martìnez, C. de Molina, M. Desco, and M. Abella, "Calibration-free method for beam-hardening compensation: preliminary results," in *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2017, pp. 1-3.

[8] C. Martínez, C. De Molina, M. Desco, and M. Abella, "Simple method for beam-hardening correction based on a 2D linearization function," in *The 4th International Meeting on Image Formation in X-Ray Computed Tomography (CTMeeting 2016)*, Bamberg, 2016.

[9] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE transactions on medical imaging,* vol. 21, pp. 89-99, 2002.

[10] I. A. Elbakri and J. A. Fessler, "Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation," *Physics in medicine and biology,* vol. 48, p. 2453, 2003.

[11] M. Abella, C. Martínez, M. Desco, J. J. Vaquero, and J. A. Fessler, "Simplified Statistical Image Reconstruction for X-ray CT with Beam-Hardening Artifact Compensation," *IEEE transactions on medical imaging,* 2019.

[12] C. Martinez, J. Fessler, M. Desco, and M. Abella, "Statistical image reconstruction with sample-based beam-hardening compensation for X-ray CT," in *Proc. 5th Int. Conf. Image Formation X-Ray Computed Tomography*, 2018, pp. 11-14.

[13] R. E. Alvarez and A. Macovski, "Energy-selective reconstruction in x-ray computerized tomography," *Phys. Med. Biol.,* vol. 21, pp. 733–44, 1976.

[14] I. A. Elbakri and J. A. Fessler, "Statistical Image Reconstruction for Polyenergetic X-Ray Computed Tomography," *IEEE Trans. Med. Imaging,* vol. 21, pp. 89-99, 2002.

[15] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Physics in Medicine & Biology,* vol. 44, p. 2835, 1999.

A motion estimation and compensation algorithm for 4D CBCT with cyclic deformation model

Alexander Katsevich, Seongjin Yoon, Michael Frenkel, Peter Munro, Pascal Paysan, Igor Peterlik, Dieter Seghers, and Adam Strzelecki

***Abstract*—We propose an algorithm for motion estimation and compensation in the case of a slowly rotating gantry in the 4D CBCT setting. Our primary target application is abdomen imaging, which is challenging because of the lack of high-contrast features. The algorithm is based on alternating application of volume estimation and motion estimation sub-iterations. The algorithm uses the optical flow constraint, and the motion is described by a global in time deformation function $\mu(s,\vec{x})$. This function constrains the motion to be cyclic and allows us to easily incorporate the motion phase into the regularization functional. For improved stability, volume sub-iterations use motion vector fields derived from the current estimate of μ instead of μ itself. Results of experiments with simulated and three clinical datasets of the TrueBeam™ treatment device demonstrate that the algorithm presented in this paper can reduce artifacts and improve image quality compared to 4D CBCT clinical reconstructions currently utilized for imaging of the abdomen.**

***Index Terms*—4D CBCT, motion estimation, motion compensation, optical flow constraint, dynamic regularization strength selection, abdomen imaging.**

I. Introduction

4D cone beam computed tomography (CBCT) [1] has long been used to evaluate the motion of thoracic anatomy in radiation oncology. However, 4D CBCT reconstructions can suffer from severe view aliasing artifacts as well as motion artifacts from residual motion within each of the phase bins. Many algorithms to eliminate artifacts in 4D CBCT have been proposed (see e.g., [2]–[8]).

However, these algorithms generally rely on the presence of high contrast anatomy such as found in the thorax to perform well. Therefore, these algorithms have performed inconsistently in the abdomen. Iterative reconstruction algorithms based on enforcing the optical flow constraint (OFC) appear to be most powerful and can potentially provide clinically acceptable 4D CBCT image quality in the abdomen.

We previously proposed a motion vector field (MVF)-based OFC algorithm and tested it on simulated and clinical data [9]. The novel idea was that we change parameters of the cost functional in a dynamic fashion during optimization to steer the algorithm towards a solution with required properties, e.g., with reduced sparse view streaks. The results were promising, but still contained some false motion. After analyzing these results, we concluded that false motion tends to form a progressive wave that is stronger close to the patient boundary, where voxels located close to each other in space have different motion phases.

In an effort to reduce false motion further, we develop an OFC-based method that incorporates a cyclic deformation model. The method enforces the motion to be cyclic throughout the entire breathing cycle and gives us explicit control over the spatial phase of the each voxel motion. We tested the algorithm on simulated and clinical datasets. The results show significantly reduced false motion and sparse-view streaks.

A. Katsevich, S. Yoon, and M. Frenkel are with iTomography Corp., Texas Medical Center Innovation Institute, 2450 Holcombe Blvd, Houston, TX 77021, USA (e-mail: alexander@iTomography.com, syoon@iTomography.com, michael@iTomography.com).

A. Katsevich is also with the University of Central Florida, Mathematics Department, Orlando, FL 32816, USA.

A. Katsevich is a shareholder and chief technology officer of iTomography Corporation and, as such, may benefit financially as a result of the outcomes of the work reported in this publication.

P. Munro, P. Paysan, I. Peterlik, D. Seghers, and A. Strzelecki are with Varian Medical Systems Imaging Laboratory, Taefernstrasse 7, 5405 Daettwil, Switzerland (e-mail: Peter.Munro@varian.com, Pascal.Paysan@varian.com, Igor.Peterlik@varian.com, Dieter.Seghers@varian.com, Adam.Strzelecki@varian.com).

II. Motion estimation and compensation based on the parameterized cyclic deformation model

Let s be the time (and also phase) variable, and $\vec{x}$ the 3D space variable. The dynamic object is represented by a volume function $f(s,\vec{x})$ and a deformation $\vec{\mu}(s,\vec{x})$ so that $f(s,\mu(s,\vec{x})) = f_0(\vec{x})$. Here, $f_0(\vec{x})$ is the attenuation coefficient at reference time/phase.

In our previous work [9], we studied the Lagrangian approach, where the OFC is expressed in terms of the MVF by taking the time derivative of $f(s,\vec{x})$:

$$\frac{\partial f(s,\vec{x})}{\partial s} + \nabla f(s,\vec{x}) \cdot \vec{v}(s,\vec{x}) = 0, \tag{1}$$

where $\vec{v}(s,\vec{x})$ is the velocity vector at $\vec{x}$ at time s. While this gave us a simpler reconstruction approach, it was not sufficiently robust to capture the correct motion due to the ill-posedness of the CBCT reconstruction problem.

Instead of the MVF-based approach, we reformulate the OFC using a parameterized deformation function. Suppose the reconstruction volume grid is given by

$$\vec{x}_{\mathbf{i}} = (h_1 i_1, h_2 i_2, h_3 i_3),\ \mathbf{i} = (i_1, i_2, i_3), \quad \text{for } 0 \leqslant i_q < N_q,\ q = 1,2,3. \tag{2}$$

For all $\vec{x}$, we have $f(s,\vec{x}) \approx \sum_{\mathbf{i}} f_{\mathbf{i}}(s)\varphi_h(\vec{x} - \vec{x}_{\mathbf{i}})$, where φ_h is the tri-linear interpolation kernel.

Let s_p, $0 \leqslant p < N_p$ be the time corresponding to the pth breathing phase bin. One way to incorporate the OFC is via the ℓ-2 norm penalty:

$$\Phi_{\mathrm{f}\mu}(\mathbf{f},\vec{\mu}) = \frac{1}{2}\sum_{\mathbf{i}}\sum_{p}\left| f(s_{p+1},\vec{\mu}(s_{p+1},\vec{x}_{\mathbf{i}})) - f(s_p,\vec{\mu}(s_p,\vec{x}_{\mathbf{i}}))\right|^2. \tag{3}$$

Advantages of and possibilities for improvement with the deformation-based approach are numerous:

1) We can enforce that each particle moves periodically along a line segment (or, along an elliptical path).
2) Such a description has fewer parameters to compute, which makes the overall motion estimation and compensation problem more stable.
3) Having easy access to the phase information (as part of the deformation function), we can reduce the progressive wave that causes false circular motion by using appropriate regularization, and leave only the stationary wave that is more physical.
4) An advantage of (3) over (1) is that it does not use the derivatives of f; hence, the requirements on the smoothness of f are relaxed. This may lead to improved spatial and temporal resolution of the method.

We choose the motion model as follows so that each voxel moves along a linear path:

$$\vec{\mu}(s,\vec{x}) = \vec{x} + \left[\cos\left(2\pi\frac{s-\bar{s}(\vec{x})}{\Delta S}\right) - \cos\left(2\pi\frac{\bar{s}(\vec{x})}{\Delta S}\right)\right]\vec{u}(\vec{x}), \quad (4)$$

where the amplitude $\vec{u}(\vec{x})$ and phase $\bar{s}(\vec{x})$ are expressed as follows:

$$\vec{u}(\vec{x}) = \sum_{\mathbf{j}} \vec{u}_{\mathbf{j}}\hat{\varphi}(\vec{x}-\vec{m}_{\mathbf{j}}), \quad (5)$$

$$\bar{s}(\vec{x}) = \sum_{\mathbf{k}} \bar{s}_{\mathbf{k}}\check{\varphi}(\vec{x}-\vec{n}_{\mathbf{k}}). \quad (6)$$

Note that neither the grids $\vec{x}_{\mathbf{i}}$, $\vec{m}_{\mathbf{j}}$, and $\vec{n}_{\mathbf{k}}$ nor the interpolation functions $\varphi(\cdot)$, $\hat{\varphi}(\cdot)$ and $\check{\varphi}(\cdot)$ have to be the same. In general, $\vec{u}$ is smooth, and $\bar{s}$ is even smoother. This reflects the hypothesis that the variability of the phase of the breathing motion throughout the patient is much smaller than the variability of its amplitude. Thus, we want to use a sparse grid with a smooth interpolation function for $\vec{u}$, and an even sparser grid and a smoother interpolation function for $\bar{s}$.

Let $\mathbf{f}$, $\mathbf{u}$, and $\bar{\mathbf{s}}$ denote phase volume, motion amplitude, and motion phase parameters unrolled into 1D vectors, respectively. Adding regularization terms for the volume parameters $\mathbf{f}$ and motion parameters $\mathbf{u}$ and $\bar{\mathbf{s}}$, the total functional is as follows:

$$\Phi(\mathbf{f},\mathbf{u},\bar{\mathbf{s}}) = \Phi_{\mathrm{L}}(\mathbf{f}) + \lambda\Phi_{\mathrm{f}\mu}(\mathbf{f},\mathbf{u},\bar{\mathbf{s}}) + \kappa_{\mathrm{f}}\Phi_{\mathrm{R_f}}(\mathbf{f},\delta) + \kappa_{\mathrm{u}}\Phi_{\mathrm{R_u}}(\mathbf{u}) + \kappa_s\Phi_{\mathrm{R_s}}(\bar{\mathbf{s}}), \quad (7)$$

where $\Phi_{\mathrm{L}}(\mathbf{f})$ is the data fidelity functional (as used in the penalized weighted least-squares approach), $\Phi_{\mathrm{f}\mu}(\mathbf{f},\mathbf{u},\bar{\mathbf{s}})$ is the OFC as described in (3), $\Phi_{\mathrm{R_f}}(\mathbf{f})$ is the volume regularizer, $\Phi_{\mathrm{R_u}}(\mathbf{u})$ is the motion amplitude regularizer, $\Phi_{\mathrm{R_s}}(\bar{\mathbf{s}})$ is the motion phase regularizer, and λ, κ_{f}, κ_{u}, and κ_s are regularization weights. The four functionals in (7) are defined as follows:

$$\Phi_{\mathrm{L}}(\mathbf{f}) := \frac{1}{2}(\mathbf{G}-\mathbf{A}\mathbf{f})^{\mathsf{T}}\mathbf{W}(\mathbf{G}-\mathbf{A}\mathbf{f}), \quad (8)$$

$$\Phi_{\mathrm{R_f}}(\mathbf{f}) := \sum_q\sum_{\mathbf{i}}\phi_{\mathrm{R}}([\mathbf{C}_q\mathbf{f}]_{\mathbf{i}},\ \delta), \quad (9)$$

$$\Phi_{\mathrm{R}_u}(\mathbf{u}) =: \frac{1}{2}\sum_q\sum_r\sum_{\mathbf{j}}\left([\mathbf{C}_r\mathbf{u}_q]_{\mathbf{j}}\right)^2, \quad (10)$$

$$\Phi_{\mathrm{R}_s}(\bar{\mathbf{s}}) =: \frac{1}{2}\sum_r\sum_{\mathbf{k}}([\mathbf{C}_r\bar{\mathbf{s}}]_{\mathbf{k}})^2, \quad (11)$$

where $\mathbf{G}$ is the entire set of measured data, $\mathbf{u}_q$ is the component of $\mathbf{u}$ along the qth axis, $\mathbf{A}$ is the system matrix, $\mathbf{W}$ is the weight matrix, $\mathbf{C}_q$ is the spatial finite difference matrix along the qth axis with clamped boundaries, $\mathrm{diag}\{\mathbf{t}\}$ is the diagonal matrix with the elements of vector $\mathbf{t}$ on the main diagonal, $[\mathbf{t}]_i$ is the ith element (scalar, vector, or matrix) along the first dimension of vector or matrix $\mathbf{t}$, $\phi_{\mathrm{R}}(t,\ \delta)$ is the hyperbolic potential regularizer defined as follows

$$\phi_{\mathrm{R}}(t,\ \delta) = \delta^2\left(\sqrt{1+(t/\delta)^2}-1\right). \quad (12)$$

In general, $\mathbf{W}$ is derived from the raw projection data to account for relative noise intensity. For now, we set $\mathbf{W}$ to be the identity matrix. The dimensions of $\mathbf{C}_q$ is determined by the vector to which it applies.

As in the previous work [9], we use an alternating direction minimization scheme to find the optimal phase volume and motion parameters by repeatedly applying two separate sub-iteration steps. To find the optimal solution at each sub-iteration step, we use an approximated Newton's method with a separable surrogate function.

III. Hybrid deformation-velocity approach

In the deformation-based OFC shown in (3), pairs of consecutive phase volumes are compared using a reference grid that is regular only in the reference phase. This grid is far from regular after deformation. Even though the OFC encourages incompressibility, the estimated volume usually shows some compressibility (whether it is real or not), and the data density at the expanding region becomes sparse, causing the OFC to be applied unevenly. Consequently, the OFC fails to regularize the expanding or compressing regions properly, resulting in speckle artifacts when the data are noisy and not perfectly matching.

To address this instability due to the data grid irregularity, we modify the OFC for volume sub-iterations so that the reference grid, where two consecutive phase volumes are compared, is regular:

$$\Phi_{\mathrm{f}\mu}(\mathbf{f},\mathbf{u},\bar{\mathbf{s}}) = \frac{1}{2}\sum_{\mathbf{i}}\sum_p\left|f(s_{p+1},\vec{x}_{\mathbf{i}}+\vec{V}_p(\vec{x}_{\mathbf{i}})) - f(s_p,\vec{x}_{\mathbf{i}})\right|^2. \quad (13)$$

Here $\vec{V}_p(\vec{x}_{\mathbf{i}})$ is the estimated MVF field between pairs of consecutive phase volumes on a regular grid. Given $\vec{\mu}(s_p,\vec{x}_{\mathbf{i}})$, we compute the MVF $\vec{V}_p(\vec{x}_{\mathbf{i}})$ in a form analogous to (5), (6) with another grid $\vec{x}_{\mathbf{i}'}$ by solving the following optimization problem:

$$\vec{V}_p = \arg\min_{\vec{v}} \frac{1}{2}\sum_{\mathbf{i}}\left|\vec{v}(\vec{\mu}(s_p,\vec{x}_{\mathbf{i}})) - (\vec{\mu}(s_{p+1},\vec{x}_{\mathbf{i}}) - \vec{\mu}(s_p,\vec{x}_{\mathbf{i}}))\right|^2. \quad (14)$$

For motion sub-iterations, we use the original deformation-based OFC (3), as the motion parameters need to be estimated using the entire cycle.

IV. Method of Alternating Minimization with respect to Volume and Velocity

The total cost functional in (7) is nonlinear, nonconvex, and ill-conditioned. To avoid undesirable local minima, we use the same method of alternating minimization with dynamic parameter updates that we introduced in the previous work [9].

To briefly explain the method, each global iteration consists of a set of volume sub-iterations and a set of motion sub-iterations. The cost functional is minimized with respect to either $\mathbf{f}$, or $\mathbf{u}$ and $\bar{\mathbf{s}}$ in an alternating fashion, while keeping the other variable(s) fixed.

Starting volume sub-iterations with a large λ when all motion parameters are initialized as zero ends up underestimating the motion. On the other hand, a small λ will cause false motion detection due to the sparse angle artifacts. We use dynamically changing λ and δ to address both effects: motion underestimation and false motion.

In our approach, the value of λ oscillates throughout the global iterations. With an oscillating λ, global iterations have another chance to catch small motions that could have been lost in previous iterations, which the static parameter setting cannot achieve. Also, we select δ to be large initially, and then gradually decrease its value as the global iterations progress. This approach initially identifies large features and ignores minor details in phase volumes to estimate global motion. In later iterations, the algorithm adds more details to phase volumes and estimates local motion.

V. Test Results

We present reconstruction results for one simulated and one clinical dataset. The former was generated and latter was acquired using Varian TrueBeam treatment device [10] with half-fan beam geometry. For the simulated dataset, we use a synthetic, sinusoidal in time motion.

The amplitude of the simulated motion is maximum at a point inside the patient, and goes to zero in a radial fashion away from the maximum following the Gaussian curve. The maximum peak-to-peak amplitudes of the motion in the y and z directions are 15 voxels (i.e., 18.75 mm), while there is no motion in the x direction. We also assume the half-fan beam geometry of the Varian TrueBeam treatment device.

For the clinical dataset, chest motion amplitude was recorded using TrueBeam's respiratory gating system and then converted to the breathing phase.

Fig. 1 shows the transverse slices of the reconstructed object of the simulated dataset using our proposed algorithm (the top row), the ground truth images (the middle row), and the difference between the reconstruction results and the ground truth (the bottom row). The reconstruction results of the body are nearly identical with the ground truth.

Fig. 2 shows the comparison of the transverse slices between the existing clinical reconstruction software *Advanced 4D* [8] of the TrueBeam treatment device (two left images 2(a)–(b)) and our proposed reconstruction algorithm (two right images 2(c)–(d)) obtained using the same clinical dataset (the results for two additional clinical cases as well as comparisons with our previous method [9] will be shown at the conference). By comparing 2(a)–(b) with 2(c)–(d), one can observe that the proposed algorithm significantly reduces the streak artifacts, allowing for performing more focused on-the-target treatment of patients.

VI. Conclusions

We developed a robust motion estimation and compensation algorithm for 4D CBCT. The algorithm uses the cyclic motion model incorporated into the OFC. The experimental results show that the use of a cyclic motion model significantly reduces sparse-view streaks and false motion artifacts.

For the simulated data, we are able to get the reconstructed phase volumes that are nearly error-free.

The reconstruction results from the clinical data demonstrate that presented in this paper algorithm can significantly reduce artifacts and improve image quality compared to imaging results of clinical reconstructions.

References

[1] J.-J. Sonke, L. Zijp, P. Remeijer, and M. van Herk, "Respiratory correlated cone beam CT," *Medical physics*, vol. 32, no. 4, pp. 1176–1186, 2005.

[2] T. Li, A. Koong, and L. Xing, "Enhanced 4D cone-beam CT with inter-phase motion model," *Medical physics*, vol. 34, no. 9, pp. 3688–3695, 2007.

[3] Q. Zhang, Y.-C. Hu, F. Liu, K. Goodman, K. E. Rosenzweig, and G. S. Mageras, "Correction of motion artifacts in cone-beam CT using a patient-specific respiratory motion model," *Medical physics*, vol. 37, no. 6Part1, pp. 2901–2909, 2010.

[4] Z. Qi and G. Chen, "Performance studies of four-dimensional cone beam computed tomography," *Physics in Medicine and Biology*, vol. 56, pp. 6709–6721, 2011.

[5] M. Brehm, P. Paysan, M. Oelhafen, and M. Kachelrieß, "Artifact-resistant motion estimation with a patient-specific artifact model for motion-compensated cone-beam CT," *Med. Phys.*, vol. 40, no. 10, pp. 1–13, 2013.

[6] C. M. Rank, T. Heußer, M. T. Buzan, A. Wetscherek, M. T. Freitag, J. Dinkel, and M. Kachelrieß, "4D respiratory motion-compensated image reconstruction of free-breathing radial MR data with very high undersampling," *Magn. Reson. Med.*, vol. 77, no. 3, pp. 1170–1183, 2017.

[7] H. Zhang, J. Ma, Z. Bian, D. Zeng, Q. Feng, and W. Chen, "High quality 4D cone-beam CT reconstruction using motion-compensated total variation regularization," *Phys. Med. Biol.*, vol. 62, pp. 3313–3329, 2017.

[8] J. Star-Lack, M. Sun, M. Oelhafen, T. Berkus, J. Pavkovich, M. Brehm, M. Arheit, P. Paysan, A. Wang, P. Munro, D. Seghers, L. Carvalho, and W. Verbakel, "A modified McKinnon-Bates (MKB) algorithm for improved 4D cone-beam computed tomography (CBCT) of the lung," *Med. Phys.*, vol. 45, no. 8, pp. 3783–3799, 2018.

[9] S. Yoon, A. Katsevich, M. Frenkel, P. Munro, P. Paysan, D. Seghers, and A. Strzelecki, "A motion estimation and compensation algorithm for 4D CBCT of the abdomen," *15th Int. Meet. Fully Three-Dimensional Image Reconstr. Radiol. Nucl. Med.*, vol. 11072, no. May, p. 83, 2019.

[10] S. Lang, M. Zamburlini, A. Stüssi, J. Hrbacek, and S. Klöck, "Ep-1298: Evaluation of dosimetric and geometric stability of a new digital linear accelerator over a period of 3 years," *Radiotherapy and Oncology*, vol. 106, p. S490, 2013.

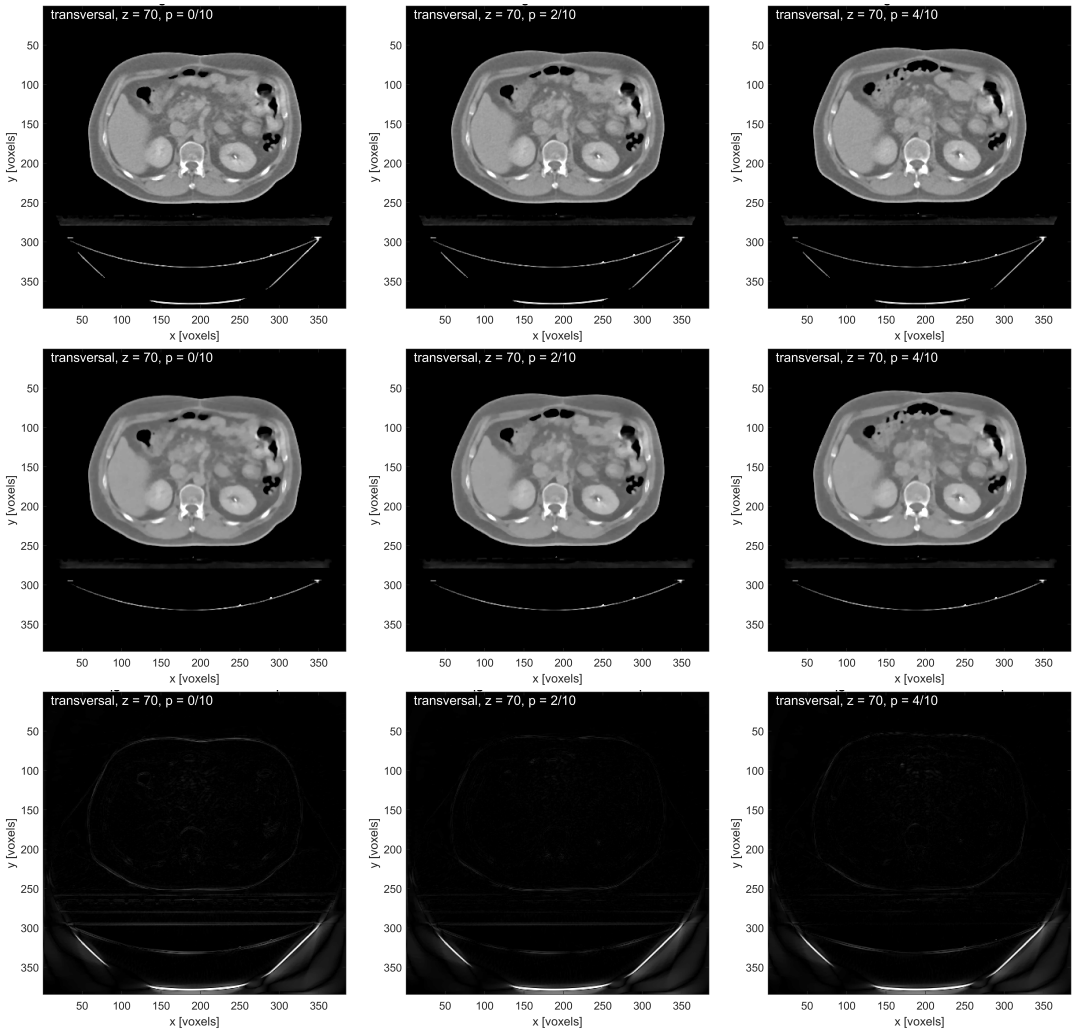
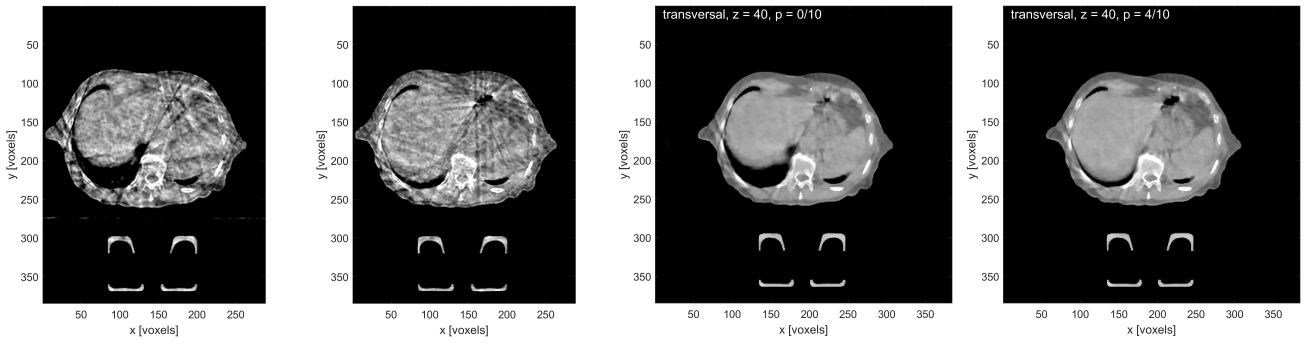


Fig. 1: Transverse slice of the volume $f_{p,\mathbf{i}}$ reconstructed from a simulated dataset based on sinusoidal discrete motion. First half of the breathing cycle is shown. From left to right: p/N_p = 0/10 (left), 2/10 (middle), and 4/10 (right). From top to bottom: ground truth (top), reconstruction (middle), and the difference $|\text{ground truth} - \text{reconstruction}|$ (bottom). WL = 0.22, WW = 0.16 for the top and middle row images. WL = 0.075, WW = 0.075 for the bottom row images.



(a) (b) (c) (d)

Fig. 2: (a)–(b) Transverse slice of the existing clinical reconstruction result (Advanced 4D) at p/N_p = 0/10 and 4/10, respectively. (c)–(d) Corresponding transverse slice of the reconstruction result with the proposed algorithm at p/N_p = 0/10 and 4/10, respectively. WL = 0.22, WW = 0.16 for all images.

Capability of a Novel Head Motion Correction Algorithm for X-ray CT Imaging

Kihong Son, Doil Kim, Seokhwan Jang, and Kyoung-Yong Lee*

***Abstract*— Characteristics of a novel head motion correction algorithm have been explored where a patient motion is virtually transformed into source and detector trajectory of a CT system. Experiments were conducted using the Hexapod for various cases of two dimensional (2D) motions. The sharpness is utilized as a quantitative measure. Results demonstrate that the novel motion correction algorithm delivers significantly higher sharpness on average (proposed: 2889, FBP: 2219, $p < 0.001$). The proposed is capable of correcting motions of amplitude up to 20 mm translations along *x*- and *y*-axes, respectively, and 10° rotation around z-axis. An effective temporal resolution and multi-motion case have also been addressed. The algorithm is accelerated by using GPU and thereby it can delivers 16 slices images in 5 minutes on P5000.**

***Index Terms*—Computed tomography (CT), GPU acceleration, head motion correction, Hexapod, optimization, rigid motion**

I. INTRODUCTION

HEAD motion of patients during CT scan makes a proper diagnosis difficult and it may lead to misdiagnosis or rescans. According to [1], it happens every forth patient during CT perfusion studies. In another study, four of 20 patients showed perceptible motion artifacts [2]. When it comes to pediatric CT scans, sedative drugs are usually administrated to prevent motions, which make the workflow complicated. Especially, brain CT is occupying 75% of the whole pediatric scans [3].

Hardware approaches, for example, using a CT system with a dual source or a fast gantry rotation speed could address the problem of head motion to some extent. Besides the high material costs, however, there exist physical limitations as well.

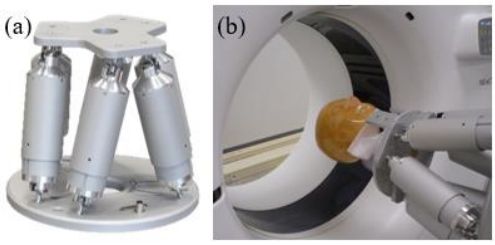


Fig. 1. (a) Hexapod and (b) experimental setup on a prototype CT scanner.

Kihong Son, Doil Kim, and Kyoung-Yong Lee are with Health and Medical Equipment (HME) Business, Samsung Electronics Co., Ltd., 16677 Republic of Korea (e-mail: kihong.son@samsung.com; doil1026.kim@samsung.com; ky71.lee@samsung.com)

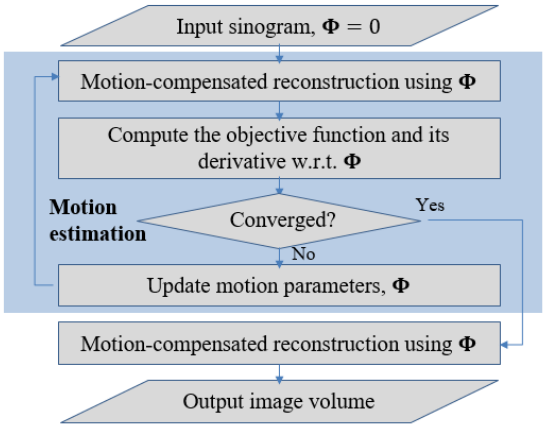


Fig. 2. Overall structure of HMC algorithm. **Φ** denotes motion parameters on translations and rotation defined on a time domain.

Moreover, the hardware approaches are not able to prevent motion artifacts completely. It is obvious that motion artifacts are more prominent in mobile CT scanners having slower gantry rotation speeds.

From these reasons, software approaches that could make the gantry rotation speed faster in a virtual way [4-8] are desired and have been explored. However, one of the major limitations of the software approaches [4-6] is that most of them require a heavy computational cost.

In order to tackle this problem, we recently developed a novel head motion correction algorithm (HMC) [7, 8]. It is designed to be computationally efficient while enjoying a compelling image quality even at a tight display window in contrast to the aforementioned approaches. We demonstrate the capability of HMC by using Hexapod experiments as described in Fig. 1.

II. METHOD

In order to come up with rigid head motion, we developed a software approach transforming a patient motion into a virtual source-detector trajectory [7, 8]. The patient motion is estimated by a sophisticatedly designed optimization process as depicted in Fig. 2. For more details, please refer to [7, 8].

In this paper, we intended to figure out the characteristics of the HMC. To understand its behavior, we imposed various

Seokhwan Jang is with Samsung Advanced Institute of Technology (SAIT), Samsung Electronics Co., Ltd., Suwon 16678 Republic of Korea (e-mail: swan.jang@samsung.com).

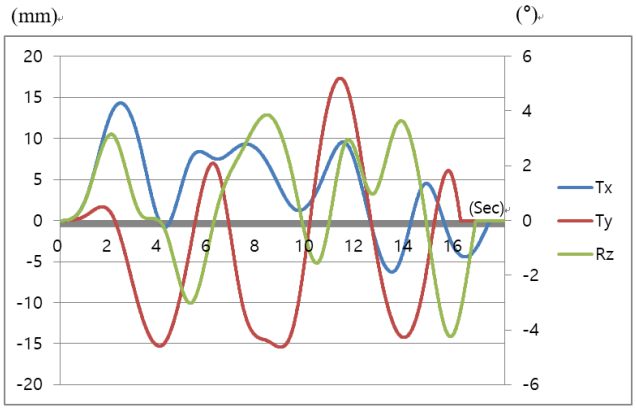


Fig. 3. An example of motion curve generated by maximum amplitude of 20mm for translations along *x*- and *y*-axes, respectively, and 5° for rotation along *z*-axis.

experimental conditions and evaluated the performance in quantitative and qualitative ways, respectively.

A. *Experimental equipment*

A CT Head Phantom ACS (Kyoto Kagaku Co., Ltd, Kyoto, Japan) was attached to the customized Hexapod (H-820, Physik Instrument, Germany). Hexapod is capable of providing motion in six degrees of freedom. In combination with absolute measuring sensors, software and motion controllers make the motion curves. In this procedure, motion along time axis is defined only by maximum amplitudes on rotation and translation. Namely, the motion curves are randomly generated by considering only the maximum preset values of the motion amplitudes as depicted in Fig. 3. Scans were taken by using a 128 slices prototype CT scanner as shown in Fig. 1. We imposed 2D motions (translations along *x*- and *y*-axis, rotation around *z*-axis) to the Hexapod and obtained axial CT scan images under 120 kVp, 100 mA, 44 CTDIvol, 10 mm collimation and 60 revolutions per minute (RPM) of gantry rotation speed. We determined the scan conditions, particularly the rotation speed, in order to simulate mobile CT scanners of Neurologica, namely CereTom, BodyTom, and OmniTom.

B. *Capability of the HMC algorithm*

To figure out the capability and limitations of the HMC algorithm, we set up total 25 cases by combination of translational and rotational movements with a constant speed. Five values of amplitude on translation were defined from 0 to 20 mm with 5 mm interval, and another five levels of rotation were imposed from 0 to 10° with a 2.5° interval. The reconstructed images were obtained by using a conventional filtered back-projection (FBP) algorithm and the HMC with 1.25 mm slice thickness, respectively. The result images were compared by means of a sharpness measure given as below:

$$S = \frac{1}{N_\Omega}\|\nabla f(\mathbf{x})\|_2^2 = \frac{1}{N_\Omega}\sum_{\mathbf{x}\in\Omega}\left\{\left(\frac{\partial f(\mathbf{x})}{\partial x}\right)^2 + \left(\frac{\partial f(\mathbf{x})}{\partial y}\right)^2\right\},$$

where $f(\mathbf{x})$ denotes a reconstructed image and N_Ω denotes the number of voxels in region of interest (ROI), Ω. A one sided t-test was performed to compare the means of the sharpness of FBP and HMC. A p-value less than 0.05 is considered statistically significant.

C. *Behavior of the HMC solution in case of multi-motion*

In order to address the multi-motion problem [4], we placed a carbon-fiber-based table and PVC-based injection line in FOV and we imposed no motion on them. Scans were taken while the static table and injection line were placed together along with the moving head phantom. In addition, a part of the injection lines was attached to the top of the head phantom to evaluate its influence on the result images of the HMC. Total 9 cases were tested with 2D motions of different patterns. The images were reconstructed by using a conventional FBP and the HMC with slice thickness of 0.625 mm, respectively. The severity of the motion artifacts and the existence of side effects of other foreign objects were qualitatively evaluated.

D. *Effective temporal resolution of the HMC algorithm*

To examine an effective temporal resolution of HMC, we set

TABLE I
SHARPNESS VALUES OF FBP AND HMC FOR VARIOUS MOVEMENTS

case	Degree of motion		Sharpness	
	Rotation (°)	Translation (mm)	FBP	HMC
1	0	0	3040	3040
2		5	2174	2988
3		10	2067	2988
4		15	2213	2939
5		20	2014	2966
6	2.5	0	2903	3033
7		5	2275	2796
8		10	2126	2983
9		15	2091	2929
10		20	2065	2751
11	5	0	2712	3010
12		5	2324	2926
13		10	2109	2964
14		15	2037	2856
15		20	1980	2805
16	7.5	0	2609	2942
17		5	2167	2756
18		10	2174	2931
19		15	2103	2841
20		20	2072	2508
21	10	0	2573	2983
22		5	2122	2946
23		10	2133	2973
24		15	2138	2857
25		20	2083	2671
Average			2219	2889
p-value			$p < 0.001$	

up gantry rotation speeds of 60, 120, 180, and 240 RPM and took scans for a linear motion of constant velocity for a fair comparison. We applied the HMC solution only to the raw data acquired from 60 RPM and FBP to all cases. The image quality was assessed by the sharpness measure. Based on the quantities we intend to measure the effective temporal resolution achieved by the HMC.

III. Result

A. Capability of the HMC algorithm

In all 25 cases, the HMC provides significant improvements of image quality in terms of the sharpness (HMC: 2889, FBP: 2219, $p < 0.001$) as presented in Table I. The average improvement by means of sharpness corresponds to 30%. FBP images suffer from typical motion artifacts such as doubling edges around skull and smeared structures in the phantom images (orange arrows) as in Fig. 4. On the other hand, HMC delivers images with a significantly improved quality, except for some residual streaks near the skull (yellow arrow).

B. Behavior of the HMC solution in case of multi-motion

It seems that the influence of other objects undergoing a different motion than the head in the motion correction is negligible. In Fig. 5, the FBP images show severely degraded images (indicated by orange arrow) while the structures of the static table and the injection line are plausible as shown in Figs. 5 (b) and (c). Note here that the injection line attached on the top of the head phantom shows motion artifacts. In addition, the HMC provides a remarkable improvement in visualization of the structure of the head phantom. It needs to be also pointed out that the virtual scan trajectory as a result of the motion estimation makes the static subjects distorted. As shown in Figs. 5 (d) and (e), intensity distortions of the table and injection line (indicated by yellow arrows) are observed. It seems that this kind of side effect is insignificant in the sense that phantom images are at the midpoint of the evaluation and no prominent influence of the other objects is observed inside the phantom images. We leave more thorough tests considering a realistic clinical practice as future work.

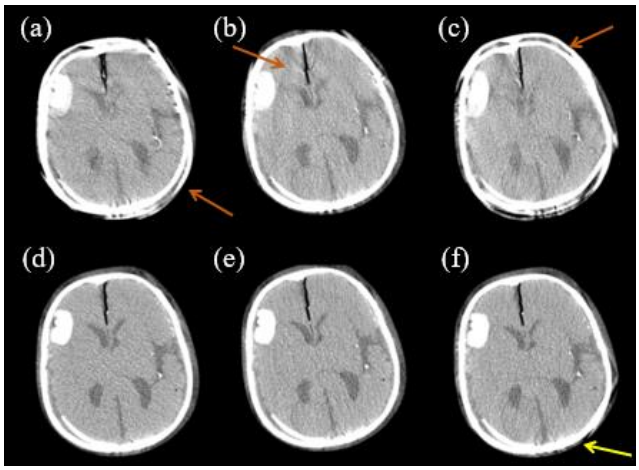


Fig. 4. Images with various linear motions. (a) FBP (translation 20 mm only), (d) HMC (translation 20 mm only), (b) FBP (rotation 10° only), (e) HMC (rotation 10° only), (c) FBP (translation 20 mm and rotation 10°), (f) HMC (translation 20 mm and rotation 10°). L/W 0/200, slice thickness: 0.625mm, FOV 300mm.

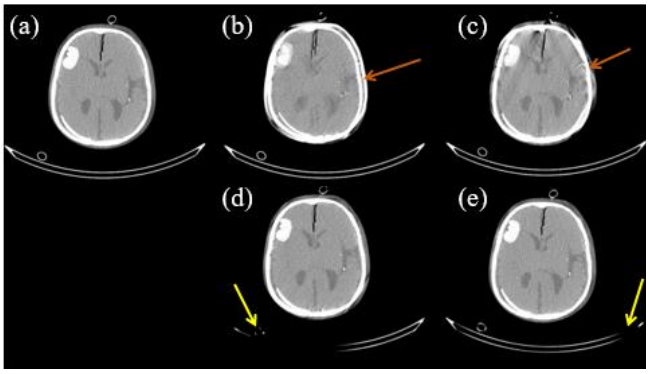


Fig. 5. Images with a patient table and an injection line placed on the table and the moving head phantom. (a) Static, (b) Case 1: FBP 2D motion (Amp: Tx: 7.4 mm, Ty: 14.1 mm, Rz: 1.6°), (c) Case 2: FBP 2D motion (Amp: Tx: 1.5 mm, Ty: 11.6 mm, Rz: 3.9°), (d) Case 1: HMC 2D motion, (e) Case 2: HMC 2D motion. L/W 0/200 HU, slice thickness: 0.625 mm, FOV 300 mm.

C. Effective temporal resolution of the HMC algorithm

The average sharpness of total 48 cases by referring to that of static image shows a drastic improvement from 74.35 to 95.39 after motion correction, as provided in Table II. It is even higher than that of 240 RPM and very close to the static case. Fig. 6 shows the noticeable improvement of the image quality. Note that the degradation due to the linear movement becomes severer as rotation speed is slower, which can be observed in both qualitative and quantitative evaluations. The result of HMC with 60 RPM case is even better than that of 240 RPM, which can be interpreted that HMC is capable of rotating the intrinsic gantry of the CT system more than four times.

IV. Discussion

Although the proposed HMC algorithm seems to be appropriate so as to deliver a compelling image quality compared to FBP algorithm, the realistic head motion could be

TABLE II
NORMALIZED SHARPNESS VALUES OF W/ AND W/O HMC ALONG WITH VARIOUS GANTRY ROTATION SPEEDS

RPM	Normalized Sharpness (%)
Static	100.00
60 w/ FBP	74.35
120 w/ FBP	80.55
180 w/ FBP	81.44
240 w/ FBP	85.17
60 w/ HMC	95.39

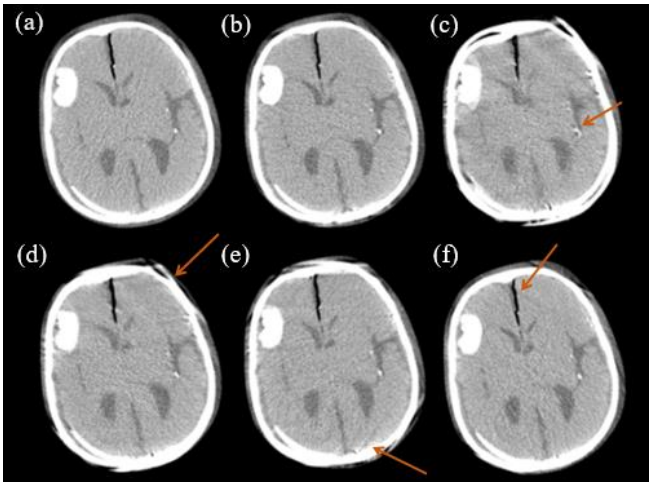


Fig. 6. Images with various gantry rotation speed conditions. (a) static, (b) 60 RPM w/ HMC, (c) 60 RPM w/ FBP, (d) 120 RPM w/ FBP, (e) 180 RPM w/ FBP, (f) 240 RPM w/ FBP (Translations: 15.0 mm/sec, Rotation: 5.0 deg./sec., L/W 0/200 HU, slice thickness: 0.625 mm, FOV 300 mm).

more complicated in its speed and amount [1, 2]. It thereby needs to be evaluated using more clinical cases. We leave this as future work. In addition, test for more challenging cases beyond the speed limit of the Hexapod will let us know the capability of HMC in terms of velocity and the amount of head motion. Moreover, if wider collimation conditions are considered in a CT scan, we expect that there are some rooms to handle 3D motion [8] by using the HMC algorithm.

The solution was accelerated by using GPU and it takes less than 5 minutes for 16 slices on P5000 graphics card. Last but not least, low dose performance of the solution is also important when it comes to pediatric CT scans. We do not know yet, how the motion estimation will perform in case of higher image noise. Therefore, more systematical study is necessary to figure out the limitation of the solution in terms of X-ray exposure. To complement its performance with respect to the low dose scans, noise reduction techniques may be combined with the proposed method. We look forward to conducting clinical studies.

V. Conclusion

We have demonstrated the capability of HMC algorithm through various phantom experiments. The results show a significant improvement of image quality compared to the conventional FBP algorithm (HMC: 2889, FBP: 2219, $p < 0.001$). The patient table and injection lines with different motions (static in this work) seem to have limited influence on the image quality. The novel HMC algorithm can deliver a four times higher effective temporal resolution than that of an intrinsic CT system in case of linear motion.

Acknowledgment

We appreciate Dal Huh, Changlae Lee, and Namwoo Kim for their help for the experiments. This work was funded by Neurologica a subsidiary of Samsung.

References

[1] F. Fahmi *et al.*, "Head movement during CT brain perfusion acquisition of patients with suspected acute ischemic stroke," *Eur. J. Radiol*., vol. 85, no. 12, pp. 2334-2341, Aug. 2013.

[2] A. Wagner *et al.*, "Quantification and Clinical Relevance of Head Motion During Computed Tomography," *Invest. Radiol.*, vol. 38, no. 11, pp. 733-741, Nov. 2003.

[3] J. Vassileva *et al*., "IAEA survey of pediatric CT practice in 40 countries in Asia, Europe, Latin America, and Africa: Part 1, frequency and appropriateness," *Am. J. Roentgenol*., vol. 198, no. 5, pp. 1021-1031, May 2012.

[4] A. Sisniega *et al.*, "Multi-motion compensation for High-Quality Cone-Beam CT of Head," in *Proc. 5th international conference on image formation in X-ray computed tomography*, Salt lake City, UT, USA, 2018, pp. 73-77

[5] T. Sun, J.-H. Kim, R. Fulton, and J. Nuyts., "An iterative projection-based motion estimation and compensation scheme for head x-ray CT," *Med. Phys.*, vol. 43, no. 10, pp. 5705-5716, Oct. 2016.

[6] H. Bruder, C. Rohkohl, K. Stierstorfer, and T. Flohr, "Compensation of skull motion and breathing motion in CT using data-based and image-based metrics, respectively," in *Proc. SPIE Med. Imag.*, vol. 9783, San Diego, CA, USA, 2016, Art. no. 97831E.

[7] K. Son *et al.*, "Workstation, medical imaging apparatus comprising the same and control method for the same," KR patent 1020180072716, Jun. 25 2018.

[8] S. Jang, S. Kim, M. Kim, K. Son, K.-Y. Lee, and J. B. Ra, "Head motion correction based on filtered backprojection in helical CT scanning," *IEEE Trans. Med. Imag*., to be published. DOI: 10.1109/TMI.2019.2953974.

Performance of a Novel Cardiac Motion Correction in CT scans of 0.33 seconds rotation

Junyoung Park, Jiyoung Choi, Duhgoon Lee, Minkook Cho, and Kyoung-Yong Lee*

***Abstract*— Performance of a novel cardiac motion correction algorithm in CT scans of 0.33 seconds rotation was explored. Coronary artery phantoms were installed on a commercially available Mocomo phantom. The moving phantom was scanned under two different heart rate conditions, 60 and 70 beat per minute, on a prototype CT scanner in 0.33 sec. The performance was evaluated by a quantitative measure at three cardiac phases for each bpm case. The results of center-to-edge distance of the coronary arteries (of diameters 3.91 mm and 3.92 mm) shows a significant improvement achieved by the proposed compared to FBP (proposed: 2.40 mm, FBP: 4.96 mm, $p < 0.001$). In addition, the visualization of the landmarks, namely calcified regions, in the coronary artery phantom was qualitatively assessed. Result images show that motion contaminated structures of the coronary arteries by the conventional filtered back-projection were remarkably recovered by the proposed.**

***Index Terms*—Computed tomography, cardiac scan, Mocomo phantom, partial angle reconstruction (PAR), PAR-based motion estimation**

I. INTRODUCTION

X-RAY CT imaging of cardiothoracic organs, such as coronary arteries, myocardial wall, and aorta, is still very challenging. This is because of the dynamic motion of heart structures. For example, coronary arteries have small lumen diameters of 3~5 mm [1] and a fast movement [2] during a cardiac cycle. To maximize the outcome of CT scans, sophisticated workflow for coronary computed tomographic angiography (CCTA) has been established such as administration of beta-blocker or nitroglycerin, adoption of ECG gating. Due to these challenges and the important role of CT exams for the patients having chest pain, advances which could lead to enhance diagnostic accuracy of CCTA are still of great interest.

Reducing scan time by using CT systems with of a fast gantry rotation, a wider detector or dual source is one of the conventional approaches to address the problem. However, these hardware approaches attempting to increase temporal resolution require higher material cost. They are confronting physical limitation as well. Meanwhile, motion correction algorithms have been introduced as a software approach [3, 4].

Junyoung Park, Jiyoung Choi, Minkook Cho, and Kyoung-Yong Lee are with Samsung Electronics Co., Ltd., Suwon-si, Gyeonggi-do 16677 Republic of Korea (e-mail: jun52.park@samsung.com, zzero.choi@gmail.com, minkook.cho@samsung.com, ky71.lee@samsung.com).

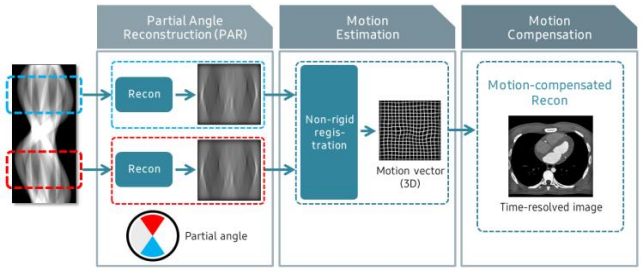


Fig. 1. Conceptual drawing of the SCULLI. Motion estimation and compensation performed based on two conjugate PAR images.

They adopt a processing step of segmentation to extract coronary arteries, which is suboptimal from the workflow point of view.

Recently, a novel motion correction algorithm, sub-cycle universal linear low-dose imaging (SCULLI), has been developed [5, 6]. Two conjugate partial angle reconstruction (PAR) images are utilized for motion estimation via non-rigid registration. The motion between two time points is estimated by a linear weighting of the motion vector field (MVF) which is generated as a result of the motion estimation. A conceptual drawing is described in Fig. 1. For more details please refer to [5, 6]. Different from the aforementioned approaches [3,4], it is fully automatic and free of segmentation. Moreover, it required scan data less than a single rotation which means less x-ray exposure than [3].

Various studies have been conducted and demonstrated the performance of SCULLI [7-11]. However, all the physical experiments so far have been based on 0.25 sec scans of the prototype CT scanner. It is of great interest to explore the performance of SCULLI in other rotation speeds, for example

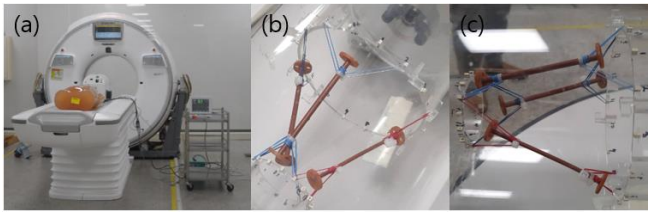


Fig. 2. (a) The experimental setup on the prototype CT scanner. (b)-(c) Mocomo phantom with coronary arteries are installed. The vessel phantom with red rubbers is excluded from the CT scan. Other two vessels were scanned.

Duhgoon Lee is with Samsung Neurologica, MA 01923 USA (e-mail: duhgoon.lee@neurologica.com).

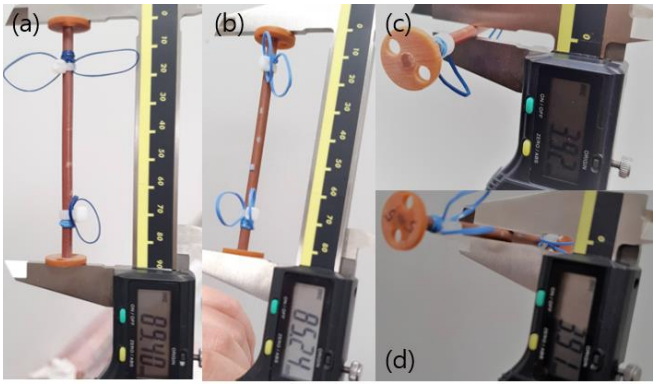


Fig. 3. Two coronary arteries which are installed on the Mocomo phantom. Length and diameter for of each artery are indicated. (a) Vessel for landmark identification. (LM-vessel) (b) Vessel for CED measurement. (CED-vessel) (c)-(d) The corresponding diameters respectively.

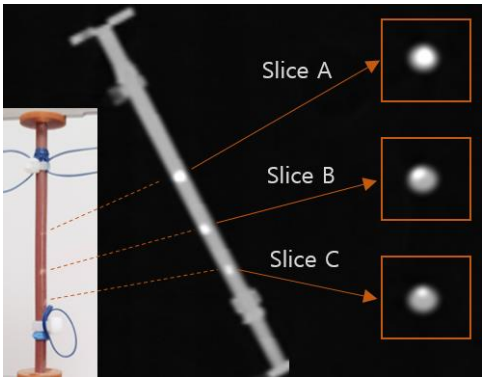


Fig. 4. Multiplanar image of the LM-vessel at a static. The zoomed transaxial image of the three landmarks at three slice locations are displayed in orange boxes. Note the different amount and location of the calcium blocking the lumen of the vessel. L/W -300/1500

0.33 sec, which is still widely adopted in medical CT scanners for cardiac scans. In this paper, we explore the performance of SCULLI under 0.33 sec rotation speed. Both quantitative and qualitative assessments were performed.

II. Method

A. Experimental setup

A physically moving phantom, Mocomo (Mocomo 1, Fuyo, Japan) was utilized for the experiment. It is dedicated for mimicking motion of coronary arteries. The rotational (15 degrees) and translational (20 mm in Z-direction) movements simulate the motion of the heart. It also generates ECG signals, which makes ECG gated cardiac scans possible. Two vessel phantoms (Gammex, USA) having 3.91 mm and 3.92 mm diameters respectively were installed in the Mocomo phantom. Each of these phantoms has calcium inserts of different sizes as shown in Fig. 3 and Fig. 4. They were tied with rubber bands to the hooks of the plastic case of the Mocomo phantom as in Fig. 2. The vessel in Fig. 3 (a) is for landmark identification, in short we call it LM-vessel. The one in Fig. 3 (b) is for center-to-edge distance (CED) measurement, so we call it CED-vessel.

We used a standard prospective ECG gating protocol with 5~10% padding option for the experiment. Due to the limited length of the vessel phantoms (8.94 cm and 8.52 cm respectively) only a single axial scan was taken with 128 × 0.625 mm slice collimation. Two heart rates, 60 and 70 BPMs, were imposed to the Mocomo phantom. In case of 70 BPM, the padding option and the target phase were slightly modified due to a limitation of the system. Detailed scan conditions are provided in Table I.

TABLE I
SCAN CONDITIONS OF THE MOCOMO PHANTOM

BPM	Phase (R-R %)	Target (R-R%)	kVp	mA	Rotation speed
60	35~45	40	120	300	0.33 sec
	55~65	60			
	70~80	75			
70	35~45	40			
	55~65	60			
	70~75	73			

B. Image reconstruction and performance evaluation

Both FBP and SCULLI were carried out for all six cases. Images were reconstructed with 0.625 mm slice thickness and 200 mm FOV. We adopted the center-to-edge distance [12] as a quantitative measure. Three imaging engineers having experiences of 9 years, 15 years, and 18 years respectively, measured it at three slice locations of the CED-vessel in Fig. 3 (b). These locations were identified by the three landmarks of LM-vessel as described in Fig. 4. CED was measured manually by the "Straight" tool (freehand lines) in ImageJ software (NIH) under fixed display window setting, namely level/width -275/1450 HU. Moreover, the images were zoomed 400% for the measurement. Average of the three measurements was

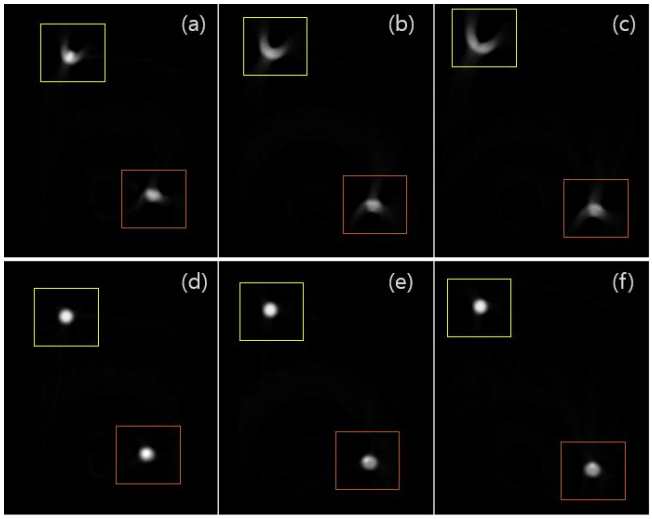


Fig. 5. Example of three slice locations for image quality assessment. 70 BPM, -73% case. (a), (d): Results of FBP and SCULLI at slice A. (b),(e): Results of FBP and SCULLI at slice B. (c), (f): Results of FBP and SCULLI at slice C. CED-vessels are indicated by the yellow boxes and LM-vessel in orange boxes. L/W: -275/1450 HU

TABLE II
CED OF THE CED-VESSEL AT VARIOUS SCAN CONDITIONS.

		60 BPM			70 BPM			Average	p-value
		40%	60%	75%	40%	60%	73%		
Slice A	FBP	2.5	2.6	4.6	4.1	3.6	5.6	**3.8**	**0.013**
	SCULLI	2.2	2.4	2.4	2.3	2.4	2.4	**2.3**	
Slice B	FBP	4.3	4.2	4.7	5.7	4.8	6.6	**5.1**	**0.001**
	SCULLI	2.4	2.3	2.4	2.3	2.4	2.4	**2.4**	
Slice C	FBP	4.9	4.8	5.9	6.3	5.4	8.5	**6.0**	**0.001**
	SCULLI	2.4	2.4	2.4	2.4	3.1	2.4	**2.5**	

Unit: mm

calculated and compared. A single sided paired t-test was performed between the average values from FBP and SCULLI. A value of $p < 0.05$ was considered statistically significant. Note that the vessels are obliquely tied, so the transaxial images of them are close to an elliptical shape rather than a circular. In addition, each vessel phantom has a different degree of motion artifacts depending on the location due to the different turning radius. Based on the installation of the vessels, we have the following order of the turning radii:

Radius of Slice A < Radius of Slice B < Radius of Slice C.

On the other hand, the LM-vessel has three calcium insertions inside as described in Fig. 4. We make use of them as landmarks. The visualization of each landmark is also evaluated in a qualitative way. Based on the reference images of LM-vessel acquired in a static state of the Mocomo phantom, we assess the location and amount of the calcium at three slice locations.

III. RESULT

A. *CED measurement*

CED values of SCULLI show a significantly improved performance. First, the average of all 18 CED measurements of SCULLI is remarkably reduced compared to FBP (SCULLI: 2.40 mm, FBP: 4.96 mm, $p < 0.001$). As shown in Table II,

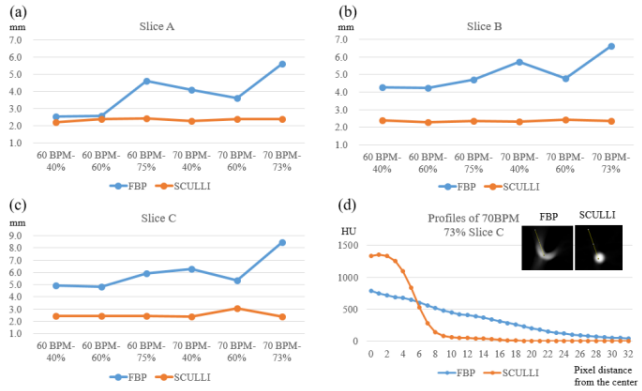


Fig. 6. (a)-(c) CED values of at three slice locations. SCULLI demonstrates a consistent performance throughout the heart rates and cardiac phases. (d) Line profile of the case of 60 BPM 40% at slice C shows the sharper drop HU values by SCULLI.

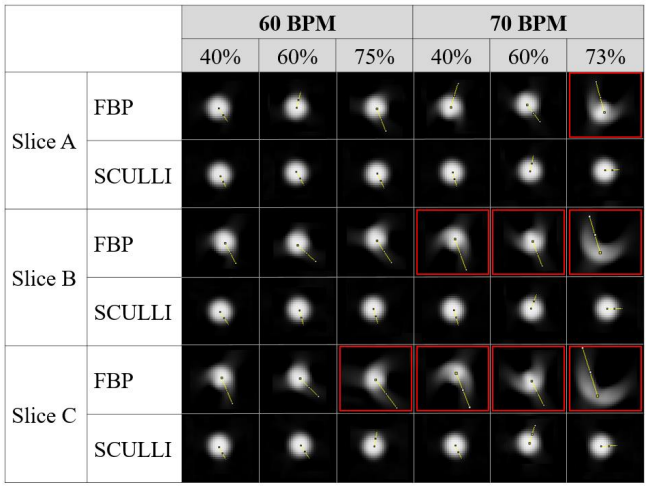


Fig. 7. Zoomed images of the CED-vessels for all cases. Yellow lines indicate that the direction along which the distance was measured by one of three evaluators. Red boxes are for images where the lumen of the vessel is severely distorted. SCULLI delivers a consistent lumen structure unlike FBP. L/W: -275/1450 HU

CED values of FBP fluctuate violently depending on the location, heart rate, and cardiac cycles, while that of SCULLI is remarkably consistent throughout all conditions. For each slice location, SCULLI delivers significantly more accurate estimations than FBP. This tendency is well displayed in Fig. 6. The result of SCULLI stays tightly at around 2.4 mm, which is very close to the ground truth taking into account the oblique positioning and some blurring. In contrast, most of the values of FBP reaches over 4.0 mm which is a huge deviation from the ground truth. Seemingly the most challenging case is 70 BPM-73%, whose result images are displayed in Fig. 5. The visualized CED-vessels of FBP show extremely severe motion artifacts which make the measurement very erroneous. On the other hand, SCULLI shows a consistent lumen shape of the vessel regardless of the location. The line profiles of the vessel at slice C in Fig. 6 (d) show the clear difference of the sharpness. Zoomed images which were utilized for the real measurement task by an evaluator are provided in Fig. 7. The drastic difference of the performance between FBP and SCULLI is well appreciated.

B. *Visualization of the landmarks*

The visualization of the three landmarks were assessed qualitatively. As indicated in Fig. 4, the three slice positions, namely Slice A, B, and C, have different patterns of calcium insertions. The Slice A is totally blocked, Slice B partially at 11 o'clock, and Slice C minimally at 12 o'clock.

As described in Fig. 8, SCULLI makes the identification of the landmarks almost perfect. In all cases, the result images allow the location and amount of the calcium to be identified exactly. Unlike this, a complete loss of the structural information is observed in around 50% of FBP images as marked by the red boxed in Fig. 8.

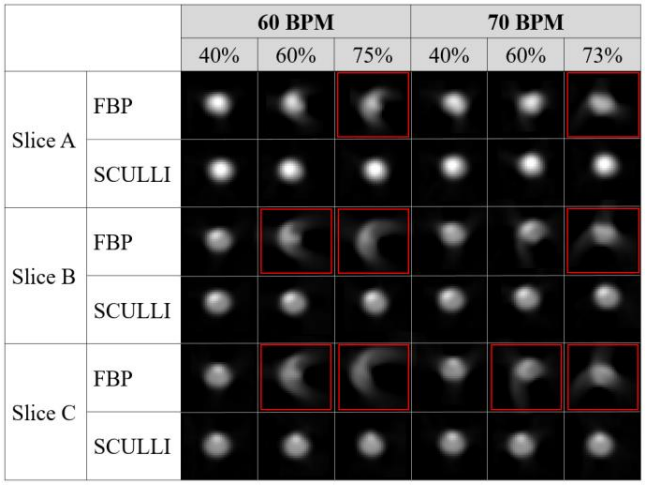


Fig. 8. Zoomed images of the three landmarks, Slice A, B, and C. The existence and the location of the calcium is noticeable only few cases in FBP images. Red boxes indicate a complete loss of the landmarks. In contrast, the landmarks are identified in all results of SCULLI. L/W: -275/1450 HU

IV. Discussion

There are some limitations of this study. First, although the movement of the Mocomo phantom is dynamic enough, which generates turbulent motion artifacts in FBP images, its pattern is limited by the mechanical movement. A thorough clinical study is necessary in order to find out the limitations of SCULLI. We are looking forward to conducting a cohort study covering various patient groups having different degrees of stenosis or heart rates. In addition, we need to come up with a quantitative measure regarding the identification of the landmarks. Furthermore, heart rates of only 60 and 70 BPMs have been tested in this paper. Studies on even higher heart rates are the next step.

The performance of the SCULLI for even slower gantry rotation speeds is also of great interest. Figuring out the effective temporal resolution achieved by SCULLI out of various rotation speeds would be a very interesting research topic.

Last but not least, if the consistent performance of SCULLI throughout all heart cycles is maintained even for clinical data, it could open the door to innovations such as cardiac CT scan free of ECG-gating.

V. Conclusion

We have explored the capability of SCULLI in terms of a moving phantom experiment. The results demonstrate an excellent performance of SCULLI for 0.33 sec rotation, in which the novelty of this paper lies. SCULLI provides a consistent CED values throughout the locations the vessel, heart rates, and heart cycles. In addition, SCULLI recovers the important structural information which was lost in FBP images due to motion.

This result implies that SCULLI has a great potential to shed light on disruptive application in cardiac CT imaging. We are looking forward to conducting more clinical studies.

Acknowledgment

We appreciate May Kim and Giselle Choi for their help for the experiments.

References

[1] J.T. Dodge, B.G. Brown, E.L. Bolson, and H.T. Dodge, “Lumen diameter of normal human coronary arteries,” *Circulation* **86**, 232-246, 1992

[2] M.B. Hofman, S.A. Wickline, and C.H. Lorenz, “Quantification of in-plane motion of the coronary arteries during the cardiac cycle: implications for acquisition window duration for MR flow quantification,” *JMRI* **8**(3):568-576, 1998

[3] R. Bhagalia, J.D. Pack, and J.V. Miller, “Non-rigid registration-based coronary artery motion correction for cardiac computed tomography,” *Medical Physics* **39(7)**, 4245–4254, 2012

[4] J. Hahn, H. Bruder, T. Allmendigner, K.Stierstorfer, T. Flohr, and M. Kachelries, “Reduction of motion artifacts in cardiac CT based on partial angle reconstructions from short scan data,” *SPIE Medical Imaging Conference Record* **97831A**:1-9, 2016

[5] J.B. Ra *et al.*, “Tomography apparatus and method for reconstructing tomography image thereof,” US9959631, 2018

[6] S. Kim, Y. Chang, and J.B. Ra, “Cardiac motion correction based on partial angle reconstructed images in x-ray CT,” *Medical Physics* **42**(5), 2560-2571, 2015

[7] J. Yi, J. Jung, D. Lee, K.Y. Lee, and S. Dutta, “Evaluation of systolic and diastolic phase motion artifacts at various heart rates during ECG gated CT acquisitions with a novel reconstruction algorithm,” RSNA 2015, PH232-SD-MOB4 (RSNA ID: 15009043)

[8] J. Jung *et al.*, “Improvement in stenosis measurement accuracy in presence of cardiac motion using a novel temporal enhanced volume reconstruction technique,” SCCT 2015, Poster #163

[9] J. Jung *et al.*, “Estimation of effective temporal resolution from high heart rate ECG gated images with a novel temporal enhanced volume reconstruction technique,” SCCT 2015, Poster #44

[10] J. Jung *et al.*, “Assessment of calcium quantification in presence of motion with a novel temporal enhanced reconstruction technique,” ECR 2016, Poster C-1084

[11] D. Lee, J. Choi, K.Y. Lee, J. Shin, and D. Kim, “Practical simulation with realistic coronary artery motion for validation of motion compensated reconstruction algorithm in CT,” ECR 2017, Poster C-1701

[12] S. Dutta and D. Okerlund, “Cardiac CT advancement – Evolution of snapshot freeze technology,” GE, white paper

High-Fidelity CT Reconstruction Based on an Unsupervised Gaussian Mixture Model

Danyang Li, Dong Zeng, Sui Li, Zhaoying Bian, Jing Huang, and Jianhua Ma

***Abstract*—Supervised learning networks show a great potential in computed tomography (CT) imaging field. They usually requires large-scale datasets with reliable and clean CT images to train the network. However, collecting large-scale CT dataset is time-consuming and expensive. In addition, the performance of the pre-trained network highly relies on the similarity between the distribution of the training datasets in the source domain and testing datasets in the target domain, and this is known as the problem of domain adaption. In this work, to address the issues of pre-collection large scale of labeled data and domain adaption, we propose a novel high-fidelity CT reconstruction method based on an unsupervised Gaussian mixture model (GMM). Specifically, at first a supervised network is trained based on a certain number of training datasets, and then the network is transferred to the target datasets by learning the distribution shift between the training and test datasets in an unsupervised manner with a small amount of test datasets. Moreover, an explicit physical model of the test datasets is incorporated into the network layers and the GMM is introduced to characterize the information from the distribution shift between the training and test datasets to promote the unsupervised network learning performance. Finally, the unsupervised CT reconstruction can be achieved by learning the network weights according to the physical model and GMM of the test data, and we call this method as GMM based transfer learning network (GMM-TLNet). Numerical studies are conducted to validate and evaluate the performance of the presented GMM-TLNet.**

***Index Terms*—CT, supervised learning, unsupervised learning, Gaussian mixture model, physical model.**

I. Introduction

THE increase in computed tomography (CT) examination has risen concern about the increased cancer risk from medical radiation exposure [1]. Various techniques are developed to reduce the radiation dose including lowering tube current and reducing the number of projection views. However, without adequate control procedures, the CT images reconstructed by the convolutional filtered backprojection (FBP) method would suffer from the severe noise and artifacts from these two techniques.

Recently, deep learning (DL) based methods have been widely used in the CT imaging field, and achieve promising reconstruction performance [2]–[5]. However, there are still some intrinsic limitations in the DL-based methods. First, most of them are supervised-based and need to pre-collect a large quantity of low-dose/high-dose CT image pairs, i.e. training inputs and labels. However, due to the radiation dose and ethical review, it is quite time-consuming and cumbersome to attain sufficient training sample pairs as expected. In addition, it is difficult to obtain well-registered data pairs, especially for the torso region. And the issue of structure mismatching between low-dose and high-dose CT image pairs should be addressed before network training. Additionally, most of the training datasets for current supervised-based DL methods are generated based on the simulation study, and the simulation study can not represent real CT scanning procedure. And the image-based networks usually assume the noise in the low-dose CT images follows simple Gaussian distribution, and in practice the noise distribution is much more complex. This might lead to over smoothed results in the CT image reconstruction tasks. Therefore, in this work, we present Gaussian mixture model to model the complex noise distribution in the low-dose CT images, and this is one contribution in this work.

Second, the pre-collected datasets are typically available for a certain reconstruction task instead of variety, bias relating to supervision, and the probabilities of DL-based reconstruction methods take into consideration of other corner cases during problem solving [6]. The CT images from different vendors often vary in scanning protocols and reconstruction algorithms, leading to different noise distributions and artifact characteristics in the CT images. This is known as the problem of domain adaption in the DL application where a network trained on one dataset (source domain) is difficult applied to a new situation (target domain). Although transfer learning strategy can be used to fine-tune the per-trained network to learn features and distributions of the data in the target domain, the labeled dataset in target domain are not always easily to collect. It should be noted that lots of the low-dose CT images only (or equivalently unlabeled CT dataset) can be enrolled into the network training, because abundant information in the unlabeled CT datasets are much easier to obtain and they represent more general real scenarios. Based on this observation, in this work, we use unsupervised learning strategy to solve the issue of domain adaption, and this is another contribution in this work.

In this work, we propose a novel high-fidelity CT reconstruction method based on an unsupervised GMM [7], which transfers the network from source domain to target domain where only unsupervised dataset are used, and we call this method as GMM based transfer learning network (GMM-TLNet). Fig. 1 shows the Architecture of the presented GMM-TLNet for low-dose CT image reconstruction. Specifically, a supervised network is firstly trained based on a certain number

This work was supported in part by NSFC under Grants U1708261, 81701690, 61701217 and 61571214. *(Corresponding author: Jianhua Ma.)*

D. Li, S. Li, Z. Bian, and J. Huang are with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China.

D. Zeng is with College of Automation Science and Engineering, South China University of Technology, Guangzhou, 510641, China.

J. Ma is with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China (e-mail: jhma@smu.edu.cn).

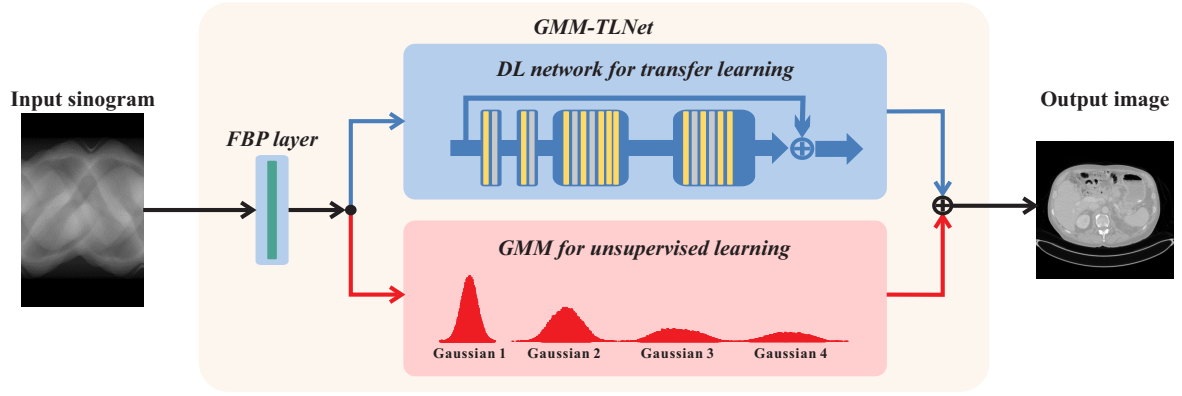


Fig. 1. Architecture of the presented GMM-TLNet for low-dose CT image reconstruction. The pipeline with blue arrows denotes the supervised DL network trained with the labeled CT images in the source domain. The pipeline with red arrows denotes the noise characterized by the GMM in the unsupervised DL network. At the unsupervised training stage, only low-dose measurements are used.

of training datasets, and then the network is transferred to the target datasets by learning the distribution shift between the training and test datasets in an unsupervised manner with a small amount of test datasets. Moreover, an explicit physical model of the test datasets is incorporated into the network layers to indo the model reconstruct CT images from sinogram and the GMM is introduced to characterize the information from the distribution shift between the training and test datasets to promote the unsupervised network learning performance. Finally, the unsupervised CT reconstruction can be achieved by learning the network weights according to the physical model and GMM of the testing data. Experiments demonstrate that the presented GMM-TLNet can deliver considerably better performance compared to the existing supervised DL network.

In summary, the main contributions of the presented model are:

- We present an unsupervised DL network with a small amount of unlabeled CT images to address the issue of domain adaption in low dose CT imaging field.
- We further introduce a GMM into the network construction to characterize the complex nose distribution in the CT images instead of simple Gaussian distribution.
- The advantages of the presented GMM-TLNet are evaluated with various assessments. Experimental results demonstrate that the presented GMM-TLNet achieve competitive performance compared to the other methods.

II. Methods

Assuming independent measurements, the CT imaging model can be expressed as follows:

$$y = Ax, \tag{1}$$

where y is the CT sinogram, x is the desired CT image, and A denotes the linear projection operator. In addition, to further improve the image quality of CT reconstruction, image postprocessing techniques are usually adopted. Among then, the DL-based techniques can obtain promising performance with high efficiency, and they can be expressed as follows:

$$\hat{x} = \emptyset\Big(\left(A^T A\right)^{-1} A^T y, \boldsymbol{\theta}\Big), \tag{2}$$

where $\emptyset$ represents the network with parameters $\boldsymbol{\theta}$, $\hat{x}$ is the predicated image, and $\left(A^T A\right)^{-1} A^T y$ can be the FBP reconstruction image as the network input. $\left(A^T A\right)^{-1}$ corresponds to an ideal ramp filter operation and A^T is a back-projection operator [8]. In the supervised DL-based methods, the parameters $\boldsymbol{\theta}$ are optimized by minimizing the loss function between the label x^{label} and the predicated one $\hat{x}$ as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} L\left(\hat{x}, x^{label}\right), \tag{3}$$

and the loss function can be mean square error (MSE) loss, L1 loss, and negative structural similarity (SSIM) loss.

In the supervised DL-based models, the problem of domain adaption sometimes exists, leading to degraded performance. The main reason is that there is always a distribution shift between the training datasets in the source domain and the test data in the target domain. For example, the network trained on the images at 140 kVp might fails to reconstruct the images at 80 kVp due to the attenuation coefficients vary between the two kVps. It should be noted that transfer learning techniques can be used to address the issue of domain adaption. Most of them focus on the test datasets with a small amount of labeled data. Meanwhile, the labeled data in the CT imaging field are usually generated artificially and lots of unlabeled low-dose CT data fail to be included in the network training. Therefore, in this work, we enroll the unlabeled low-dose CT sinogram only in the transfer network training in an unsupervised manner as follows:

$$\mathcal{L}_{sinogram} = \left\| A\emptyset\Big(\left(A^T A\right)^{-1} A^T y, \boldsymbol{\theta}\Big) - Y \right\|_F^2 \tag{4}$$

where $Y = [y_1, y_2, ..., y_N]$, N is the number of the unpaired training batch size. It should be noted that Eq. (4) is used to describe the distribution shift between the training and test datasets, and CT imaging model is added in the training process.

In addition, the conventional DL-based techniques in CT image reconstruction task usually assume the noise in the CT image usually follows simple Gaussian distribution. However, the noise distribution of CT image in actual situation is very complicated far from simple Gaussian distribution. Therefore,

in this work, we utilized GMM to characterize the noise distribution in the unlabeled CT images as follows:

$$\mathcal{L}_{GMM}(\varepsilon;\mathbf{\Pi},\mathbf{\Sigma}) = -\sum_{n=1}^{N}\log\sum_{k=1}^{K}\pi_k\mathcal{N}\left(\varepsilon|\mu_k,\Sigma_k\right) \quad (5)$$

where K is the number of the Gaussian components, π_k is the mixture coefficient of the kth Gaussian component, where $\pi_k > 0$, and $\sum_{k=1}^{K}\pi_k = 1$. μ_k and Σ_k are the mean and variance of the Gaussian component, respectively.

In this work, the total variation regularization is added to preserve the structure details. Therefore, the total loss function can be expressed as follows:

$$\begin{aligned}\mathcal{L}_{unsupervise} &= \mathcal{L}_{sinogram} + \mathcal{L}_{GMM} + \mathcal{L}_{TV} \\ &= \left\|A\emptyset\left(\left(A^TA\right)^{-1}A^Ty,\boldsymbol{\theta}\right) - Y\right\|_F^2 \\ &- \alpha\sum_{n=1}^{N}\log\sum_{k=1}^{K}\pi_k\mathcal{N}\left(\varepsilon_n|\mu_k,\Sigma_k\right) \\ &+ \beta\left\|\emptyset\left(\left(A^TA\right)^{-1}A^Ty,\boldsymbol{\theta}\right)\right\|_{TV}\end{aligned} \quad (6)$$

where $\varepsilon_n = \left(A^TA\right)^{-1}A^Ty_n - \emptyset\left(\left(A^TA\right)^{-1}A^Ty_n,\boldsymbol{\theta}\right)$ represents the noise learned by the network, α and β are the trade-off parameters. This method can be termed as GMM based transfer learning network (GMM-TLNet) in this work.

The parameters in the presented GMM-TLNet mainly contain two parts, i.e. parameters in the GMM and in the network. Specifically, the parameters in the GMM can be optimized by the EM algorithm [9] and the parameters in the network can be tuned by the Adam [10]. The whole procedure for optimization of Eq. (6) can be summarized in Algorithm 1.

Algorithm 1 Algorithm for Solving Eq. (6)

Input: Dataset $\mathbf{\Omega}$, Per-trained Net $\emptyset_{pre-trained}$.
Output: Unsupervised network $\emptyset_{unsupervise}$.
1: $\emptyset_{unsupervise} = \emptyset_{pre-trained}$.
2: **while** not convergence **do**
3: Update GMM parameters by EM algorithm [9].
4: Update $\emptyset_{unsupervise}$ parameters by Adam [10].
5: **end while**

III. Results

A. Implementation details

The dataset authorized for *"2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge"* by Mayo Clinic [11] are used. In experiments, we randomly selected 2000 image pairs (low-dose images/high-dose ones) at 120 kVp as the training dataset for network training. The trained network is defined as "Pre-trained Net". In the network training, learning rate is $1e^{-4}$, batch size is 25, and training epoch is 100. The perceptual loss based on VGG-19 network [12] and MES loss are selected as the loss function in the network.

With consent for chest and abdomen CT studies, 100 CT images at normal dose with 140 kVp were obtained by a commercial CT scanner. The low dose projection data are simulated based on the previous study [13], and they serve as testing dataset used for the presented GMM-TLNet training. In the network training, learning rate is $1e^{-5}$, batch size is 2, and training epoch is 100. The parameter α and β is empirically set to be 1 and $5e^{-5}$, respectively. In the experiment, the residual network [14] is selected as the backbone network for the Pre-trained Net and the presented GMM-TLNet. Both the networks are implemented in Python with PyTorch package on a NVIDIA Tesla K40c GPU.

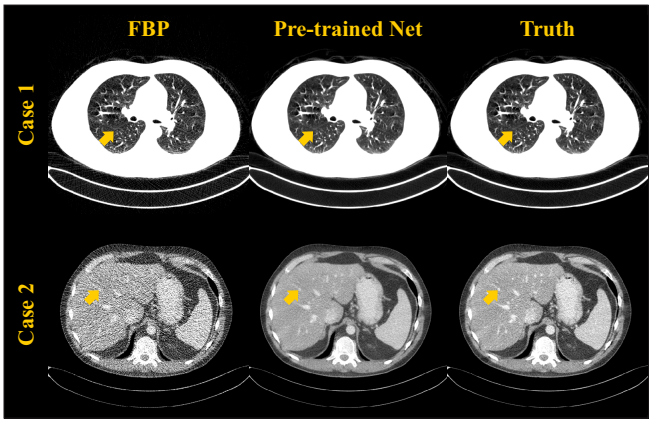


Fig. 2. Low dose reconstruction results on the Mayo dataset.The display window for Cases 1 and 2 are [-1000, -500] and [-300, 50] HU, respectively. "FBP" is the reconstructed images by filtered-backprojection (FBP) algorithm at quarter dose, "Pre-trained Net" is the denoise results of the Pre-trained Net, "Truth" is FBP reconstructed images for normal dose.

B. Performance of the Pre-trained Net

Fig. 2 illustrates the low dose reconstruction results on the Mayo clinic dataset. Rows 1 and 2 show the chest and abdomen CT images, respectively. It can be observed that the Pre-trained Net trained on the Mayo dataset can effectively remove the noise-induced artifacts and recover the images details at the case of Mayo clinic dataset, as indicated by the yellow arrows. Table I illustrates the quantitative measurements, i.e., peak signal-to-noise (PSNR) and SSIM, and it also demonstrates the reconstruction performance of the Per-trained Net on the Mayo clinic dataset. This experiment demonstrates the training and test data from the same dataset can help the network yield promising reconstruction results.

TABLE I
Quantitative results on the Mayo dataset at 120 kVp.

	PSNR	SSIM
FBP	30.24 ± 3.24	0.8401 ± 0.0648
Pre-trained Net	$\mathbf{41.26 \pm 2.69}$	$\mathbf{0.9820 \pm 0.0104}$

C. Qualitative analysis

Fig. 3 shows visual comparisons of the different methods for the chest and abdomen CT images at low dose with 140 kVp. One selective ROI by the red box at each case is zoomed in for better visual inspection. The first row shows the chest CT images reconstructed by the different methods wherein the image at normal dose is chosen as the ground truth. The second row shows the corresponding difference images. The

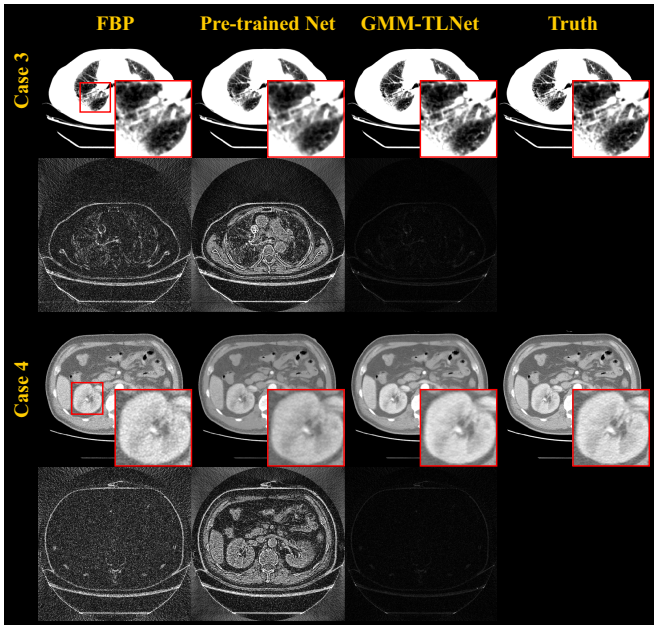


Fig. 3. Low dose reconstruction results on the 140 kVp spectrum dataset. Row 1 and 3 show the reconstructed images by FBP, Pre-Trained Net, the presented GMM-TLNet and ground truth, respectively. The display windows for Case 3 and 4 are [-800, -300] and [-300, 50] HU, respectively. One selective ROI is displayed for better visualization. Row 2 and 4 show the corresponding difference images, and the display window is [0, 100] HU.

third row shows the abdomen CT images reconstructed by the different methods and the forth row shows the corresponding difference images. It can be seen that the Pre-trained Net trained on Mayo dataset fails to recover the low dose images at 140 kVp. For example, at Case 3, the presented GMM-TLNet successfully recover the structure details in the lung region as shown in the zoomed-in ROI, while the Pre-trained Net is prone to produce blurry results. In the Rows 2 and 4, we can see that the presented GMM-TLNet can reconstruct more accurate CT images than the Pre-trained Net wherein the CT values in Pre-trained Net results shift a lot. The main reason is that the Pre-trained Net is trained on the Mayo Clinic dataset, and there is lots of distribution shift between the Mayo Clinic dataset at 120 kVp and test dataset at 140 kVp. Meanwhile, with the physical model and GMM of the test dataset, the presented GMM-TLNet can effectively transfer the network learning from Mayo Clinic dataset in the source domain to the test dataset at 140 kVp in the target domain, and it can not only remove the noise and artifacts, but also preserve structure details and avoid value shifting.

TABLE II
QUANTITATIVE MEASUREMENTS ON THE RECONSTRUCTION RESULTS AT 140 KVP FROM THE DIFFERENT METHODS

	PSNR	SSIM
FBP	28.45 ± 5.63	0.7140 ± 0.0421
Pre-trained Net	27.91 ± 7.79	0.7174 ± 0.1243
GMM-TLNet	$\mathbf{39.18 \pm 3.41}$	$\mathbf{0.9845 \pm 0.0924}$

D. Quantitative analysis

In order to quantify the reconstruction performance of different methods, PSNR and SSIM are also utilized. Table II lists the quantitative measurements on the reconstruction results at 140 kVp from the different methods. From the results, it can be seen that the presented GMM-TLNet achieves better results in terms of both PSNR and SSIM, compared with Pre-trained Net and FBP algorithm.

IV. DISCUSSION AND CONCLUSION

DL based methods have been widely used in CT imaging and produce promising reconstruction performance. However, most of them are supervise-based and still have some still some intrinsic limitations such as the pre-collection of large quantity training dataset and issue of domain adaption. To address those issues, in this work, we presented a novel high-fidelity CT reconstruction method based on an unsupervised GMM called "GMM-TLNet". Specifically, GMM-TLNet first initialized by a supervised network, then, the network is transferred to the target datasets in an unsupervised manner with a small amount of test datasets. Moreover, an explicit physical model of the test datasets the GMM is introduced to characterize the information from the distribution shift between the training and test datasets to promote the unsupervised network learning performance. Simulation studies are conducted to demonstrate the practical utility of the presented GMM-TLNet method. In the future, more patient and scene studies would included to further demonstrate the transfer learning performance of the presented GMM-TLNet.

REFERENCES

[1] Y. Sagara *et al.*, "Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients," *American Journal of Roentgenology* vol: 195, no. 3, pp. 713-719 2010.
[2] K. Jin, M. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Transactions on Image Processing*, vol. 26, pp. 4509-4522, 2017.
[3] Y. Wang, Y. Liao, Y. Zhang, J. He, S. Li, Z. Bian, H. Zhang, Y. Gao, D. Meng, W. Zuo, D. Zeng, and J. Ma, "Iterative quality enhancement via residual-artifact learning networks for low-dose CT," *Physics in Medicine and Biology*, vol. 63, p. 215004, 2018.
[4] J. Adle and O. Oktem, "Learned Primal-Dual Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 1322-1332, 2019.
[5] J. He, Y. Yang, Y. Wang, D. Zeng, Z. Bian, H. Zhang, J. Sun, Z. Xu, and J. Ma, "Optimizing a Parameterized Plug-and-Play ADMM for Iterative Low-Dose CT Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 38, pp. 371-382, 2019.
[6] K. Raza, and N. Singh, "A tour of unsupervised deep learning for medical image analysis," *arXiv preprint* arXiv:1812.07715, 2018.
[7] G. J. McLachlan and K. E. Basford, "Mixture Models: Inference and Applications to Clustering," Marcel Dekker, 1988.
[8] G. L. Zeng, "Medical image reconstruction: a conceptual tutorial," Springer, 2010.
[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (methodological)*, pp. 1-38, 1977
[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint*, arXiv: 1412.6980, 2014
[11] AAPM. (2017). *Low Dose CT Grand Challenge*. [Online]. Available: http://www.aapm.org/GrandChallenge/LowDoseCT/#.
[12] K. Simonyan and A. Zisserman, (2014). "Very deep convolutional networks for large-scale image recognition," [Online]. Available: https://arxiv.org/abs/1409.1556.
[13] D. Zeng, J. Huang, Z. Bian, *et al.*, "A simple low-dose x-ray CT simulation from high-dose scan," *IEEE Transactions on Nuclear Science* vol. 62, no. 5, pp. 2226-2233, 2015.
[14] K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

Self-supervised Reconstruction of Cerebral Perfusion CT Images based on Noise Properties

Sui Li, Manman Zhu, Danyang Li, Qi Gao, Zhaoying Bian, Dong Zeng, and Jianhua Ma

***Abstract*—Cerebral perfusion computed tomography (CPCT) introduces high radiation exposure to patients due to the repeated dynamic sequential scans. Lowering radiation dose would increase the amount of noise and hence deteriorates diagnosis accuracy. The advanced noise reduction methods can suppress the CPCT noise significantly. However, some of them assume that the noise is independent and identically distributed, and the noise distribution might be much more complicated in practice. In this work, we first investigate the noise properties of CPCT images, and find that (1) the noise in each frame is complex far from a simple distribution. (2) The noise in different frames is obviously different from each other. Based on the observations, we present a self-supervised CPCT image reconstruction (SS-CPCT) method to produce high-quality CPCT images from low-dose measurements. Different from the conventional supervised network, the SS-CPCT method utilizes the low-dose CPCT measurements only for network training. Real patient data are used to evaluate the presented SS-CPCT method, and the corresponding experimental results demonstrate that the presented SS-CPCT method can effectively reduce the noise and improve spatial resolution based on the noise properties of the CPCT images.**

***Index Terms*—Cerebral perfusion computed tomography, radiation dose, noise properties, self-supervised learning, image reconstruction.**

I. Introduction

CEREBRAL perfusion computed tomography (CPCT) is an imaging technique that allows functional evaluation of tissue vascularity with hemodynamic parameter maps, i.e., cerebral blood flow (CBF), cerebral blood volume (CBV) and mean transit time (MTT) [1]. Meanwhile, in the CPCT scanning, the protocol involves repeated three-dimensional acquisitions on the selective anatomical coverage in a cine mode and the CPCT imaging is associated with high radiation dose which might increase potential cancer risks of patients [2]. Therefore, it is necessary to minimize the radiation dose of CPCT scanning and achieve reliable diagnosis.

To obtain accurate CPCT reconstruction and hemodynamic parameters estimation at low dose, lots of approaches have proposed. They can be mainly divided into two categories. The first category is traditional model-based algorithms [3-4]. All these algorithms can obtain acceptable results, but most of them are patient-specific, and lack good generalization performance. The second category is the data-based algorithms, especially the deep learning based algorithms [5-6]. These networks perform much better than the conventional model-based methods. However, they all need high-dose CPCT images as label datasets to train the networks which is not easy to implement in clinics, and most of the methods usually assume the noise of CPCT images follows simple Gaussian distribution which might more complex in practice.

In this work, we first investigate the noise properties of CPCT images, and find that (1) the noise in each frame is very complicated far from a simple distribution, and it can be modeled by mixture of Gaussians (MOG). (2) The noise in different frames is obviously different from each other. Based on the observations and previous study [7], we present a self-supervised CPCT image reconstruction (SS-CPCT) method to yield high-quality CPCT images from low-dose measurements. Specifically, similar to the work [7], the SS-CPCT method is designed to estimate the high-quality CPCT images in a full Bayesian framework. First, considering the rich spatio-temporal correlation among CPCT images intrinsically, we use the adjacent frames to estimate the "clean" CPCT images as the labeled images. Then, we train the network in [7] to map between the low-dose CPCT image at each frame and its estimated "clean" image. The experimental results on clinical patients demonstrate that the presented SS-CPCT method can significantly improve CPCT image quality and produce more accurate hemodynamic maps compared with the competing network.

II. Methodology

A. CPCT image noise distribution

Fig. 1 shows the normal-dose CPCT images, low-dose CPCT images, the difference images between the normal and low -dose images, and the histogram map of the difference images. They are from the #5, #15, and #20 frames in the CPCT data. We can see that (1) the noise in each frame is complex, which obviously does not obey a simple distribution. For example, in Fig. 1 and Fig. 2, the noise at brain area is much heavier than the noise at the air area in a CT image. This refer to the fact that

S. Li, M. Zhu, D. Li, Q. Gao, and Z. Bian are with the School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China, and also with the Guangzhou Key Laboratory of Medical Radiation Imaging and Detection Technology, Southern Medical University, Guangzhou 510515, China.

*D. Zeng is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China. (E-mail: zengd@scut.edu.cn)

*J. Ma is the corresponding author with Guangdong Provincial Key Laboratory of Medical Image Processing, Department of Biomedical Engineering, Southern Medical University, Guangzhou, 510515, China. (Email: jhma@smu.edu.cn).

the variance of noise in a CT image it usually depended on the observation. (2) The noise in different frames is obviously different from each other. For example, the noise in the #5 frame is heavier than the other two frames. This may refer to the fact that the chemical composition of the observation is often temporally varying.

It has been demonstrated that mixture of Gaussians (MOG) models outperform single Gaussian distributions (or Laplace distributions) on tasks with complex noise. Thus, instead of modeling the noise in CT images by a simple distribution, we exploit the following MOG to model the noise E:

$$P_E\left(E|\pi,\Sigma\right)=\pi_k\sum_k N\left(E|0,\Sigma_k\right), \quad (1)$$

where π_k represents the mixing proportion with $\pi_k>0$ and $\Sigma_k \pi_k=1, k=1,....,K$ denotes the k th Gaussian component. $N\left(\cdot|0,\delta\right)$ represents a Gaussian distribution with zero mean and variance as δ.

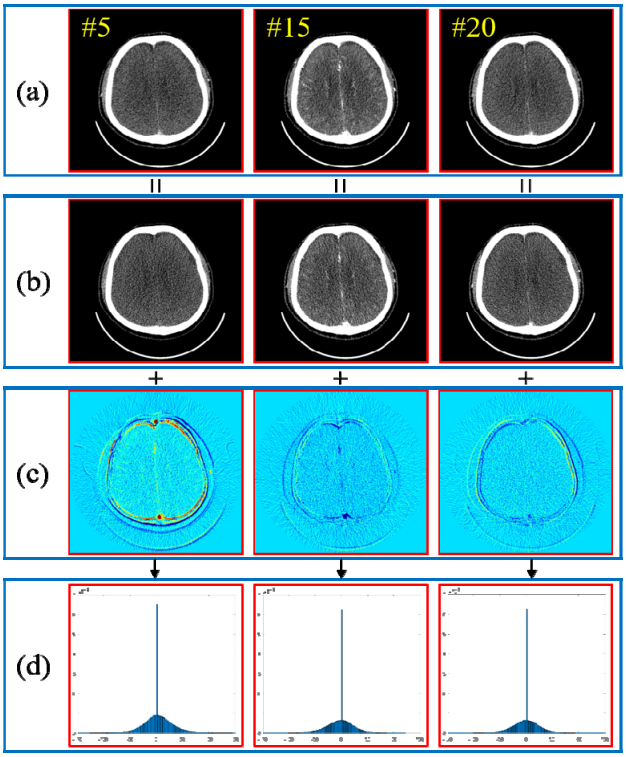


Fig. 1. The normal-dose CPCT images (a), low-dose CPCT images (b), the difference images between the normal and low-dose images (c), and the histogram map of the difference images (d).

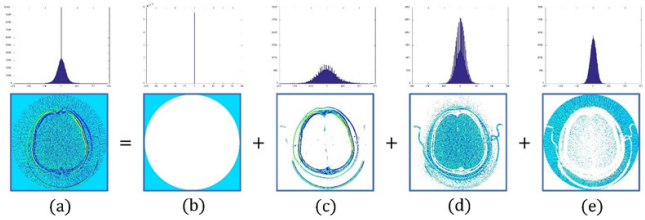


Fig. 2. (a)Illustration of the noise in a frame of CPCT image and its histogram. (b)-(e) Illustration of 4 different Gaussian components and their histogram, obtained by adopting an MoG model with 4 Gaussian components to the noise data in (a).

Based on the above analysis, we can easily find that it is not very rational to model this noise by a simple distribution, e.g., the most commonly used Gaussian distribution or Laplace distribution. However, many previous works, though has been demonstrated to be effective, still adopt a simple Gaussian-distribution assumption on this noise. Thus, there is still a large room to improve performance by better modeling this noise.

B. Self-supervised CPCT image reconstruction model

Due to the special scan protocol of CPCT imaging, it is hard to obtain the high-dose and low-dose CPCT images at the same time for supervised network training. Therefore, in this work, inspired by the previous study [7], we present a self-supervised CPCT image reconstruction method to estimate the high-quality CPCT images from low-dose measurements only. Specially, different from the traditional supervised deep learning methods which need a large number of labeled training data, the SS-CPCT method only enrolls the low-dose CPCT measurements to train network. In this work, considering the rich spatio-temporal correlation in CPCT frame images intrinsically, we use the CPCT images at adjacent frames to estimate the "clean" CPCT image as the target image. In CPCT imaging, adjacent frames are acquired in short time interval and have rich spatio-temporal correlation, so the simulated label data can be obtained by $y(t)=\left(x(t-1)+x(t+1)\right)/2$, t represents the frame index. It has lower noise and artifacts than target frame $x(t)$. Then, we train the network in [7] to map between the low-dose CPCT image at each frame and its estimated "clean" image.

III. EXPERIMENTS

A. Data

Under the written consent, real CT perfusion images scanned on human brains were used in our experiment, and patients with brain deficits were acquired by GE Discovery CT750 HD system, and before scanning, approximately 45 mL nonionic iodinated contrast was administered intravenously at 4 mL/second. Then the cine (continuous) enhanced normal-dose scan was performed with following protocol: 120kVp and 200mA. There are total 602 CPCT sequence images with size of 512 × 512 × 27. Some typical CPCT images are shown in Fig. 2. Based on these normal-dose CPCT images, we simulated the low-dose CPCT images using the simulation technique [8]. In order to ensure data consistency, we remove skull of all CPCT images, as a high-attenuation material, the CT value of skull is too higher than the CT value of brain tissue. In this work, we randomly select 572 sequence images for training and the remaining for testing, we also selected 3 images from the testing dataset for validation.

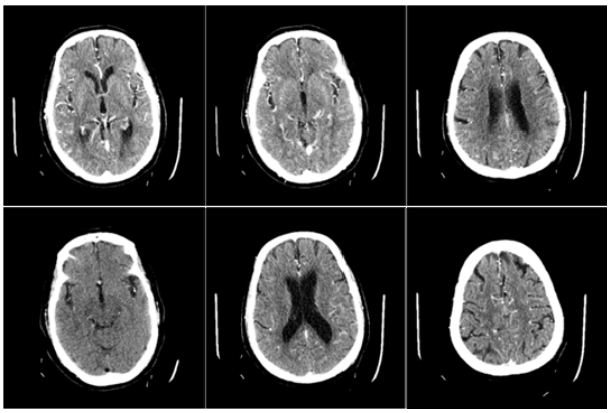


Fig. 3 Some examples from the normal-dose CPCT images

B. *Implementation details*

The network was trained for 100 epochs and in each epoch, we randomly crop $N = 128\times1000$ patches with size 128×128 for training. The adaptive moment estimation (Adam) algorithm is adopted to optimize the network parameters through minimizing the loss function. And the initial learning rate sets as 2×10^{-4} and linearly decreased in half every 10 epochs. The whole network was trained with an Nvidia Tesla P40 by using Pytorch [9].

C. *Evaluation merits*

We apply the root mean square error (RMSE) and structural similarity (SSIM) to quantitatively evaluate the results. The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(I_i - \bar{I}_i\right)}, \qquad (2)$$

where n is the total number of the pixels in image and I_i and $\bar{I}_i$ denote the pixel value at the i-th pixel in denoised image and reference image, respectively. The SSIM is calculated as:

$$SSIM = \frac{\left(2\mu_{I_i}\mu_{\bar{I}_i} + C_1\right)\left(2\sigma_{I_i}\sigma_{\bar{I}_i} + C_2\right)}{\left(\mu_{I_i}^{\ 2} + \mu_{\bar{I}_i}^{\ 2} + C_1\right)\left(\sigma_{I_i}^{\ 2} + \sigma_{\bar{I}_i}^{\ 2} + C_2\right)} \qquad (3)$$

where C_1, C_2 are small constants, μ_{I_i} and $\mu_{\bar{I}_i}$ represent the mean value of I_i and $\bar{I}_i$. σ_{I_i} and $\sigma_{\bar{I}_i}$ represent the variance value of I_i and $\bar{I}_i$, respectively.

IV. Result

To validate and evaluate the performance of the presented SS-CPCT method, Autoencoder was used for comparison. Fig.3 shows the normal-dose CPCT images and low-dose CPCT images obtain by Autoencoder and presented SS-CPCT algorithms at three representative frames (#5, #15 and # 25), respectively. In Fig. 4, it can be seen that the low-dose images in both frames suffered from high noise and several structures cannot be distinguished. Autoencoder and SS-CPCT can suppress most of the noise and artifacts. Furthermore, the presented SS-CPCT method preserved the structural features in the best way than Autoencoder, as indicated by red arrows in the Fig. 4. Moreover, to quantitatively evaluate the capability of the presented method, we assume the normal-dose images as reference and calculate the RMSE and SSIM at each frame as shown in Fig. 5. Overall, the RMSEs of low-dose images are 8.0045±0.0003, which reduce to 5.3872±0.0319 for Autoencoder images, and reduce to 3.9098±0.0539 for the presented SS-CPCT, and the SSIMs of low-dose images are 0.8119±0.0116, for Autoencoder are 0.9060±0.0152, for the presented method are 0.9401±0.0023.

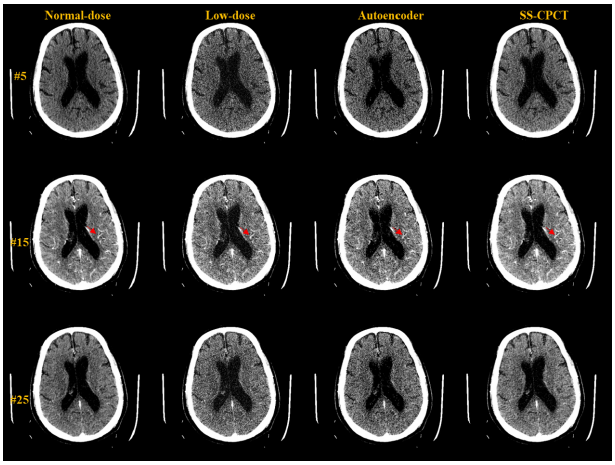
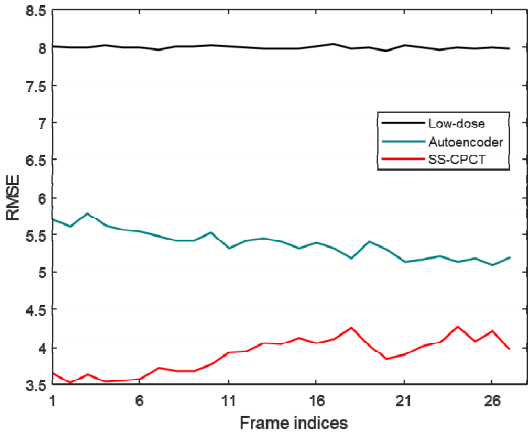
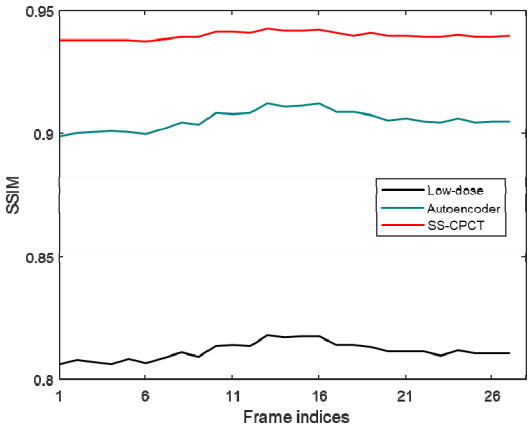


Fig. 4 The CPCT images denoising by the different algorithms All the images are displayed in the same window: [10 80]HU.



(a)



(b)

Fig. 5 RMASE (a) and SSIM (b) from different algorithms for each frame.

To further demonstrate the performance of the presented SS-CPCT method, we also evaluate the corresponding CBF maps from the different methods, and the results are displayed in Fig. 6. Compared with the normal dose CBF map, the presented method significantly improves the CBF map quality by reducing noise and preserve structural details and the result shows strong agreement with the normal-dose one in visual inspection than Autoencoder result. Meanwhile, the SSIM and RMSE metrics are used to evaluate the performance quantitatively are listed in Table I. And the presented SS-CPCT method obtain highest SSIM and lowest RMSE values, indicating the presented method produce better performance than the Autoencoder in terms of on noise reduction and local structures preservation.

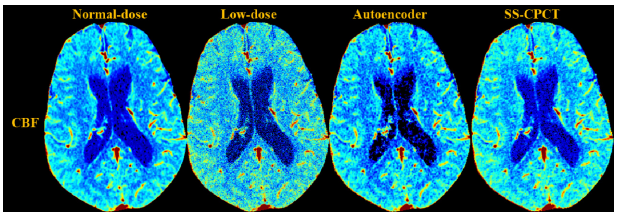


Fig.6 The CBF maps from the CPCT images reconstructed by the different algorithms. The unit is $ml/100g/min$

TABLE I
QUANTITATIVE RESULTS (MEAN±SDS) ASSOCIATED WITH DIFFERENT COMPETING ALGORITHMS ON THE CBF MAPS

	SSIM	RMSE
Low-dose	0.6515	7.8514
Autoencoder	0.8808	6.1258
SS-CPCT	0.9079	5.6371

V. CONCLUSION

Based on the CPCT noise distribution analysis, we find that it is not very rational to model this noise by a simple distribution. In this work, we model the noise as a mixture of Gaussians (MoG) distribution, and different with the traditional supervised methods which need a large number of labeled training data, we consider the rich spatio-temporal correlation in CPCT frame images intrinsically, and then use the adjacent frames to estimate the "clean" CPCT image as labeled images. Finally, we train the network in [7] to map between the low-dose CPCT image at each frame and its estimated "clean" image. The performance of SS-CPCT method is verified by comprehensive comparative studies against Autoencoder through the visual and quantitative evaluations. In the future work, extending the presented SS-CPCT method for evaluating other tasks or datasets will be conducted.

ACKNOWLEDGMENT

This work was supported in part by the NSFC under Grant U1708261, Grant 81701690, Grant 61701217, and Grant 61571214, in part by the Science and Technology Program of Guangzhou, China, under Grant 201705030009, in part by the Science and Technology Program of Guangdong, China, under Grant 2015B020233008 and in part by the China Postdoctoral Science Foundation funded project under Grant 2019M652910.

REFERENCES

[1] Benjamin, Emelia J., Salim S. Virani, Clifton W. Callaway, Alanna M. Chamberlain, Alexander R. Chang, Susan Cheng, Stephanie E. Chiuve et al. "Heart disease and stroke statistics-2018 update: a report from the American Heart Association,:" Circulation, vol. 137, no. 12, pp. e67, Mar. 2018.

[2] Wintermark, M., and M. H. Lev. "FDA investigates the safety of brain perfusion CT," American Journal of Neuroradiology January, vol. 31, no. 1, pp. 2-3, Jan, 2010.

[3] J. Ma, H. Zhang, Y. Gao, J. Huang, Z. Liang, Q. Feng, and W. Chen, "Iterative image reconstruction for cerebral perfusion CT using a precontrast scan induced edge-preserving prior," Phys. Med. Biol., vol. 57, pp. 7519-7542, Nov 21 2012.

[4] D. Zeng, X. Zhang, Z. Bian, J. Huang, H. Zhang, L. Lu, W. Lyu, J. Zhang, Q. Feng, W. Chen, and J. Ma, "Cerebral perfusion computed tomography deconvolution via structure tensor total variation regularization." Medical Physics, vol. 43, pp. 2091-2107, 2016.

[5] Y. Xiao, P. Liu, Y. Liang, S. Stolte, P. Sanelli, A. Gupta, J. Ivanidze, and R. Fang, "STIR-Net: Deep Spatial-Temporal Image Restoration Net for Radiation Reduction in CT Perfusion," Frontiers in Neurology, vol. 10, p. 647, 2019.

[6] V. S. Kadimesetty, S. Gutta, S. Ganapathy, and P. K. Yalavarthy, "Convolutional Neural Network-Based Robust Denoising of Low-Dose Computed Tomography Perfusion Maps," IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 3, no. 2, pp. 137–152, Jul 2018.

[7] Yue, Zongsheng, et al. "Variational denoising network: Toward blind noise modeling and removal," Advances in Neural Information Processing Systems. 2019.

[8] J. Ma, Z. Liang, Y. Fan, Y. Liu, J. Huang, W. Chen, and H. Lu, "Variance analysis of x-ray CT sonograms in the presence of electronic noise background," Medical Physics, vol. 39, pp. 4051-4065, 2012.

[9] Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen et al. "PyTorch: An imperative style, high-performance deep learning library," In Advances in Neural Information Processing Systems, pp. 8024-8035, 2019.

Deep Learning Acceleration of OS-SIRT in X-ray Cone-Beam Computed Tomography

Stefan Haninger
ImFusion GmbH
Munich, Germany
haninger@in.tum.de

Matthias Wieczorek
ImFusion GmbH
Munich, Germany
wieczorek@imfusion.com

Wolfgang Wein
ImFusion GmbH
Munich, Germany
wein@imfusion.com

Tobias Lasser
Technical University of Munich
Munich, Germany
lasser@in.tum.de

***Abstract*—This paper presents an acceleration technique for iterative algorithms to jump from a low convergence state to a higher one by the application of deep learning. This leapfrogging strategy [1] is applied to an OS-SIRT algorithm on real Cone-Beam Computed Tomography animal scans. Furthermore, we introduce a post-processing step into the pipeline in order to reduce noise in high convergence states. We show that our U-Net architecture manages to learn a highly accurate prediction, which also appears to generalize towards heterogeneous datasets.**

I. Introduction

X-ray Computed Tomography (CT) is a key technology for non-invasive visualization of the human body in clinical routine. In clinical scanners, analytical methods like Filtered Backprojection are predominantly used for the reconstruction. Their main benefit is that the reconstruction comes at a low computational cost. However, iterative methods have been popular in the past due to their higher flexibility, especially when it comes to regularization. Regularization allows for improved handling of noise and artifacts which might be contained in the data. However, this comes at comparably high computational costs. In recent years, researchers have integrated deep learning into the tomographic reconstruction pipeline. Arridge et al. [2] recently provided an overview of applying machine learning techniques (or "data-driven models") to Computed Tomography. Many recent works focus on two application areas, first low dose CT (e.g. [3], [4]), i.e. the task of denoising the reconstruction, and second sparse data CT (e.g. [5], [6]), i.e. the task of removing artifacts from the reconstruction. These problem domains need different consideration as low dose noise can be treated locally, whereas sparse data streaking artifacts are structures which extend across the volume. There appear to be two main ways how deep learning is incorporated into the tomographic reconstruction pipeline: The first approach uses a learned model as a post-processing step (e.g. [3], [5]). The second one is to use data-driven models for iterative algorithms as regularization or optimization scheme to push the reconstruction to a more desirable outcome (e.g. [4], [6]). Besides using deep learning for improving image quality, it can also be employed to reduce run-time of high performant but slow iterative methods. This idea was first implemented by Cheng et al. [1]. In their work they used a "leapfrogging" strategy in order to jump from a low convergence state of a BSREM algorithm to a higher one with a simple three layer Convolutional Neural Network (CNN). In this work we apply a similar leapfrogging strategy (see Fig. 1) to an OS-SIRT approach on real cone-beam data. We use advanced deep architectures for the leapfrogging approach, like the U-Net [7] architecture (see Fig. 2). Furthermore, we apply an anisotropic filter to the high convergence state to learn a leapfrogging and post-processing at the same time. We evaluate this method on similarity and convergence metrics. Additionally, we provide visual results of the network's performance on various datasets.

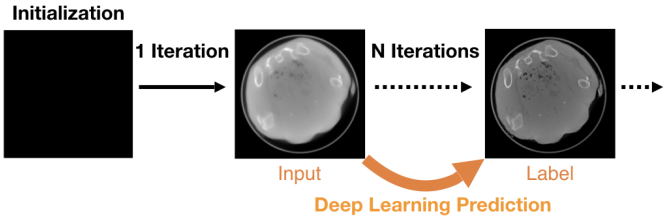


Fig. 1: Schematic of leapfrogging an iterative reconstruction algorithm by deep learning prediction.

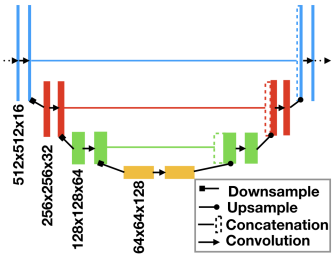


Fig. 2: Schematic of the employed U-Net architecture, using four levels and one convolution per level.

II. Methodology

In order to leapfrog an iterative algorithm using deep learning, we first generate the training data and then train a specific model. The model will then be integrated into the reconstruction pipeline. As the saved time is proportional to the number of skipped iterations, we used the result of a low number of iterations as leapfrogging input, while for the training label a rather high number is used (ideally close to

convergence). Furthermore, we added an optional application of post-processing steps, such as filtering, at the end of the data generation pipeline. This results in directly embedding the post-processing within the trained model. Thus the data generation consists of generation of the input from a low iteration state, the target from a high iteration state, and an optional application of post-processing to the labels. During the training phase we train a CNN from the input images to the labelled images. This is an image-to-image translation/pixel-wise regression task, and in general any corresponding architecture fits within this framework. Finally, during the reconstruction phase we apply as many iterations of the iterative reconstruction as we used for the input data, and then apply the trained model to leapfrog. After the deep learning prediction the iterative algorithm can be reapplied, correcting potential shortcomings of the model.

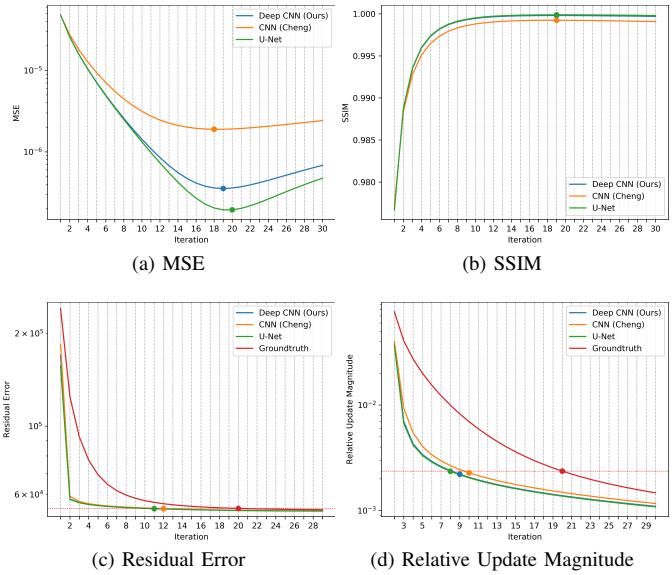
III. Experiments

The dataset for the experiments contains 12 cone-beam CT scans from 6 different ex-vivo animals from the same species (2 scans per animal), all of which were scanned with the same scanner setup. Additionally, for evaluation we included one cone-beam scan of another animal species with a different scanning setup for evaluation on a heterogeneous dataset. The training data was taken from 8 scans, the validation data from 2 scans and the testing data from 2 scans of the first animal, plus the one from the second animal. From each scan we take slices for one data point, so that the deep learning model is a 2D image to 2D image mapping. The same method can be applied to 3D, but we wanted to test for deeper architectures with only a few training datasets, and we had to crop out cone artifacts in the axial dimension. All the datasets were reconstructed to a resolution of $512 \times 512 \times 512$ and split in axial dimension into 512 slices. The reconstructed volumes of the first animal have a spacing of 0.443359 mm and the volume of the second one a spacing of 0.3 mm. For the reconstruction we used an OS-SIRT implementation (c.f. [8]) with a subset size of 30 projections. Furthermore, a positivity constraint was applied. The number of OS-SIRT iterations was fixed for the input to 1 and for the label to 20. For the input we selected the first iteration since it shows many image details but with a strong blur. On the other hand, for the labels we selected the 20th iteration as a trade-off between image details and noise. The first and last slices which had cone artifacts were removed from the training data, such that from each volume 356 of the 512 slices were taken. Furthermore, we generated additional labels by clipping all label data to positive values (as is done in each OS-SIRT iteration), and by including a Perona-Malik anisotropic diffusion [9] filter into our data generation phase, in order to filter out all noise structures. For the network architecture we experimented with different deep CNNs and U-Nets, where we selected one specific architecture for each network type after some experimentation. All networks were wrapped with one first layer with 32 channels, 5x5 kernels and one last layer with one channel and 7x7 kernels. Furthermore, all layers in the middle have a fixed channel size of 16 and a fixed kernel size of 3x3. Accordingly, the first architecture we selected, a deep CNN, has a total number of 12 layers. During the experimentation we found that for deep CNNs, an architecture with 10-20 layers had the best performance. The second selected architecture, the U-net, consists of 6 levels and 2 convolutional layer per level with average pooling for downsampling and a nearest neighbor interpolation followed by a 2x2 convolution for upsampling. In each level the resolution of the features was down/upsampled by a factor of 2. Thereby, the number of channels was increased/decreased by the same factor. We trained with the Adam optimizer in a standard configuration (learning rate 10^{-4}, β_1=0.9, β_2=0.999) and batch size 1. In our experiments only this particular batch size let the model train properly. We further used a learning rate decay $lr(epoch) = \frac{1}{1+lr(epoch-1)\cdot dr}$ with decay rate (dr) of 0.05 and trained for 200 epochs.

IV. Results

A. Quantitative Analysis

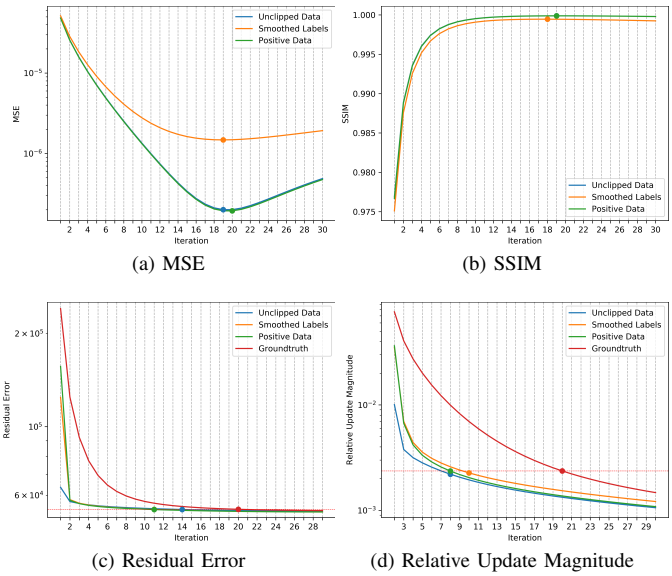
For quantitative analysis, we selected different similarity and convergence metrics and took time measurements of different models and of single OS-SIRT iterations. With similarity metrics we examine which convergence state is most similar to the deep learning prediction. Therefore, we measure the SSIM and MSE between prediction and the iterations ranging from 1 to 30. Since we trained on a cropped volume to leave out cone beam artifacts, we calculate metrics also on cropped versions of the test data. Furthermore, we evaluate the actual convergence state of the prediction using convergence metrics we compute using the OS-SIRT as ground truth and each deep learning model with further 29 OS-SIRT iterations applied to it. The first convergence metric we use the Residual Error which is the objective function of the CT optimization, i.e. the squared L2 norm of the residual. However, in this case the cropping is not straightforward, since any change in the volume, such as cropping, results in a high change to the forward projection. Hence we apply a cropping in projection space. As a second convergence metric, we calculate the relative change in the cropped volume between consecutive iterations as "Relative Update Magnitude" (RUM$(i) = \frac{\left\| x^i - x^{i-1} \right\|}{\left\| x^i \right\|}$). The MSE in Fig. 3a shows that the prediction of the U-Net has its optimum at iteration 20 (marked by a point). Furthermore, the MSE at this point is smaller than the one of the deep CNN and Cheng's CNN [10]. This is reflected by the SSIM in Fig. 3b, which, however, shows a less significant difference between the models. The Residual Error (Fig. 3c) only shows a small difference after the prediction of each model (iteration 1 in the graph) to the ground truth, but then drops quickly after one further OS-SIRT iteration. The model graphs then form a plateau and do not differ much on higher iterations. However, the point where the models first reach the convergence state of 20 OS-SIRT iterations, is at 11 iterations for the U-Net, one iteration earlier than the model of Cheng (marked by the red dashed line and the point in each graph). On the other hand, the RUM shows a more



(a) MSE (b) SSIM

(c) Residual Error (d) Relative Update Magnitude

Fig. 3: Similarity (a),(b) and convergence (c),(d) metrics with positive labels for different architectures.



(a) MSE (b) SSIM

(c) Residual Error (d) Relative Update Magnitude

Fig. 4: Similarity (a),(b) and convergence (c),(d) metrics with U-Net for different labels.

significant difference between the models. The crossing point of the U-Net is at 8 iterations, while the one of Cheng's model is at 10 iterations. The plot reveals an advantage of our deep models over the shallow CNN of Cheng. Fig. 4 shows the similarity metrics trained only on the U-Net architecture with different data processing (labels with no clipping, clipping to positive values, filtering). The MSE (Fig. 4a) shows that there is only a small difference between positive and unclipped data, while the positive data models have a slight advantage.

TABLE I: Mean and standard deviation of the runtime of different model predictions and a single SIRT iteration of 100 runs in milliseconds.

	Mean	Standard Deviation
1 OS-SIRT iteration	7009 ms	81 ms
Cheng's CNN	9186 ms	148 ms
Deep CNN	12774 ms	148 ms
U-Net	16200 ms	228 ms

The smoothed labels, on the other hand, show a much higher error, which is expected since it was trained on a filtered version of the labels. That the other models are better in the MSE indicates that the models reproduce at least some of the noise structure correctly. A similar picture arises in the SSIM (see Fig. 4b), where the prediction of models trained on unclipped and positive labels completely overlap and the smoothed labels are slightly worse. The Residual Error in Fig. 4c shows that the unclipped data prediction starts at a much smaller value than the ground truth and the other models. This behavior is expected, since the labels of the other models were processed. However, at the second iteration all model curves are at the same value, so the plot suggests that the OS-SIRT compensates for the clipping and smoothing in only one iteration. Furthermore, the point where the smoothed and positive model first cross the value of 20 OS-SIRT iterations lie on the same point. So the iterative algorithm can even compensate for the smoothing and manages to catch up with the positive clipped prediction. However, the unclipped model approach crosses this value three iterations later. In Fig. 4d, we can observe that the unclipped model starts at a much smaller value, but the positive model converges to it. On the other hand, the smoothed labels model starts at a similar value as the positive one, but decreases slower. Finally, the run time has been evaluated for a single OS-SIRT iteration versus the application of each model with the ImFusion framework on all 512 slices. We ran the model on a GeForce GTX 1080 and the results are listed in table I. While the shallow model of Cheng et al. [1] only takes a bit longer than one SIRT iteration, the U-Net takes a bit more time than two SIRT steps.

B. *Qualitative Analysis*

We further provide a qualitative analysis with slices and line plots of the U-Net predictions on the test data. Here, we compare our U-Net against the architecture of Cheng et al. [1] on line plots of the second animal data (that was not trained on) to evaluate how good the different models generalize. Furthermore, we show the visual performance of the U-Net on the test data trained on the positive data. The prediction of the models, trained on the positive data (Fig. 5), shows a significant difference between the U-Net and Cheng's model. While both models predict the big spikes well, the shallow CNN shifts the prediction in flat areas a bit upwards compared to its label. The U-Net manages to predict all areas well and almost completely overlaps the label. The same effect can be observed when the models are trained with the smoothed labels (Fig. 6). Here, the model of Cheng further inserts some

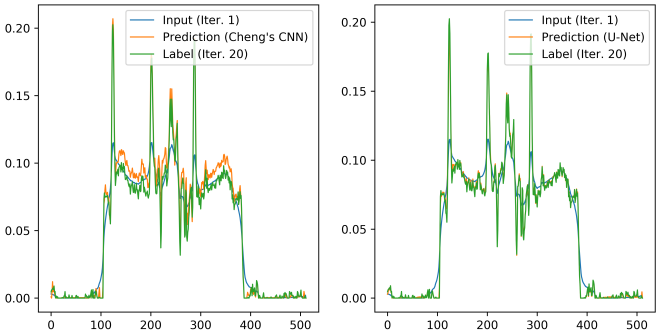


Fig. 5: Comparison of line plots from the model of Cheng et al. [1] and our U-Net trained on positive labels on the second animal dataset.

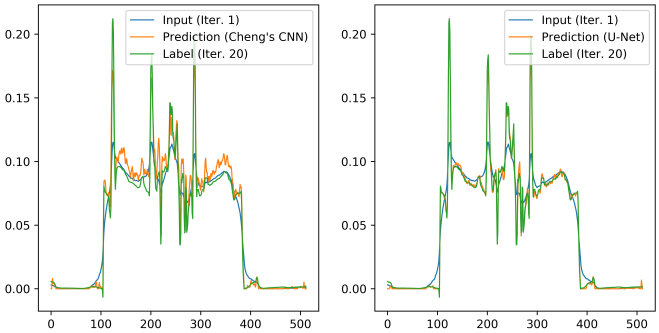
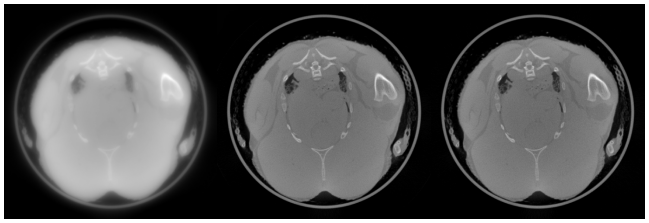


Fig. 6: Comparison of line plots from the model of Cheng et al. [1] and our U-Net trained on smoothed labels on the second animal dataset.

noisy structures which are not present in the label. The same happens to the prediction of the U-Net, but less severe. In summary, the U-Net predicts all image areas equally well and clearly outperforms shallower networks on all labels. Finally, the visual appearance of the U-Net prediction on the first animal test data is shown in Fig. 7. While the input shows a very blurry appearance, the labels have sharp structures, which the prediction completely recovers, such that it cannot be visually distinguished from the label. Since the training was only performed on reconstructions of one type of animal, we further evaluated on different data. Therefore, the prediction



(a) Input (b) Prediction (c) Label

Fig. 7: Slices of the first animal test data with prediction of U-Net leapfrogging.

(a) Input (b) Prediction (c) Label

Fig. 8: Slices of the second animal test data with prediction of U-Net leapfrogging.

of the same network on the second animal dataset is shown in Fig. 8. The results of this plot show that the prediction is also very close to the label on heterogeneous data except for a subtile blur in bone structures. That the prediction has such a high quality on different datasets shows that the network is learning a sharpening filter which accurately recovers the application of OS-SIRT.

V. Conclusion

In this paper we proposed a deep learning based leapfrogging approach in order to accelerate an OS-SIRT algorithm. Our results indicate that U-Nets provide the best prediction in terms of leapfrogging convergence state, image quality, and generalization for different datasets. A smoothed version of the labels can also be trained to incorporate a post processing step into the leapfrogging prediction to filter out noise.

References

[1] L. Cheng, S. Ahn, S. Ross, H. Qian, and B. De Man, "Accelerated iterative image reconstruction using a deep learning based leapfrogging strategy," *International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017.

[2] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, pp. 1–174, 2019.

[3] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

[4] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Problems*, vol. 33, no. 12, p. 124007, 2017.

[5] Y. Han and J. C. Ye, "Framing u-net via deep convolutional framelets: Application to sparse-view CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418–1429, Jun. 2018.

[6] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "Learn: Learned experts' assessment-based reconstruction network for sparse-data CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1333–1347, Jun. 2018.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[8] A. C. Kak and M. Slaney, "Principles of computerized tomographic imaging," *IEEE, New York*, pp. 284–292, 1988.

[9] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[10] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT denoising with convolutional neural network," in *Proc. IEEE 14th Int. Symp. Biomedical Imaging (ISBI 2017)*, Apr. 2017, pp. 143–146.

4D Segmentation-Based Anatomy-Constrained Motion-Compensated Reconstruction of On-Board 4D CBCT Scans

Markus Susenburger, Pascal Paysan, Adam Strzelecki, Igor Peterlik, Dieter Seghers, and Marc Kachelrieß

***Abstract*—On-board 4D CBCT scans benefit from motion-compensated (MoCo) reconstruction, where motion vector fields, that are estimated from the acquired data, are applied during reconstruction. In some cases, however, MoCo suffers from non-physiological motion artifacts such as moving bone, for example. In this work, we enhance our artifact-specific cyclic motion compensation (acMoCo) approach by adding spatially and temporally resolved prior knowledge. This is done by using a four dimensional segmentation. The algorithm obtained is the anatomy-constrained acMoCo (acacMoCo) algorithm. The advantage of the anatomical prior knowledge is that it allows us to incorporate additional constraints into the registration process such that motion estimation can now be handled differently for different anatomical regions. The effect on the image quality of the spine, the ribs and the lung border is demonstrated and found to improve the overall 4D image quality.**

***Index Terms*—Radiotherapy, 4D CBCT, motion-compensation, image reconstruction, 4D segmentation**

I. Introduction

RESPIRATION-CORRELATED cone-beam CT (CBCT) data can be used to increase the spatial accuracy of stereotactic body radiation therapy (SBRT) and therefore decrease the undesired dose deposit in healthy tissue [1], [2]. Current commercially available 4D reconstructions approaches are based on a respiration-gated (binned) reconstruction [3]. However, current 4D acquisition protocols tend to be of rather low image quality and advanced approaches that improve the overall impression may lead to motion underestimation or non-physiological motion patterns [4]. Moreover, gating does not make use of the full x-ray dose because each image only contains the dose of the respective gate, say, 10 or 20% of the total dose.

The aim of MoCo reconstruction is to use the full dose for each reconstructed respiratory phase. This allows to obtain high quality images even for short scan times. One of the MoCo algorithms which showed very good results in terms of dealing with sparse data from one minute scans (reasonable level of dose and short acquisition time) is the artifact-model based cyclic motion-compensation (acMoCo). It features a patient-specific, 3D artifact model as well as a cyclic constraint to ensure the periodic behavior of respiratory motion [5], [6]. The algorithm was further extended to deal with sliding organ motion, such as between the lung cavity and the ribs, with motion vector field upsampling and with irregular breathing patterns [7], [8], [9]. However, it turned out that even then the inherent deformable image registration algorithm has problems to deal with certain, moving anatomical structures. One prominent example is the spine which tends to move in MoCo images from very sparse data, although the true solution for patients positioned supine is a zero-filled motion vector field. Another problem is the required regularization of the motion vector fields that corresponds to a local smoothing which does, however, not account for anatomical borders.

To tackle those problems, we developed the anatomically-constrained acMoCo (acacMoCo) algorithm, a method based on temporally-resolved anatomical knowledge. This prior knowledge allows us to constrain or regularize the motion estimation in a physiologically more realistic way. We here demonstrate the potential of acacMoCo on the spine, the ribs and the pleural cavity. These represent regions of stationarity, rigid motion and sliding organ motion, respectively. We compare the results with acMoCo and with the clinical references respiratory-gated FDK and modified McKinnon-Bates (mMKB) algorithm [10]. 10 patient cases are analyzed.

M. Susenburger and Prof. Dr. M. Kachelrieß are with the German Cancer Research Center, Heidelberg, Germany, and with the University of Heidelberg, Heidelberg, Germany. Dr. P. Paysan, Dr. I. Peterlik, Dr. A. Strzelecki and Dr. D. Seghers are with Varian Medical Systems, Baden-Dättwil, Switzerland. Corresponding author: Markus Susenburger (markus.susenburger@dkfz.de)

II. Methods

A. Data Acquisition

The 4D CBCT scans are acquired on a Varian TrueBeam in full arc and shifted detector mode within one minute scan time, two scans are acquired within two minutes with a lower frame rate. The respiration signals are either measured during treatment (extrinsic), or intrinsically estimated in a reconstruction software (iTools Reconstruction, Varian Medical Systems) or an available respiration and cardiac gating method [11]. The parameters are collected in table I.

B. 4D Reconstruction

Given a CBCT scan acquired on an on-board imaging system and a respiration signal, either measured or intrinsically estimated, we sort the projection into 20 overlapping respiration bins (width = 10%, sampling distance = 5%). Each phase is separately reconstructed with the Feldkamp-David-Kress (FDK) algorithm [12]. Since those gated data are angularly sparse the phase-gated images contain severe reconstruction artifacts. In MoCo reconstruction, all the acquired projections will be used for each phase. Projections from other phases are deformably warped into the space of the destination phase after

backprojection. This process is repeated for every respiratory phase, i.e. 20 times in our case.

C. Deformable Image Registration

Given a source volume $\boldsymbol{f}$ which should be deformably registered to the destination volume $\boldsymbol{g}$, we are looking for a deformation vector field $\boldsymbol{d}(\boldsymbol{r})$, such that

$$T_{\boldsymbol{d}}\boldsymbol{f}(\boldsymbol{r}) = \boldsymbol{f}(\boldsymbol{d}(\boldsymbol{r})) = \boldsymbol{g}(\boldsymbol{r}). \quad (1)$$

The Demon's algorithm is a well-known method to find a deformable motion vector field satisfying the transformation above [13]. In our implementation, we follow the symmetric forces extension together with the diffeomorphic framework [14]. The non-parametric Demon's framework aims at minimizing the cost function between image similarity and smoothness of the motion vector fields. This cost function can be solved as a two step optimization: First, find an update for the motion vector field, second, regularize the (new) motion vector field. As regularizers, Gaussian smoothing is chosen in the unmodified algorithm for a pre-regularization on the update motion vector field and a second Gaussian smoothing for a post-regularization on the newly composed motion vector field. The Demon's algorithm is implemented to work on a multi-resolution approach, starting on a coarse grid and subsequent registrations on finer grids until the final voxel size of $1 \times 1 \times 1$ mm^3 on a $512 \times 512 \times 210$ grid is reached.

D. Cyclic Motion Compensation

The set of motion vector fields is obtained by a deformable registration of the phase-gated images in a cyclic fashion. The deformation from phase n to the adjacent phase $n+1$ is denoted by T_n^{n+1}, hence the deformation from an arbitrary phase n to any other phase m in the cycle can be understood of a composition of adjacent motion vector fields:

$$T_n^m = T_{m-1}^m \circ \ldots \circ T_{n+1}^{n+2} \circ T_n^{n+1}. \quad (2)$$

Since the respiratory motion is cyclic T_n^{n+N} should equal the identity tranform. To make sure that this is the case a cyclic constraint is applied during the motion vector field estimation process, penalizing deviations from the identity transform for a full cyclic vector field composition (see details in reference [5]).

E. Artifact Model-Based Cyclic Motion Compensation

The motion vector fields estimated by the cyclic motion compensation comprise the desired patient motion vectors and an undesired contribution from the artifact motion. The artifact model is to estimate the latter. A simple threshold-based patient-specific coarse segmentation is performed to obtain a motion-free 3D image. This volume is forward projected into the same geometry as the original measurement and a gated reconstruction of these virtual rawdata is performed. The images now only contain the artifact motion and no patient motion. After two separate cyclic motion estimations, we subtract the artifact motion vector field from the total motion vector field. This gives a motion vector field describing just the patient motion.

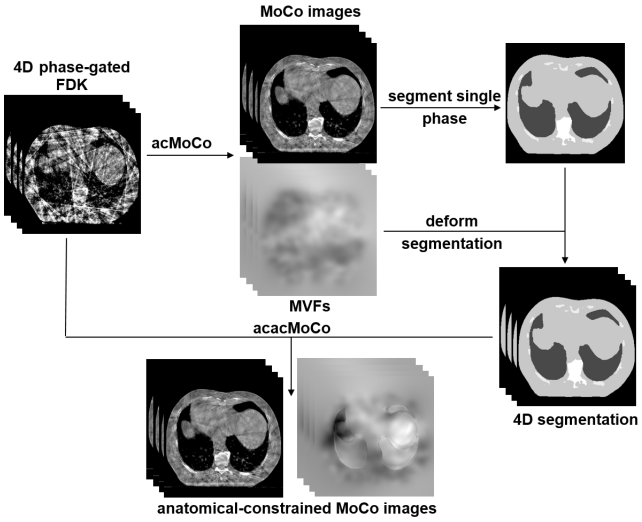


Fig. 1. Workflow scheme of the proposed method. An initial acMoCo is performed to gain a high quality image of a certain respiration phase (phase 0 in the case of this study). After the segmentation of the end of inhale phase, we apply the motion vector fields estimated in the initial acMoCo to transform the segmentation into the remaining phases. These segments are used as additional constraints in the acacMoCo approach.

F. 4D Segmentation

Given an initial acMoCo reconstruction, i.e. 20 nearly artifact-free volumes and 20 cyclically-constrained motion vector fields, an advanced organ segmentation is performed on one of the phase (we have chosen phase 0 which corresponds to the end of inhale). The segmentation is performed with an autosegmentation tool (MIRS, Varian Medical Systems). It provides us with a mask for each segmented anatomical region. The anatomical regions are body, bones, spine and lungs and an estimate for the body region close to the couch (see figure 2). We then use the motion vector fields to deform these masks into the other respiratory phases (see figure 3).

G. Sliding Organ Border Motion

To prevent the sliding organ border motion, a guided bilateral filter is used [7]. Given the 4D prior segmentation of the lung, we assume sliding motion at lung chest wall interface, so that we limit the motion transfer introduced otherwise by the conventional Gaussian regularization, which is isotropic. The filter works as follows: The lung mask is morphologically processed to decompose each motion vector into a component tangential to the lung surface and into a component perpendicular thereto. The tangential direction is the one where the sliding may occur. The perpendicular component is filtered with a Gaussian filter G of standard deviation σ, while the tangential component of the motion vector field is convolved with a bilateral filter consisting of two Gaussian with standard deviations σ and ω and the latter using the the prior image P as its argument:

$$\boldsymbol{d}(\boldsymbol{r})_{\parallel} \propto \sum_{\boldsymbol{r}'} G_\sigma(\boldsymbol{r}-\boldsymbol{r}') \cdot G_\omega(P(\boldsymbol{r}) - P(\boldsymbol{r}')) \cdot \boldsymbol{d}(\boldsymbol{r}')_{\parallel}$$

$$\boldsymbol{d}(\boldsymbol{r})_{\perp} \propto \sum_{\boldsymbol{r}'} G_\sigma(\boldsymbol{r}-\boldsymbol{r}') \cdot \boldsymbol{d}(\boldsymbol{r}')_{\perp}$$

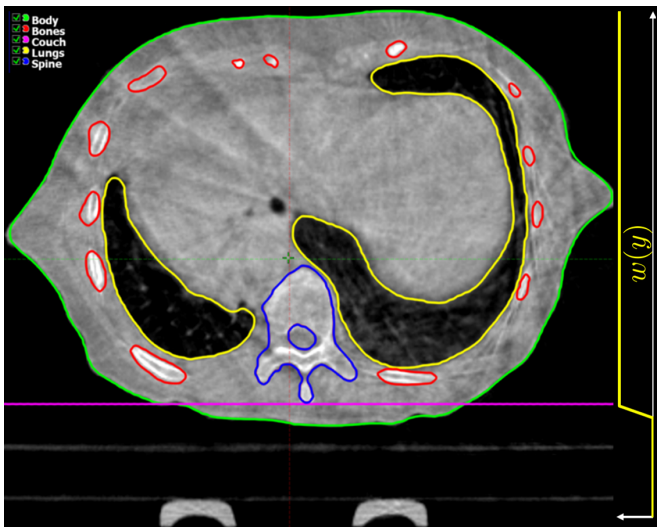


Fig. 2. A segmentation performed on the end of inhale phase of an initial acMoCo image, divided into the structures body, bones, couch, lungs and spine. The body is used to regularize the motion vectors of the registration outside of the body, with ribs and lung is dealt with the sliding lung motion filter, artifacts in the back are controlled by the couch segmentation and the spine is kept stationary. On the right, the y dependency of the weight $w(y)$ is shown (see section II-H).

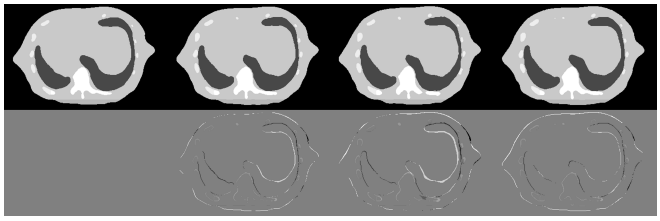


Fig. 3. Four of 20 segmented respiratory phases are shown as an example. The anatomical structures patient, lung, ribs, spine and patient's back are displayed in different shades of gray. Differences between the phases are small for the shown slice in the transversal view. The second row shows the difference image to the end of inhale phase (reference phase for the segmentation).

with the constant of proportionality chosen such that the filters are properly normalized.

H. Anatomically-Constrained Motion Estimation

The anatomical constraints are implemented with different boundary conditions on the motion vector field $\boldsymbol{d}$:

- Motion vectors belonging to the spine region are set to zero: $\boldsymbol{d}(\boldsymbol{r}) = 0$ if $\boldsymbol{r} \in \boldsymbol{\Omega}_{\text{spine}}$.
- Motion vectors at the sliding lung interface are treated as described in the previous section.
- The patient outline is used to minimize the volume of interest for the registration by setting the motion vectors to zero and omitting the registration: $\boldsymbol{d}(\boldsymbol{r}) = 0$ if $\boldsymbol{r} \notin \boldsymbol{\Omega}_{\text{body}}$.
- The motion of the patient's back is suppressed by weighting the motion vector field linear to its distance to the couch $w(y)\boldsymbol{d}, w(y) \in [0, 1]$. The reasoning is motion suppression close to the couch ($w = 0$) and a smooth transition to a full motion weight in the interior ($w > 0, w = 1$ at the violet line).

TABLE I
SCANS CONSIDERED IN THIS STUDY WITH INFORMATION ON SCAN TIME, THE CURRENT OF THE USED X-RAY SOURCE AND THE GATING METHOD.

Patient	RPM	Gating	Scan Time	Current	FPS
B00	19	intrinsic	60 s	80 mA	15
B01	17	intrinsic	60 s	20 mA	15
B02	21	intrinsic	60 s	20 mA	15
B03	21	intrinsic	60 s	80 mA	15
BT2	27	extrinsic	60 s	20 mA	15
M00	13	extrinsic	120 s	20 mA	11
M01	18	extrinsic	60 s	112 mA	15
M03	21	extrinsic	60 s	112 mA	15
H02	27	unknown	60 s	80 mA	15
V01	18	extrinsic	120 s	40 mA	7

Each boundary condition is evaluated in each iteration of the Demons motion estimation in the four dimensional domain of the source image $\boldsymbol{f}$, as it undergoes the deformation encoded in the motion vector field towards the destination image $\boldsymbol{g}$. The highest priority of motion constraints in overlapping or neighboring structures have those who enforce zero motion (spine and body mask). A smooth transition between those regions is ensured over the smoothing operation of involved filters (bilateral or Gaussian, depending on the lung segmentation prior).

Additionally, the cyclic constraint and the patient-specific artifact model are used in the acacMoCo.

III. RESULTS

We tested the 4D segmentation and the acacMoCo algorithm using ten patient data sets that were provided by Varian Medical Systems. The acquisition parameters for every data set can be found in table I, the result of scan (B00) is shown in figure 4. The acacMoCo improves the result of the acMoCo in the spine, as we introduced the stationary constraint. This effect is best seen in the 4D respiration sequence, as the motion occurs between the phases. It is also indicated by the blurring of the spine in the acMoCo. The sliding lung motion filter prevent cross organ border motion transfer, this results in a recovery of the rib shape and slightly sharper diaphragm. The linear weighted motion suppression in the back mitigates the artifacts in the patient's back. Tissue structures can be seen in the acacMoCo which are impaired by artifacts in the acMoCo. All 4D reconstructions can reduce the sparseness artifacts while preserving the image quality.

IV. DISCUSSION AND CONCLUSION

The use of 4D anatomical constraints empowers the acacMoCo to achieve improved results compared to the acMoCo. The advantage is given by the precise anatomical prior knowledge which enables us to guide the registration with more realistic physiological motion patterns. The motion estimation is split from a global problem to several smaller motion estimation problems, which enables an organ-specific motion handling. Restrictions for the acacMoCo are given with insufficient prior segmentation or motion occurring in truncated

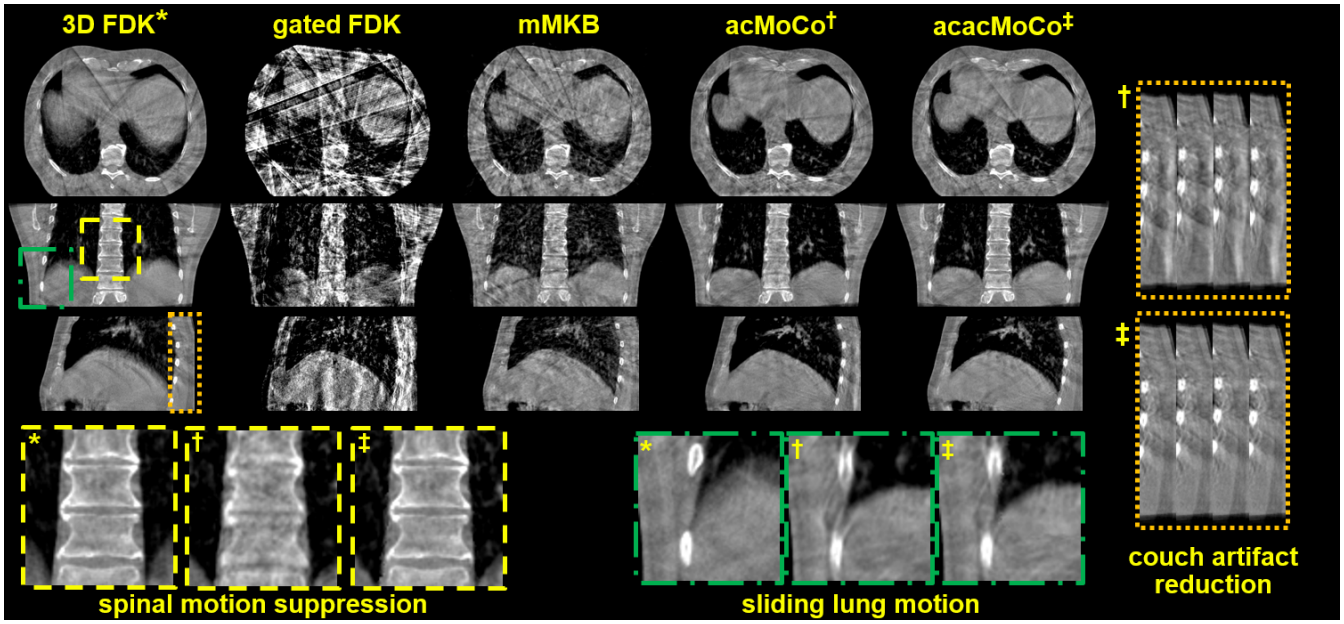


Fig. 4. Comparison of reconstruction methods for an example one minute CBCT patient scan: 3D reconstruction, simple 4D gated reconstruction and the 4D reconstructions modified McKinnon-Bates, acMoCo and acacMoCo. The 3D FDK suffers from motion blurring and the gated FDK has a high temporal resolution but the sparseness reconstruction artifacts limit the image quality. Three zooms show aspects of the acacMoCo improvements, compared to 3D FDK (*), acMoCo (†) and acacMoCo (‡): The recovery of the spine in the acacMoCo achieved with the spinal motion suppression (yellow, long dashed zoom), the effect of the sliding lung motion filter, indicated by the restoration of the rib shape and the couch artifact reduction, here shown with four different phases, which results in a homogenous back and temporal stability.

parts of the body, which is a common problem in MoCo with deformable image registration. However, it seems that small motion artifacts in the initial motion vector fields for the 4D segmentation do not influence the acacMoCo results.

The acacMoCo can be extended with further constraints to consider knowledge of other anatomies as well. Such extensions can feature affinity constraints on the motion vector fields in regions of low contrast or other regularizers on the vector fields.

Acknowledgment

Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.

References

[1] A. P. Santoso, K. H. Song, Y. Qin, S. J. Gardner, C. Liu, I. J. Chetty, B. Movsas, M. Ajlouni, and N. Wen, "Evaluation of gantry speed on image quality and imaging dose for 4D cone–beam CT acquisition," *Radiat. Oncol.*, vol. 11, no. 1, p. 98, Jul. 2016.

[2] C.-C. Shieh, Y. Gonzalez, B. Li, X. Jia, S. Rit, C. Mory, M. Riblett, G. Hugo, Y. Zhang, Z. Jiang, X. Liu, L. Ren, and P. Keall, "SPARE: Sparse–view reconstruction challenge for 4D cone-beam CT from a 1–min scan," *Med. Phys.*, vol. 46, no. 9, pp. 3799–3811, Sep. 2019.

[3] J.-J. Sonke, L. Zijp, P. Remeijer, and M. van Herk, "Respiratory correlated cone beam CT," *Med. Phys.*, vol. 32, no. 4, pp. 1176–1186, Apr. 2005.

[4] F. Bergner, T. Berkus, M. Oelhafen, P. Kunz, T. Pan, R. Grimmer, L. Ritschl, and M. Kachelrieß, "An investigation of 4D cone–beam CT algorithms for slowly rotating scanners," *Med. Phys.*, vol. 37, no. 9, pp. 5044–5053, Sep. 2010.

[5] M. Brehm, P. Paysan, M. Oelhafen, P. Kunz, and M. Kachelrieß, "Self–adapting cyclic registration for motion–compensated cone–beam CT in image–guided radiation therapy," *Med. Phys.*, vol. 39, no. 12, pp. 7603–7618, Dec. 2012.

[6] M. Brehm, P. Paysan, M. Oelhafen, and M. Kachelrieß, "Artifact–resistant motion estimation with a patient–specific artifact model for motion–compensated cone–beam CT," *Med. Phys.*, vol. 40, no. 10, p. 101913 (13 pages), Oct. 2013.

[7] S. Sauppe, J. Kuhm, M. Brehm, P. Paysan, D. Seghers, and M. Kachelrieß, "Sliding organ motion regularisation for motion–compensated cone–beam CT (CBCT) in image–guided radiation therapy (IGRT)," *Insights into Imaging, ECR Book of Abstracts*, vol. 9, no. Suppl. 1, pp. S339, B–1432, Mar. 2018.

[8] S. Sauppe, C. Rank, M. Brehm, P. Paysan, D. Seghers, and M. Kachelrieß, "Motion vector field upsampling for improved 4D cone-beam ct motion compensation of the thorax," *SPIE Medical Imaging Proc.*, vol. 10132, p. 101322E 6, Mar. 2017.

[9] S. Sauppe, J. Kuhm, M. Brehm, P. Paysan, D. Seghers, and M. Kachelrieß, "Motion vector field phase–to–amplitude resampling for 4D motion–compensated cone–beam CT," *Phys. Med. Biol.*, vol. 63, p. 035032 (14 pages), 2018.

[10] J. Star-Lack, M. Sun, M. Oelhafen, T. Berkus, J. Pavkovich, M. Brehm, M. Arheit, P. Paysan, A. Wang, P. Munro, D. Seghers, L. M. Carvalho, and W. F. a. R. Verbakel, "A modified McKinnon–Bates (MKB) algorithm for improved 4D cone–beam computed tomography (CBCT) of the lung," *Med. Phys.*, vol. 45, no. 8, pp. 3783–3799, Jun. 2018.

[11] A. Hahn, M. Knaup, and M. Kachelrieß, "A robust fully automatic method for intrinsic respiratory and cardiac gating for cone–beam CT scans of the thorax region," *Program of the 104th Scientific Assembly and Annual Meeting of the RSNA: SSJ23-05*, Nov. 2018.

[12] L. Feldkamp, L. Davis, and J. Kress, "Practical cone–beam algorithm," *Journal of the Optical Society of America*, vol. 1, no. 6, pp. 612–619, Jun. 1984.

[13] J.-P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.

[14] H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," *Phys. Med. Biol.*, vol. 50, no. 12, pp. 2887–2905, Jun. 2005.

Projection-based cardiac phase estimation: towards ECG-less prospective cardiac CT gating

Eri Haneda, Bernhard Claus, Jed D. Pack, Bruno De Man

***Abstract*—In cardiac CT, triggering a scan during a quiescent cardiac phase is important in order to minimize motion artifacts. In the current workflow, cardiac phases are determined by an ECG (electrocardiogram) device. Typically, the minimum cardiac motion phase is observed at end-systole or mid- to end-diastole. ECG devices are routinely used, with the inherent workflow overhead and sometimes poor signal issues. For example, reinstallation is required if the signal from the leads is not reliably detected. In addition, the current methodology is somewhat limited in the sense that it relies on only a single sample (the R peak) per heartbeat.**

In this paper, a proof-of-concept study to estimate cardiac phase directly from CT projection data using a neural network is presented. We trained a neural network to estimate cardiac phase based solely on CT projections. The basic idea is to perform an ultra-low-dose monitoring scan followed by a standard diagnostic scan. Projection data are obtained during the monitoring stage with sparse view sampling, a low-dose technique and a wide collimator opening. By continuously analyzing the sparse sequence of projection views, we can in real-time estimate the cardiac phase and use that information to prospectively gate the diagnostic scan. In this preliminary study, we used a small number of clinical datasets for training and testing to show the feasibility of extracting cardiac phase information directly from projection data.

***Index Terms*—Computed tomography, Cardiac CT gating, Cardiac phase, Sinogram, Projection data, Neural networks**

I. Introduction

Cardiac CT scanner performance has improved dramatically over the last two decades, with effective temporal resolution now approaching 30 ms thanks to motion compensation algorithms, and several commercial CT scanners now capable of covering the entire heart in one rotation. However, cardiac CT exams are still challenging to perform due to the need to carefully time the scan based on cardiac phase to avoid motion artifacts and also based on when the peak iodine contrast enhancement is reached. To find the minimum cardiac motion, cardiac scans typically require the use of an ECG device, which measures electrical impulses traveling across heart muscle to trigger contractions. With ECG, the cardiac phase within a cardiac cycle (such as end-diastole or end-systole) can be estimated from the time of the R peaks (Fig.1). Typically, minimum cardiac motion is observed in mid- or end-diastole and end-systole phases [1]. In traditional cardiac CT scans, the R peaks are recorded by ECG, and the quiescent phase during which the scan is performed is defined for example as 70-80% of the R-R interval.

While ECG is routinely used, ECG installation to patients adds a substantial time and effort to the exam, comes with additional cost (patches and equipment) and requires additional technologist training. Moreover, there is a chance that the leads provide a poor signal, resulting in aborting the exam, which is most problematic if contrast agent was already administered or radiation dose was already delivered.

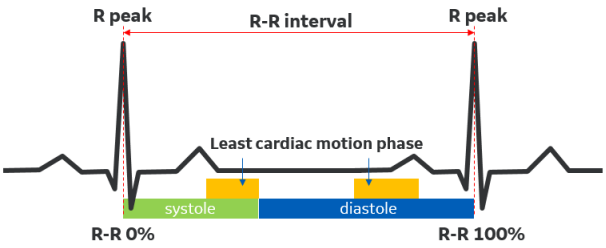


Fig. 1. R-R interval and systole/diastole in ECG signal

In this paper, we will demonstrate a proof-of-concept study of cardiac phase estimation from sparse-view CT projection data with the potential to eliminate the use of an ECG device for gating. The idea of using projection data for cardiac phase estimation was already proposed in past studies. The most relevant studies are discussed in [2, 3], where a direct projection approach was used for automatic cardiac phase selection in retrospective gating. The authors showed that the kymogram signal extracted from the projection data has a strong correlation with the ECG signal and demonstrated that it can be used to select the optimal interval for reconstruction with minimum motion artifacts. The kymogram signal was generated from densely-sampled projections. It was estimated by backprojecting the center-of-mass of the projection views [2, 3].

The ideas of characterizing static and dynamic structures of a human body without image reconstruction were also published previously. For static object detection, Chesler et al. [4] analytically showed that projection-domain SNR is greater than image-domain SNR under simple conditions where

Eri Haneda (haneda@ge.com), Bernhard Claus, Jed D. Pack, and Bruno De Man are with GE Global Research, Niskayuna, NY.

background and target are exactly known. There are also simulation studies that support this claim in the context of attenuation-coefficient estimation and lesion detectability [5], [6]. In [7], blood vessel feature detection and characterization were demonstrated directly from CT projection data.

In this paper, we present a proof-of-concept study of cardiac phase estimation directly from sparse-view CT projection data using a neural network. More specifically, we estimate the cardiac phase (represented by percent of the R-R interval (R-R%)) from a sequence of projection images with wide collimator opening to cover the entire heart. Projection-based cardiac phase estimation enables a new cardiac scanning paradigm which starts with a monitoring phase where the collimation is already wide open (ideally covering the entire heart, like for the diagnostic scan), and the patient's heart is scanned with a low-dose technique and under sparse-view conditions to minimize radiation dose. During this monitoring phase, the acquired projection data is analyzed in real-time to determine the cardiac phase (i.e. R-R%) of each new projection. Once the optimal scan timing is calculated from this signal (along with bolus arrival time), the scan parameters can be switched to a diagnosis scan with standard technique and densely-sampled views. In this paper, our focus is on demonstrating initial feasibility of cardiac phase estimation. Real-time implementation is out of scope.

II. Methods

A. Clinical datasets

For training purposes, we prefer projection data with full (16 cm) coverage, including all cardiac phases (full R-R interval) and accompanied by ECG measurements to be used as label. Such clinical data is not routinely acquired.

Dynamic perfusion data is taken over multiple heart beats, however it is gated so the acquisition is not equally spread over different cardiac phases. Timing bolus scans or bolus tracking scans only acquire data every 1-2 seconds, so the temporal sampling is not high enough. These also use a narrow collimator opening, so only a small portion of the heart is imaged. Coronary computed tomography angiography (CCTA) scans are usually already gated and therefore don't contain all cardiac phases. Un-gated CCTA scans are acquired in special cases such as arrythmia cases, transcatheter aortic valve replacement (TAVR) planning or functional studies [8].

In this study, we used 11 clinical un-gated CCTA datasets used for functional studies with long exposure (>1 sec) for training and testing. R peaks were recorded by ECG, and contrast agent was administered. All scans were acquired using a GE Revolution CT scanner with wide collimator (16cm z-coverage). Settings varied across exams. For example, tube voltage ranged from 80 kVp to 120kVp while rotation speed was 0.28 or 0.35 sec per rotation. Scan durations ranged from 1.0 sec to 2.2 sec, covering at least one full cardiac cycle. In most cases, tube current is significantly lowered outside of a selected cardiac phase (~1/40 to 1/30 relative to the standard tube current (~400mA)). Therefore, a large portion of data had a very high noise level.

Fig. 2 shows an example of Anterior-Posterior (AP) projection view (left) and the intensity changes from the 1st through the 5th rotation. The intensity change was computed by taking difference relative to the minimum value over all 5 projections. Each image is labeled with the corresponding R-R% as a reference. We can observe low and high intensity changes below the aorta region. This is probably due to the decrease and increase of blood volume in left ventricle area during systole and diastole. We observed similar trends in other patients.

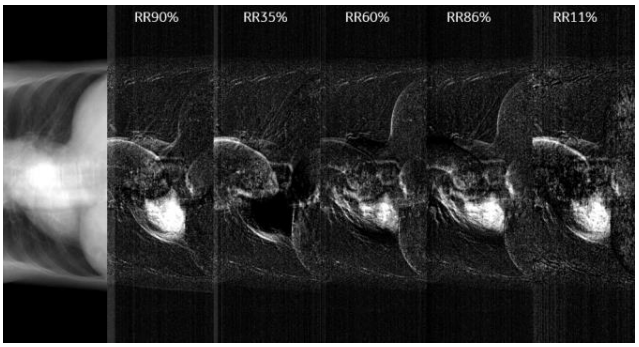


Fig. 2. The most left image is an example of AP view of projection. The rest of images are the intensity changes at every AP view from the 1st rotation through the 5th rotation. The intensity change was computed by subtracting the minimum value over the 5 images for each pixel. Contrast agent was administrated in training and testing data.

B. Architecture

There are many ways to select a sequence of projection data to estimate the R-R%. For example, one projection per rotation, multiple uniformly-distributed projections per rotation, or a few neighboring projections per rotation. Each strategy has advantages and disadvantages. In this preliminary study, we used a single projection at a fixed view angle at every rotation (e.g. AP views in Fig. 2). All resulting projection views are spatially aligned, which should make it easier for a network to learn changes in cardiac phase. The disadvantage of this strategy is the low temporal sampling density. Since the gantry rotates in 0.28 or 0.35 seconds, a projection is acquired only every 28-53% of a cardiac cycle for a regular heart beat with 60 bpm-90 bpm. Multiple projections per rotation would provide higher temporal sampling density, but the task of extracting cardiac phase estimation becomes intuitively more complex. The network requires a large amount of training data. Similarly, one snapshot with neighboring angle views will allow for capturing some type of "depth" information, similar to tomosynthesis.

We implemented four scenarios: a sequence of one, two, three, or four consecutive AP views, and in all cases we estimated the R-R% of the last view. Four AP views span approximately 0.84 to 1.05 seconds, which is close to one heartbeat.

We used simple convolutional neural network (CNN) layers combined with fully connected (FC) layers to estimate R-R%. Figure 3 shows the neural network model we used. The convolution kernel size is 3x3, and max pooling size is [2,2]. ReLU activation functions are used after all convolution and dense layers. Since we use very little training data, the depth was kept relatively shallow (4 repeated CNN layers) to reduce risk of overfitting [9]. Notice that the output of network is not a single value of R-R%. Instead, we used cos(2*pi*R-R%) and

sin(2*pi*R-R%). This is because R-R% is circularly repeated, therefore 0% and 100% are the same phase, consequently the error metric measuring the difference between 1% and 99% should be small. To accommodate the periodicity of the cardiac phase, we mapped R-R% to two values using polar coordinate system: cos(2*pi*R-R%) and sin(2*pi*R-R%). Then, the squared distance between two points in Cartesian coordinates was used as error metric for training. In the training, SGD (stochastic gradient descent) was used for the optimization. The final R-R% was computed by the angle formed by the cosine and sine. The network was implemented using Keras with a Tensorflow backend.

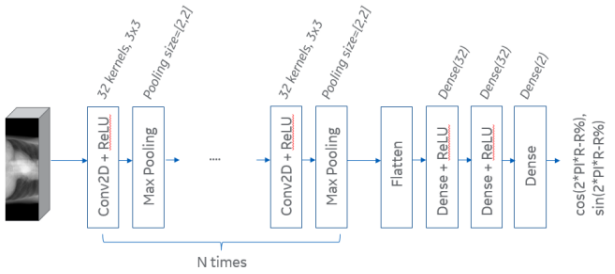


Fig. 3. Neural network architecture. The input is a sequence of 1 to 4 AP projection views, and the output is a 2-dimensional representation of the cardiac phase in the last view (see text).

C. *Preprocessing & data augmentation*

We applied preprocessing steps to each sequence of projection data. We also performed data augmentation to increase the number of training samples. First, a rectangular box which contains the heart was segmented from each patient data. The rectangular box spans all pixels in the detector row direction (i.e. body height direction), and about 42% in the column direction. The center of the box in the column direction was selected by summing all detector columns, low-pass filtering, thresholding and center of mass calculation. Figure 4 shows examples of heart region segmentation. The bounding boxes with red lines are the segmented area. The segmentation was done by using the first AP view of each projection data sequence, and the same segmentation was applied to the other projections in the sequence. The segmented data was down-sampled by 4 after low-pass filtering with a 5x5 Gaussian kernel.

To increase the variation of training samples, data augmentation was performed. We first collected projections over ±5º around each AP view (i.e. 27 views) and labeled each projection with the same R-R% as the central view, where the R-R% was computed from the time stamps at R peaks which were stored in the scan data. Then, each data was resized by factors 1.2, 1.1, 1.0, 0.9, and 0.8, and rotated in the range of ±10º in 1º increments. The data after each step was cropped to the original size. The missing values from resizing and rotation were zero-padded. Finally, the data was randomly shifted in horizontal and vertical directions up to 10%. Each dataset was normalized by dividing with the max value of the first AP view snapshot, then the mean of entire dataset in the sequence was subtracted. Note that the exact same preprocessing was always performed on each AP view within a sequence. Overall, a network was trained with ~70,000 augmented datasets, which originated from 27 datasets extracted from 10 patients and tested on ~8,000 augmented datasets from a completely separate patient dataset.

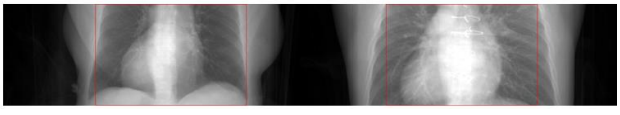


Fig. 4. Examples of heart segmentation with a rectangular box (red line) on AP view data as a preprocessing. The box height and width are fixed. The box spans all pixels in detector row direction (vertical) and spans about 42% in column direction (horizontal).

III. RESULTS

We performed 11 leave-one-out experiments. In each experiment, we excluded one patient data to preserve for testing. The rest of patient data were divided into training (90%) and validation datasets (10%). We repeated the experiment 11 times to train 11 different networks. In this way, we always used a patient dataset that the neural network had not seen for testing.

A. *Training convergence*

Figure 5 shows an example of convergence curves of loss function during training. Each plot corresponds to the case of a sequence of 4 AP views, 3 AP views, 2 AP views, and 1 AP view. The plot shows both curves for training and validation dataset. For all leave-one-out training experiments, the convergence curves were similar. While the convergence speed of 4 AP views and 3 AP views cases were similar and monotonically decreasing, the speed became slower for 2 AP views. The 1 AP view case showed the slowest convergence speed and the curve was not as smooth as other cases.

B. *R-R% results*

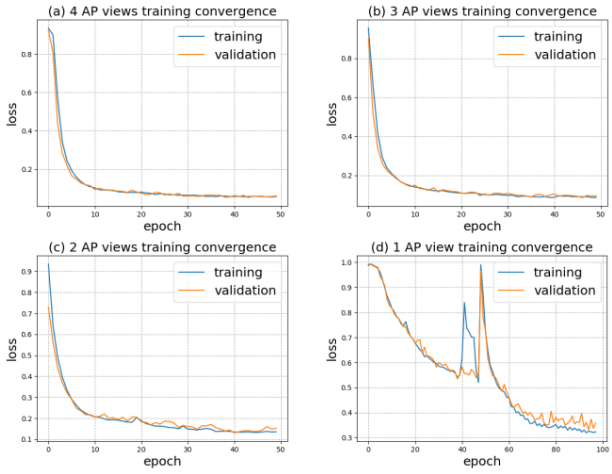


Fig. 5. Convergence of loss function during training. Each plot corresponds to a sequence of (a) 4 AP views (b) 3 AP views (c) 2 AP views, and (d) 1 AP view. SGD (stochastic gradient descent) was used for the optimization.

Figure 6 shows four plots of prediction versus ground-truth R-R% (a sequence of 4 AP views, 3 AP views, 2 AP views, and 1 AP view). Each plot is a combined test result of 11 leave- one-out experiments, therefore combined result of 11 different

trained networks. Multiple prediction points correspond to one ground-truth point because of the data augmentation we performed. Recall that R-R% is circularly continuous, which means 0 (0%) and 1.0 (100%) are the same phase. To compute the error, the prediction value y that is closer to ground-truth x was selected by using the equation: predict = mod(y-x+0.5,1)-0.5. We observed that root-mean-squared-error (RMSE) was 17.2%, 17.6%, 19.1%, and 26.0% in 4 AP views, 3 AP views, 2 AP views, and 1 AP view. As a reference, the RMSE of random predictions using uniform distribution was 28.9%. Although the error level and variance are still relatively large in these early results, we observed an obvious correlation between the ground truth and prediction, which becomes more pronounced when a larger number of AP projection views is used.

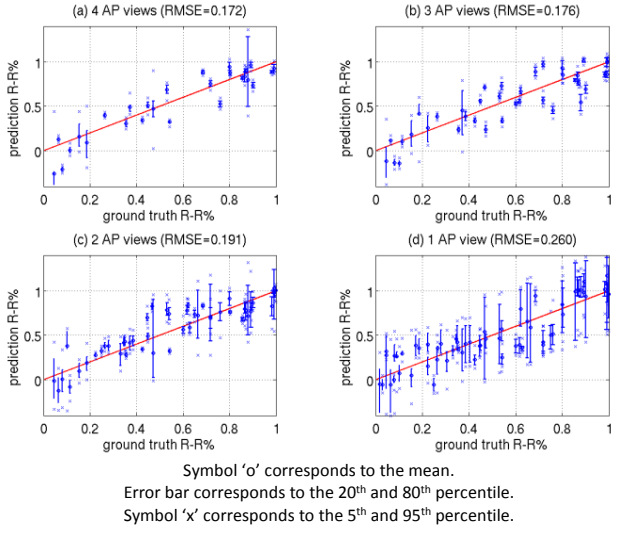


Fig. 6. Prediction vs. Ground-Truth of R-R% using a sequence of: (a) 4 AP views (b) 3 AP views (c) 2 AP views, and (d) 1 AP view. Each plot is a combined result of 11 leave-one-out experiments (test data only).

C. Cosine/Sine scatter plots

Figure 7 shows an example of estimated cosine and sine scatter plot which corresponds to one ground truth R-R% data point in III-b. The ground truth R-R% value was 89% as indicated by a black line. The angle pointing to the right direction corresponds to R-R 0% and the value increases counter-clockwise. Again, multiple predictions points correspond to one ground-truth point because of the data augmentation. As we can see, the scatter plot is more spread out as the number of AP views is reduced. This observation mirrors the results of R-R% in Figure 6.

IV. Conclusion

In this paper, we demonstrated a proof-of-concept study of cardiac phase estimation from a sequence of CT projection data (one AP view per rotation). Despite a limited amount of training data and scan constraints, the results indicate that use of four consecutive AP views gives more accurate predictions with smaller variance than fewer AP views, and results in better correlation between ground truth and estimated cardiac phase. The preliminary results indicate that cardiac phase estimation

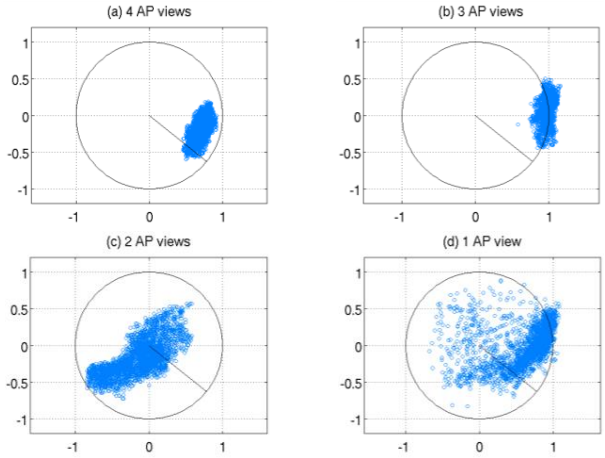


Fig. 7. An example of cosine and sine prediction vs. ground-truth: (a) 4 AP views (b) 3 AP views (c) 2 AP views, and (d) 1 AP view. The ground-truth was indicated by a black line (R-R 89%). The angle pointing to the right direction corresponds to R-R 0% and the value increases counter-clockwise. 2835 predictions points for augmented dataset are presented.

from sparse projections is possible, but more extensive training of deep learning algorithms with more data is needed for an accurate and robust estimator.

References

[1] B. Ohnesorge and T. Flohr, "Principles of Multi-slice Cardiac CT Imaging," in Multi-slice and Dual-source CT in Cardiac Imaging, 2nd ed., Springer, Berlin/Heidelberg, 2007, pp. 71–126.

[2] D. Ertel, T. Pflederer, M. Kachelrieß, S. Achenbach, P. Steffen, and W.A. Kalender, "Validation of a Rawdata-based Synchronization Signal (Kymogram) for a Phase-Correlated Cardiac Image Reconstruction," in *European Radiology*, vol 18, no. 2, pp.253-262, 2008

[3] M. Kachelrieß, D.-A. Sennst, W. Maxlmoser, and W.A. Kalender, "Kymogram detection and kymogram-correlated image reconstruction from subsecond spiral computed tomography scans of the heart," in *Medical Physics*, vol. 29, no. 7, pp. 1489-1503, 2002

[4] D.A. Chesler, S.J Riederer, and N.J. Pelc, "Noise due to photon counting statistics in computed X-ray tomography," in *J. Computer. Assisted. Tomography*, vol. 1, no.1, pp. 64-74, 1978

[5] S. Hsieh, D.A. Chesler, D. Fleischmann, and N. Pelc, "A limit on dose reduction possible with CT reconstruction algorithms without prior knowledge of the scan subject," in *Medical Physics*, vol. 43, no. 3, pp. 1361-1368, 2016

[6] R. De Man, G. Wang, M. K. Kalra, A. Otrakji, S. Hsieh and N. Pelc, "Upper-Bound on Dose Reduction in CT Reconstruction for Nodule Detection," in *IEEE Access*, vol. 4, pp. 4247-4253, 2016

[7] Q. De Man, E. Haneda, B. Claus, P. Fitzgerald, B. De Man, G. Qian, H. Shan, J. Min, M. Sabuncu, and G. Wang, "A two-dimensional feasibility study of deep learning-based feature detection and characterization directly from CT sinograms", in *Medical Physics*, vol. 46, no. 12, pp. e790-e800, 2019

[8] J.E. Scholtz and B. Ghoshhajra, "Advances in cardiac CT contrast injection and acquisition protocols," in *Cardiovasc Diagn Ther*, vol 7, no. 5, pp.439-451, 2017

[9] The Keras Blog, Building powerful image classification models using very little data [Online]. Available: https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html

Adjoint pairs of image warping operators for motion modeling in 4D-CT

Jens Renders, Jan Sijbers and Jan De Beenhouwer

***Abstract*—In 4D-CT, a 4D image (3D+time) needs to be reconstructed from projection data of a moving object. A simple approach is to acquire a number of consecutive 3D-CT scans (called subscans) and to then reconstruct them separately. When the object is moving rapidly, or a high time resolution is desired, a short acquisition time is needed for the subscans, leading to a low number of projections and a strongly underdetermined system to solve. Recently, a number of methods have been proposed to alleviate this issue using image warping operators. By warping an image from a certain time frame to a neighboring time frame, the image can be compared to the projection data of that neighboring time frame, reducing underdetermination. The motion that is used for the warp can be either estimated in advance, or it can be reconstructed jointly with the images. To invert a model that incorporates warping operators, the adjoints of these warping operators are needed. The current methods approximate these adjoint operators by warping operators for an approximate inverse flow. We propose to use the exact adjoints instead, which leads to faster convergence, less computation and less memory requirements.**

***Index Terms*—4D-CT, warping, motion.**

I. Introduction

A 4D-CT scan is a series of n consecutive, regular CT scans called subscans. It is generally assumed that the scanned object is motionless during each subscan, such that the scan can be modeled by the equation $\boldsymbol{W}_i\boldsymbol{x}_i = \boldsymbol{p}_i$, where $\boldsymbol{W}_i$ is the projection matrix, $\boldsymbol{x}_i$ is the scanned object and $\boldsymbol{p}_i$ is the projection data of the i-th subscan. The extent to which this assumption is true depends on the time that passed during the subscan, and the speed of the motion of the object. For this reason, subscans are usually fast scans with few projections. Reconstructing a 4D image from a 4D-CT scan with n subscans then corresponds to solving n linear systems

$$\boldsymbol{W}_i\boldsymbol{x}_i = \boldsymbol{p}_i, \quad i = 1, \cdots, n \quad , \tag{1}$$

which, in this case, are highly underdetermined. Equivalently, the problem can be represented as one big underdetermined system of the form

$$\begin{bmatrix} \boldsymbol{W}_1 & 0 & 0 & 0 \\ 0 & \boldsymbol{W}_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boldsymbol{W}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \\ \vdots \\ \boldsymbol{p}_n \end{bmatrix} \quad . \tag{2}$$

One way to alleviate this underdetermination is to link the time frames together using image registration or optical flow techniques [1]–[4] and image warping [5]. In [6], [7] problem (2) is regularized with terms that constrain the change between frames, i.e. with constraints on the optical flow. In [7], warping operators are involved in these regularization terms. In the MoVIT algorithm [8] and in [9], the images are warped along the flow between frames before they are projected using a standard projection geometry. In [10], [11] the optical flow between frames is accounted for in the projection operators, by using a curved ray projection geometry instead of explicitly involving image warping.

For methods that involve image warping operators in the forward model, or in the regularization terms, the adjoints of these warping operators are needed for reconstruction with gradient based methods. The current methods implement the adjoint image warping operators by warping along an approximated inverse of the flow. Computing the inverse of the flow requires computation time and memory. On top of that, since the flow is generally not exactly invertible and the adjoint of a warp is not exactly the warp along the inverse flow, it introduces inaccuracies.

In [12] a pair of adjoint warping operators with a custom interpolation method is used for respiratory and cardiac motion correction in 4D PET. It is shown that using inverse warps as an approximation for the adjoint warp leads to image degradation compared to using the exact adjoint warps. However, these adjoint warp operators can not be directly substituted for the approximated adjoint warping operators in the previously mentioned methods, because the custom interpolation method of [12] requires more information than general optical flow methods provide. In this paper, we investigate the implementation and the effect of adjoint warping operators using the more generally applicable and commonly used multivariate spline interpolation.

II. Methods

Although, our proposed method to calculate the adjoint warping operators is applicable to any method that uses image warping with multivariate spline interpolation, including [7]–[9], we will demonstrate their use on a basic method similar to [8]. In a first step, each time frame is reconstructed separately by solving (2). On these initial reconstructions, we use an optical flow algorithm to estimate the flow between the time frames. Finally, we use a least squares solver to solve the same dynamic tomographic model that MoVIT solves.

A. Warping operators

While the scanned object can change during a 4D-CT scan, we assume that the materials that make up the object do not

J. Renders, J. Sijbers and J. De Beenhouwer are with imec-Vision Lab, a research group at the University of Antwerp, Belgium (email: jens.renders@uantwerpen.be)

change, but only get repositioned[1]. Under this assumption, we can deform an image $\boldsymbol{x}_i \in \mathbb{R}^N$ into the image $\boldsymbol{x}_j \in \mathbb{R}^N$ of a different time frame, by moving its voxel values without changing them. For each voxel, a vector in $\mathbb{R}^3$ describes its displacement. Together, these displacement vectors form a displacement vector field or deformation vector field (DVF) representing the optical flow between the images.

Moving the voxel values according to the DVF results in non-grid data, because the voxels are allowed to move to non-integer coordinates. To turn the result back into an image in $\mathbb{R}^N$, resampling is required. General image warping is the combined action of moving the voxels and resampling. A standard choice of resampling method is multivariate spline interpolation (usually linear or cubic splines), used in for example [5], [8], [9]. With this choice each voxel in the warped image is a linear combination of voxels in the original image, so such warping operators are linear maps. We will write $\boldsymbol{M}_{ij}$ to denote a warping operator that transforms $\boldsymbol{x}_i$ into $\boldsymbol{x}_j$.

B. Dynamic tomographic model

Suppose that we want to reconstruct a certain time frame of interest, $\boldsymbol{x}_i$. In an ideal situation, any time frame $\boldsymbol{x}_j$ can be produced from $\boldsymbol{x}_i$ by using a suitable warping operator $\boldsymbol{M}_{ij}$, such that $\boldsymbol{x}_j = \boldsymbol{M}_{ij}\boldsymbol{x}_i$ holds for all j. This can be written as:

$$\begin{bmatrix} \boldsymbol{M}_{i1} \\ \boldsymbol{M}_{i2} \\ \vdots \\ \boldsymbol{M}_{in} \end{bmatrix} \boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} \quad . \tag{3}$$

Substituting (3) into (2) yields the dynamic tomographic model:

$$\begin{bmatrix} \boldsymbol{W}_1 & 0 & 0 & 0 \\ 0 & \boldsymbol{W}_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boldsymbol{W}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_{i1} \\ \boldsymbol{M}_{i2} \\ \vdots \\ \boldsymbol{M}_{in} \end{bmatrix} \boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \\ \vdots \\ \boldsymbol{p}_n \end{bmatrix} \quad . \tag{4}$$

or more concisely:

$$\boldsymbol{W}\boldsymbol{M}\boldsymbol{x} = \boldsymbol{p} \quad . \tag{5}$$

In practice, (3) can not be achieved exactly, but by estimating the flow between initial reconstructions, a good approximation can be obtained. System (5) is the one that is solved by MoVIT, and it can be interpreted as a factorization of the model presented in [10] and [11], where the DVFs are used to directly modify the projection matrix $\boldsymbol{W}$ instead of adding the extra factor $\boldsymbol{M}$. The new system has the same number of equations as (2), but the unknowns are reduced to only the voxels of the chosen time frame.

C. Estimating the DVFs

There are many image registration and optical flow algorithms available to estimate the DVFs for the warping operators in (4) such as [1]–[4]. We use the TV-L1 optical flow algorithm provided by scikit-image, which is detailed in [13].

[1]This assumption is only approximately valid, which puts a limit on the number of time frames we will be able to combine.

D. Least squares solvers

There are many choices of least squares solvers for the system (4). These different choices can enforce different types of constraints on the solution. In [7], [9] the Primal Dual Hybrid Gradient method of [14] is used to enforce sparsity of the gradient of the solution. MoVIT uses a modified version of SIRT with an intuitive interpretation. We choose to use a basic (projected) gradient descent to demonstrate the effect of using the exact adjoints. This is very close to what was done in MoVIT as SIRT is a slightly altered version of gradient descent [15]. We use the Barzilai–Borwein step size rule [16] and we enforce minimum and maximum constraints on the voxel values.

E. Adjoint pairs of warping operators using multivariate spline interpolation

There are two different approaches to implement image warping operators, referred to as forward and backward warping [5]. Assume we have two images, $\boldsymbol{a}$ and $\boldsymbol{b}$ and a DVF describing the flow from $\boldsymbol{a}$ to $\boldsymbol{b}$. With forward image warping, the voxels of $\boldsymbol{a}$ are first moved along the DVF to obtain non-grid data representing the warped version of $\boldsymbol{a}$, and this non-grid data is then resampled at grid points to get an image similar to $\boldsymbol{b}$. Backward image warping is an other approach, where the DVF is followed in the opposite direction. For each voxel, we look at the position it is sent to by the DVF and interpolate the regular grid data of $\boldsymbol{b}$ at that point. The result is an image similar to $\boldsymbol{a}$.

Both forward and backward warping are viable methods to implement the warping operators of (4). If we want to be able to transform $\boldsymbol{x}_i$ into $\boldsymbol{x}_j$, we can either get a DVF from $\boldsymbol{x}_i$ to $\boldsymbol{x}_j$ and use forward warping, or we can get a DVF from $\boldsymbol{x}_j$ to $\boldsymbol{x}_i$ and use backward warping. We implement our warping operators using backward warping with rectangular multivariate spline interpolation. For convenience, we will work with 2D images and bilinear splines, but the methods are easily extendable to 3D images and higher order splines. In [8] and [9], cubic splines are used.

Our image $\boldsymbol{x}$ can be thought of as a function

$$\boldsymbol{x} : [n] \times [m] \to \mathbb{R} \quad , \tag{6}$$

where $n, m \in \mathbb{N}$ and $\forall k \in \mathbb{N} : [k] = \{1, \cdots, k\}$. For each pair of integer coordinates, it yields a voxel value. Its bilinear interpolation at a non-integer position (a, b) is given by

$$x'(a, b) = c_1\boldsymbol{x}(\boldsymbol{q}_1) + c_2\boldsymbol{x}(\boldsymbol{q}_2) + c_3\boldsymbol{x}(\boldsymbol{q}_3) + c_4\boldsymbol{x}(\boldsymbol{q}_4) \quad , \tag{7}$$

where $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_4$ are the integer valued points surrounding (a, b), and $c_1, \cdots, c_4$ are the bilinear spline coefficients.

Our DVF $\boldsymbol{v}$ can be thought of as a function

$$\boldsymbol{v} : [n] \times [m] \to \mathbb{R}^2 \tag{8}$$

which yields a 2D displacement vector for each voxel. Our backward warping operator $\boldsymbol{M}$ transforms $\boldsymbol{x}$ into a new image $\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x}$ of which the (i, j)-th voxel value is obtained by sampling $\boldsymbol{x}$ at position $(i, j) + \boldsymbol{v}(i, j)$. That is:

$$(\boldsymbol{M}\boldsymbol{x})(i, j) = x'((i, j) + \boldsymbol{v}(i, j)) \quad . \tag{9}$$

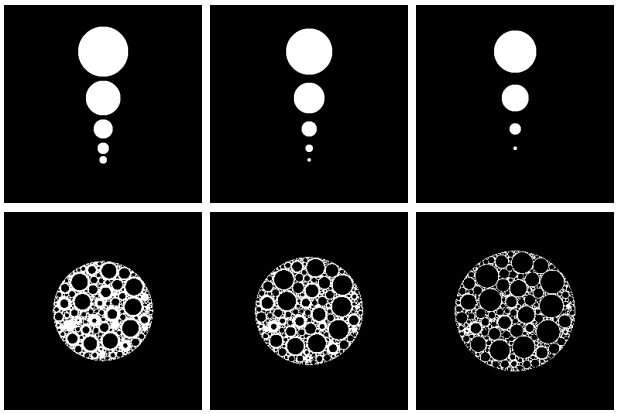


Fig. 1: Ground truths for our experiments. The top row is a 2D+time phantom consisting of 3 time frames with shrinking circles. The bottom row shows 3 time frames of a slice of a 4D simulated forming foam. The actual foam phantom has one time frame per projection: 128×3.

Since, by (7), (9) is simply a linear combination of voxel values of $\boldsymbol{x}$, we can interpret the action of $\boldsymbol{M}$ as a matrix vector product. To do this, we need to represent the image as a vector again: $\boldsymbol{x} \in \mathbb{R}^N$, where each voxel gets a single integer index. $\boldsymbol{M}$ can then be represented by a matrix, with 4 non-zero coefficients on each row, namely the bilinear spline coefficients of (7) at the corresponding voxel indices. The adjoint operator $\boldsymbol{M}^T$ is then simply given by the matrix with the rows of $\boldsymbol{M}$ as its columns. The rows of $\boldsymbol{M}$ or equivalently, the columns of $\boldsymbol{M}^T$ can by computed on the fly, so there is no need to store these matrices. If we denote the i-th row of $\boldsymbol{M}$, i.e. the i-th column of $\boldsymbol{M}^T$ by $\boldsymbol{r}_i$, then the action of $\boldsymbol{M}^T$ on a vector $\boldsymbol{y} \in \mathbb{R}^N$ can be implemented as follows:

$$\boldsymbol{M}^T \boldsymbol{y} = \sum_{i=1}^{N} y_i \boldsymbol{r}_i \quad . \tag{10}$$

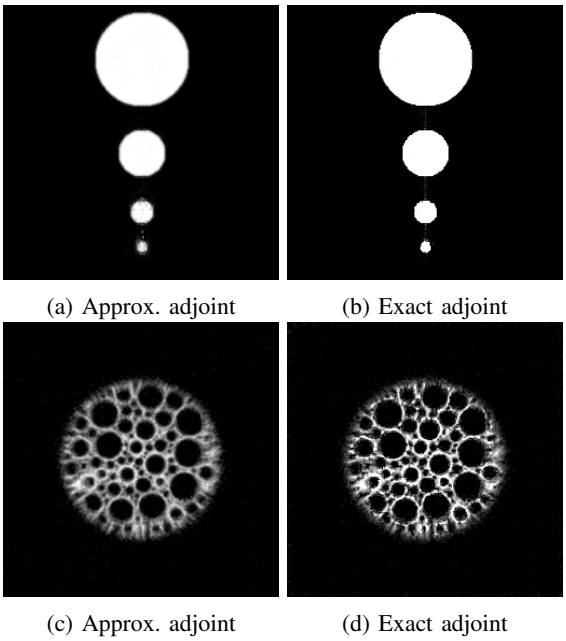
For comparison, we also implemented the approximation of the adjoint by using the inverse DVF. Like [8] and [10], we calculate the inverse DVF with the iterative method presented in [17] with 15 iterations. The action of $\boldsymbol{M}^T$ is then approximated by using the same warping method as $\boldsymbol{M}$, but with this inverse DVF.

III. Experiments

To compare the effects of the exact adjoints to the approximated adjoints, we performed two simulation experiments.

The first experiment used a simple 512×512 2D+time phantom of 5 circles of different sizes which shrink in 3 time frames. We performed 3 subscans, one for each time frame, during which the phantom is static. Each subscan consisted of 128 parallel beam projections over a range of π rad. The DVF's were calculated on the ground truth time frames.

The second experiment used a 256×256 slice of a 4D (3D+time) phantom of growing spheres in a cylinder, mimicking the formation of a foam [18]. This phantom was generated at as many time frames as projections, such that each projection sees a slightly different object. We performed a simulated scan with 3 subscans. Each subscan consisted of 128 parallel beam projections over a range of π rad. In this experiment, we first made initial reconstructions by solving (2). We then calculated the DVFs on these initial reconstructions.



(a) Approx. adjoint (b) Exact adjoint

(c) Approx. adjoint (d) Exact adjoint

Fig. 2: The resulting reconstructions after 30 iterations. Figures (a) and (b) have been zoomed in on the smaller circles to highlight the details.

In both experiments, we reconstructed the middle time frame using both of its neighbors by solving (4) with 3 time frames. The algorithm was ran for 30 iterations. Both phantoms are $\{0, 1\}$-valued, so we enforced a minimum constraint of 0 and a maximum constraint of 1 on the reconstruction.

All projection operators, both for the simulations and for the reconstruction, were implemented using the ASTRA toolbox [19]–[21]

IV. Results and discussion

The implementation using the approximate adjoints required the storage of two DVFs per neighboring time frame, and one image. A DVF consists of one displacement vector per voxel, so it takes as much storage as two images for a 2D problem, or three images for a 3D problem. In our experiments, a total storage of 9 images was required. The implementation using the exact adjoint required only the storage of one DVF per neighboring time frame, cutting the memory requirements almost in half, to a total of 5 images.

The effect of the choice of adjoints can be visually observed in Fig. 2. After 30 iterations, the reconstruction with the exact adjoints is visibly sharper than the reconstruction with the approximate adjoints. Some bubbles which can barely be discerned in Fig. 2(c) are clearly visible in Fig. 2(d), while Fig. 2(d), used the same amount of iterations and less computation time. In Fig. 2(b) the sharper edges are clearly visible.

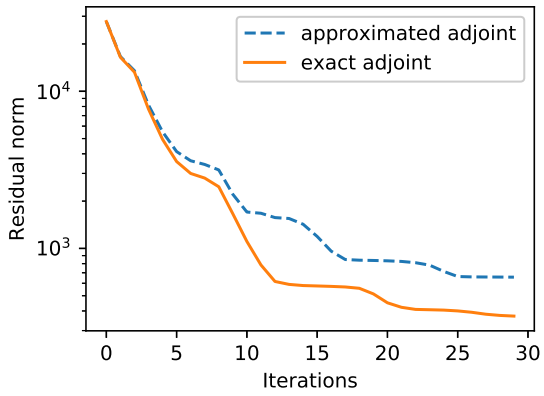


Fig. 3: The difference in convergence when using the exact adjoints versus the approximated adjoints on the circle phantom

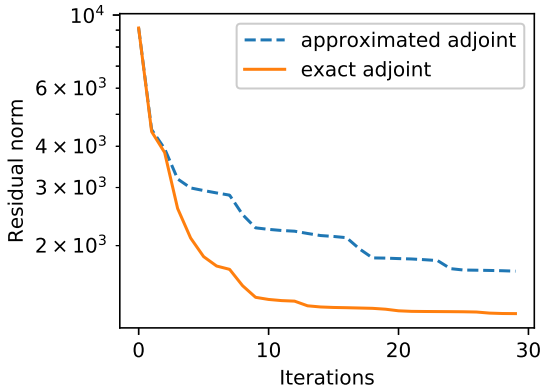


Fig. 4: The difference in convergence when using the exact adjoints versus the approximated adjoints on the foam phantom

The faster convergence can also be verified in the residual plots in Fig. 3 and Fig. 4. For each iteration, we plotted the 2-norm of the residual, i.e. $\|\boldsymbol{A}\boldsymbol{x}_i - \boldsymbol{p}\|_2$, where $\boldsymbol{A} = \boldsymbol{W}\boldsymbol{M}$. We see that the method with the exact adjoints gets the residual down much earlier, and keeps it lower at all iterations.

V. Conclusion

Although the adjoint of an image warping operator can be approximated by the same kind of warping operator with an approximate inverse of its DVF, we have shown that it is beneficial to use the exact adjoint instead. Not only does it provide a faster rate of convergence, it also avoids several problems that the approximate adjoints impose. Calculating an inverse of a DVF is computationally intensive and takes as much space as three images. When scaling to large 3D problems, this becomes an issue. The inverting of a DVF also requires the choice of an extra parameter, namely the number of iterations for the inversion algorithm. These problems are not present with the proposed approach of computing the adjoints.

Acknowledgment

This research is funded by the Fund for Scientific Research-Flanders (FWO) grant nr. S007219N and grant nr. 1SA2920N.

References

[1] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[2] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, ser. IJCAI'81. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1981, p. 674–679.

[3] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L1 optical flow," in *Joint Pattern Recognition Symposium*. Springer, 2007, pp. 214–223.

[4] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.

[5] G. Wolberg, *Digital Image Warping*. IEEE Computer Society Press Los Alamitos, CA, 1990, vol. 10662.

[6] M. Burger, H. Dirks, L. Frerking, A. Hauptmann, T. Helin, and S. Siltanen, "A variational reconstruction method for undersampled dynamic X-ray tomography based on physical motion models," *Inverse Problems*, vol. 33, no. 12, p. 124008, 2017.

[7] G. Zang, R. Idouchi, R. Tao, G. Lubineau, P. Wonka, and W. Heidrich, "Space-time tomography for continuously deforming objects," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, p. 100, 2018.

[8] V. Van Nieuwenhove, J. De Beenhouwer, J. Vlassenbroeck, M. Brennan, and J. Sijbers, "MoVIT: a tomographic reconstruction framework for 4D-CT," *Optics Express*, vol. 25, no. 16, pp. 19 236–19 250, 2017.

[9] G. Zang, R. Idoughi, R. Tao, G. Lubineau, P. Wonka, and W. Heidrich, "Warp-and-project tomography for rapidly deforming objects," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, p. 86, 2019.

[10] M. Odstrcil, M. Holler, J. Raabe, A. Sepe, X. Sheng, S. Vignolini, C. G. Schroer, and M. Guizar-Sicairos, "Ab initio nonrigid X-ray nanotomography," *Nature Communications*, vol. 10, no. 1, p. 2600, 2019.

[11] K. Ruymbeek and W. Vanroose, "Algorithm for the reconstruction of dynamic objects in CT-scanning using optical flow," *Journal of Computational and Applied Mathematics*, vol. 367, p. 112459, 2020.

[12] T. Feng, J. Wang, G. Fung, and B. Tsui, "Non-rigid dual respiratory and cardiac motion correction methods after, during, and before image reconstruction for 4d cardiac pet," *Physics in Medicine & Biology*, vol. 61, no. 1, p. 151, 2015.

[13] J. S. Pérez, E. Meinhardt-Llopis, and G. Facciolo, "TV-L1 optical flow estimation," *Image Processing On Line*, vol. 2013, pp. 137–150, 2013.

[14] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[15] P. C. Hansen, *Discrete inverse problems: insight and algorithms*. Siam, 2010, vol. 7, ch. 6, pp. 111–112.

[16] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 1988.

[17] M. Chen, W. Lu, Q. Chen, K. J. Ruchala, and G. H. Olivera, "A simple fixed-point approach to invert a deformation field," *Medical Physics*, vol. 35, no. 1, pp. 81–88, 2008.

[18] D. Pelt, K. J. Batenburg, and J. Sethian, "Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks," *Journal of Imaging*, vol. 4, no. 11, p. 128, 2018.

[19] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible X-ray tomography using the ASTRA toolbox," *Optics Express*, vol. 24, no. 22, pp. 25 129–25 147, 2016.

[20] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.

[21] W. J. Palenstijn, K. J. Batenburg, and J. Sijbers, "Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs)," *Journal of Structural Biology*, vol. 176, no. 2, pp. 250–253, 2011.

A method for reduction of axial truncation artifacts

Stanisav Žabić
UIH America, Inc.
Email: zabic.stanislav@united-imaging.com

Liuchun He and Yuan Bao
Shanghai United Imaging Healthcare Co., Ltd.
Email: liuchun.he@united-imaging.com and
yuan.bao@united-imaging.com

***Abstract*—We introduce a new, simple extrapolation and HU normalization technique for axial truncation in CT. The method is approximate, but effective as shown on the clinical cases. Biases in the extended field of view lead to errors in the PET reconstruction which uses CT images for attenuation and scatter correction. Thus, the new algorithm is evaluated using strict requirements for accuracy in PET images of water phantoms.**

I. Introduction

Gantry of a CT system rotates around the isocenter with focal spot location and center of the detector system located opposite from one another, typically at approximately the same distance from the center of rotation. This distance is referred to as the bore radius r_{bore} of a CT system. Fan angle of the detector system, however, determines the maximum field of view available for reconstruction $r_{\text{fov}} < r_{\text{bore}}$. All voxels beyond the maximum field of view, and still within the bore radius, cannot be determined precisely with any of the filtered backprojection (FPB) methods and are subject to so called truncation artifacts, especially when a scanned object extends beyond the maximum field of view.

CT voxels in the extended field of view, however, play an important role for the attenuation and scatter correction (AC and SC) in PET image reconstruction. Attenuation correction ensures that the standard uptake values (SUV) are correctly estimated. Scatter interactions can lead to errors in estimation of the initial direction in gamma-rays, thus without SC, a nonlinear blurring would appear in the PET image [1], [2].

AC and SC both need the physical information of object to perform the estimation for interactions between gamma-rays and the object. The electron density, typically represented as attenuation coefficient, are derived from the CT images [1], [2]. Consequently, assuming the AC-CT image and PET image are aligned perfectly, AC and SC rely on accurate CT images to derive the attenuation coefficient within the entire bore of the scanner, with truncation artifacts hopefully reduced as much as possible. Thus, with incorrect HU values, the quantification and quality of PET image also decreases. In other words, obvious truncation artifacts in CT can lead to global SUV errors and they can be either too high, or too low, inside and outside of the maximum field of view, even when the CT image within the maximum field of view is correct.

In CT literature, there are extrapolation methods which allow for stable reconstruction of voxels within the maximum field of view. However, the voxels outside of r_{fov} can only be approximately determined. For example, previously published methods [4], [5] achieve stable reconstruction within the maximum field of view, and another previously published method [6] does a reasonably good reconstruction even for the voxels outside of the maximum field of view. The latter method however, proposes reprojections, with two or more recursive applications of FBP which can be computationally expensive.

In this paper we examine a new method for the reconstruction outside of the maximum field of view without reprojections and we use SUV measurements from the PET image to determine the success criteria of the new method.

II. Method

Our new method for reduction of truncation artifacts consists of two steps. One is extrapolation in projection domain prior to FBP, and the other is HU correction in the image domain after the FBP reconstruction.

A. Projection domain extrapolation

We will assume that all typical pre-processing steps are performed on the CT data, including air bias subtraction, scatter correction and water beam hardening correction. We will also assume just one focal spot position, but without loss of generality, similar correction can be performed to any focal spot configuration, or even after the spots are interleaved. Another assumption is that the detectors are equally sampled in radial direction, which is also not a hindrance because reconstruction algorithms often employ rebinning to make them equally sampled.

If a scanned object was sitting in the extended area, some projections will be truncated. The extrapolation method relies of estimating two main factors: estimation of the trend of the projection (whether it is increasing or decreasing) and the extent of the truncated object past the detector fan.

Suppose projection data for a fixed view and detector data is given with vector p with the total number of channels N_{ch}. This data provides coverage of the maximum field of view radius r_{fov}. The goal is to form a new vector p_2 with an extended number of channels $N_{\text{ch2}} = N_{\text{ch}} + 2N_{\text{extra}}$, where the number of extra channels N_{extra} is large enough to provide virtual coverage of the entire bore radius r_{bore} from each side of the detector array. We first perform Gaussian smoothing with the standard deviation of 10mm to eliminate the impact of noise. Denote the smoothed vector as $\bar{p}$.

To estimate the trend of the projection, we perform the linear extrapolation of the truncated projection. Suppose the truncation happened on the left side of the vector p. Similar

right end truncation can be done by simply flipping the order of elements in vector p. The points we use for the extrapolation are the points between the first two points of the smoothed vector $\bar{p}$:

$$\bar{p}_2(l) = (l - N_{\text{ch2}}) \cdot ((\bar{p}(2) - \bar{p}(1)) + \bar{p}(1),$$

for $l = 1, ..., N_{\text{extra}}$.

To estimate the extent of the object, we have observed that the object size beyond the truncation depends on the amplitude of the object at the truncation point. The higher the amplitude is, the object is larger and it extends more.

$$E \propto \bar{p}(1),$$

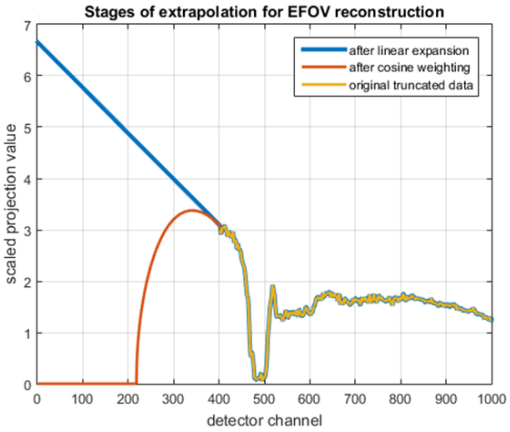
where E denotes the number of detector channels for which the projection needs to be extended before the object vanishes.

Finally, extended vector p_2 is formed using the cosine weight of the linear extrapolation, and valid measurements p:

$$p_2(l) = \begin{cases} 0 & l < E_1, \\ \cos\left(\frac{\pi(N_{\text{extra}} - l)}{2(N_{\text{extra}} - E_1)}\right) \cdot \bar{p}_2(l) & E_1 <= l <= N_{\text{extra}}, \\ p(l - N_{\text{extra}}) & l > N_{\text{extra}}, \end{cases}$$

where $E_1 = N_{\text{extra}} - E$ is the number of the channel at which the object is estimated to vanish. An illustrative example is Fig. 1.

Fig. 1. Original truncated data in gold is first extended using linear interpolation in blue, then the vanishing point is determined and cosine weighting is applied in red.



B. Image domain HU correction

Using the projection domain extrapolation alone, post reconstruction image volumes can still show streaks in the air and shading of the soft tissue. However, at this stage, image domain can be accurately segmented into air, soft tissue and bone. This segmentation is further used to *"flatten"* the HU numbers around those artifacts. Suppose the air is between -1600HU and -400HU, soft tissue is between -400HU and +400HU and bone tissue is above +400HU. For every CT image I we get three images I_a, I_s and I_b for air, soft and bone tissue respectively using the simple thresholding for segmentation.

In the next step, we create a parametric image I_p with -1000HU wherever I_a determined air, with 0HU wherever I_s determined soft tissue and I_b values elsewhere.

In the next step, we blend the low frequencies of I_p with high frequencies of the original image I:

$$I_{\text{blend}} = I + \text{lowpass}(I_p - I).$$

Particular low-frequency blending denoted by lowpass is performed by the 2D Gaussian smoothing with standard deviation of 15 mm.

Finally, we recommend to further blend original image I and I_{blend} so that 100% of original image I is used for voxels within r_{fov} and only a portion of I_{blend} is used in the truncated region. This is done in the following way. Let w be a matrix with the same dimension as the original image with values set as

$$w(i,j) = \frac{r(i,j) - r_{\text{fov}}}{p \cdot r_{\text{fov}}},$$

where $r(i,j)$ is the radial coordinate of the voxel with matrix indices i, j and $p \in (0, r_{\text{bore}}/r_{\text{fov}} - 1)$ is a tuning parameter. Additionally, set to 1 all values of w that are greater than 1 and to 0 all values that are less than 0. The final image is

$$I_{\text{adj}}(i,j) = I_{\text{blend}}(i,j) \cdot w(i,j) + I(i,j) \cdot (1 - w(i,j)).$$

III. Clinical Examples

Clinical images were acquired using helical scans with pitch of 0.9 from a CT system with 64 detector rows, bore radius of 350mm and maximum field of view of 250mm. Patient had their arms crossed over their chest and parts of their arms were hanging in the extended field of view area. Images were reconstructed using the symmetric data extrapolation [4] and our method. An example of achieved results is presented in Fig. 5 and 6. Visual image inspection shows that both methods performed equally well within 250mm radius. However, the symmetric data extrapolation had biases outside 250mm radius, not only inside of the object, but also in the air region as well. Our new method performed with reduced biases in the extended field of view.

IV. Water Phantom Testing

We scanned a water phantom on PET/CT system with additional water bottles hanging outside of the maximum field of view. Particular scanning configuration in CT had a bore radius of 350mm and maximum field of view radius of 250mm. Helical geometry with pitch factor of 0.9 on a scanner with 16 detectors was used. Theoretical attenuation correction is $0.096\text{cm}^2/\text{g}$ at 511keV and the theoretical PET activity is expected to be uniformly distributed at 1.0 SUV. We used two methods to reconstruct CT images of the whole object within the system bore, including the voxels outside of r_{fov}. One method used simple symmetric extrapolation [4] and the other method is the new one described in this paper (see Fig 2). Then we performed the PET reconstruction [3] to evaluate the SUV using two different CT reconstruction as AC-CT input (see Fig. 3). Fig. 4 shows that the mean value of SUV within the

water phantom was correctly reconstructed when we use PET reconstruction with AC-CT images from the new truncation correction method, while distribution of SUV values is clearly shifted to a low number in symmetric data correction. These low numbers are mainly due to the scatter over-correction. The artifact in the extended field of view CT lead to overestimation of SUV by wrong AC, thus resulting in wrong estimation of out-of-object scatter events.

Fig. 2. CT reconstruction of a water phantom with two bottles on each side. On the left, we see the reconstruction performed using the previously published method with symmetric data extrapolation. CT number biases outside of r_{fov}, inside and outside of the object lead to wrong SUV in PET image. On the right side, we see the reconstruction performed with the new method. Display window is 1800HU with the center at 1000HU.

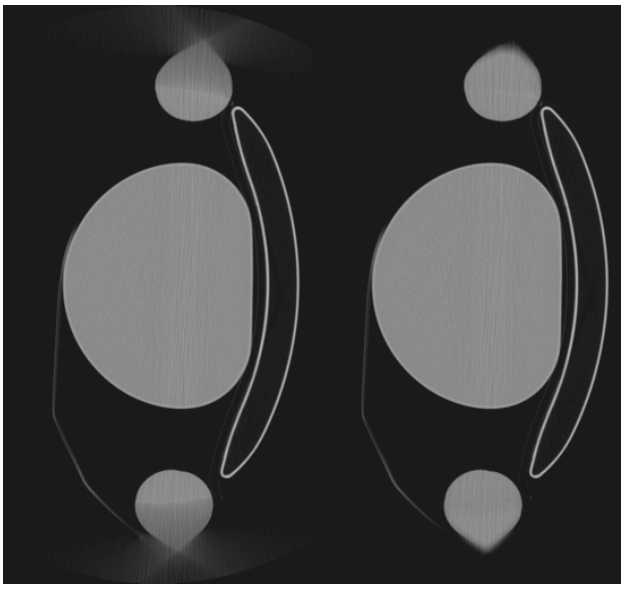
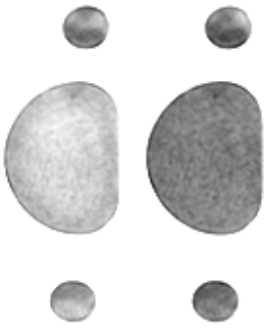


Fig. 3. PET reconstruction with AC-CT coming from the symmetric data extrapolation method (on the left) and from the new extrapolation method (on the right). Center is 1SUV, window is 0.7SUV.

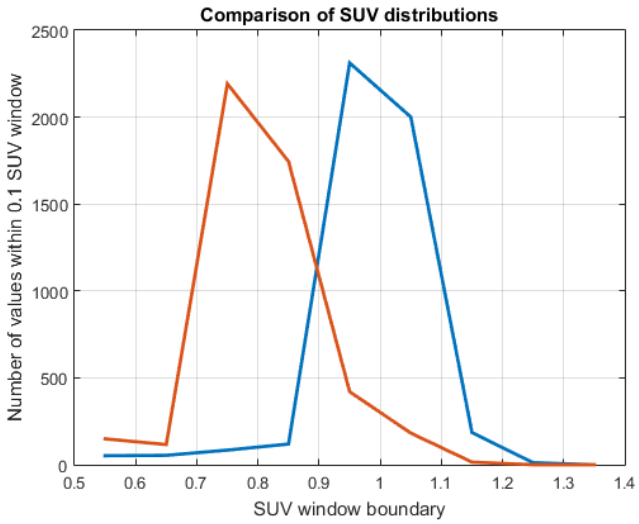


V. Conclusion and Discussion

The proposed extrapolation method with image processing delivers a simple and efficient reconstruction of CT images within the entire bore, without reprojections. When compared to symmetric extrapolation, the new method delivers much more accurate SUV results in water phantom experiment. Central area of the water phantom experiment is assessed accurately, and there are still slight deformations in the shoulder region of the phantom experiment (bottles), which we plan to address in our future research.

Another topic of further research is to compare this method more carefully with other published correction methods. We also mention, that reprojection of the corrected image after the reconstruction with the new method could possibly lead to further refinement of the extended FOV image, albeit with more processing. This is also left as followup research activity.

Fig. 4. Comparison between the SUV distributions with AC images reconstructed using the new approach (blue) and using the symmetric data extrapolation (red).



References

[1] Bonvento M J. Physics in Nuclear Medicine, 4th Edition[J]. Health Physics, 2013, 105(1).

[2] Peter E. Valk, Dale L. Bailey, David W. Townsend, and Michael N. Maisey. Positron Emission Tomography, Basic Science and Clinical Practice, New York: Springer-Verlag; 2004

[3] Badawi R D, Shi H, Hu P, et al. First Human Imaging Studies with the EXPLORER Total-Body PET Scanner. The Journal of Nuclear Medicine, 2019, 60(3): 299-303.

[4] Ohnesorge, B., Flohr, T., Schwarz, K., Heiken, J. P., Bae, K. T. (2000). Efficient correction for CT image artifacts caused by objects extending outside the scan field of view. Medical Physics, 27(1), 39 - 46.

[5] Hsieh, J., Chao, E., Thibault, J., Grekowicz, B., Horst, a., McOlash, S., Myers, T. J. (2004). A novel reconstruction algorithm to extend the CT scan field-of-view. Medical Physics, 31(9), 2385.

[6] Hsieh, S. S., Nett, B. E., Cao, G., Pelc, N. J. (2014). An algorithm to estimate the object support in truncated images. Medical Physics, 41(7), 071908.

Fig. 5. CT reconstruction of a clinical case. Above, we see the reconstruction performed using the method with symmetric data extrapolation. Note the CT number biases in the air region. On bottom, we see the reconstruction performed with the new method. Elbows were represented more accurately and biases in the air are no longer present. Window is 1100HU and center is at 825HU.

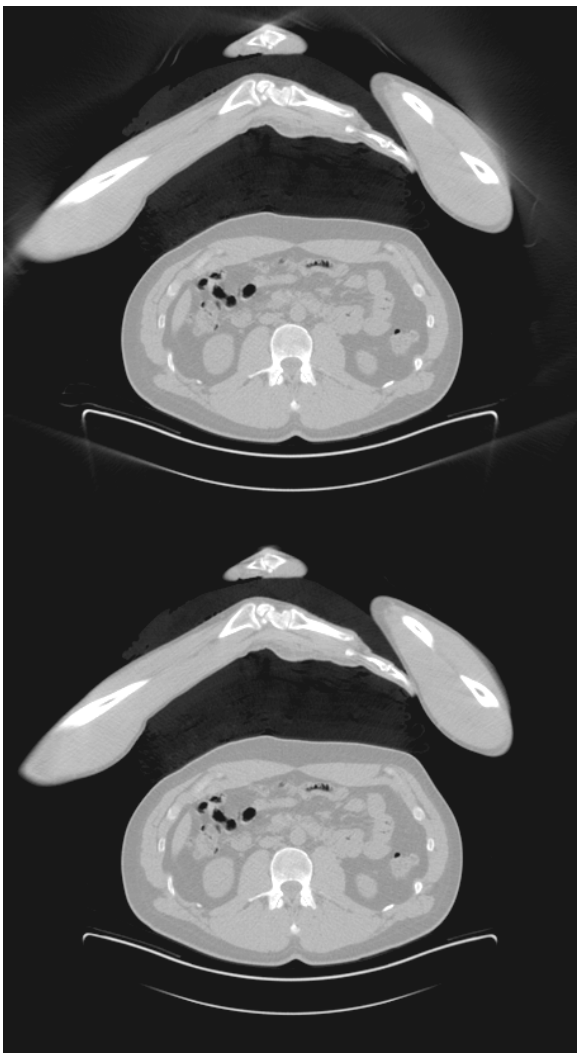
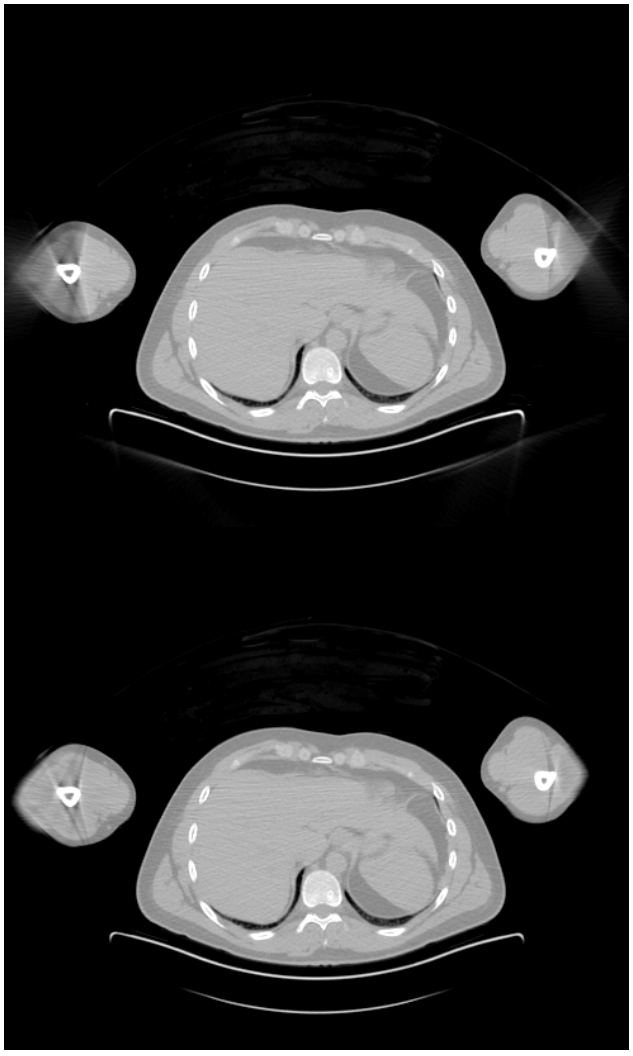


Fig. 6. A different image location of the same clinical case as in Fig. 5, with the same ordering and display setting.



Three different ways of implementing cycloidal computed tomography: a discussion of pros and cons

Oriol Roche i Morgó, Fabio Vittoria, Marco Endrizzi, Alessandro Olivo, and Charlotte K. Hagen

***Abstract*—We present three implementation strategies for cycloidal computed tomography. The latter refers to an imaging concept that enables the acquisition of high-resolution tomograms in a flexible manner (e.g. with x-ray sources with a relatively large focal spot and detectors with relatively large pixels). In cycloidal computed tomography, the sample is rotated and laterally translated simultaneously; with this scheme, each sample feature follows a cycloidal trajectory. This has been shown to reduce scanning time and delivered dose, while maintaining a high resolution. The different ways of implementing this method are: step-and-shoot, continuous unidirectional and continuous back-and-forth translation. While step-and-shoot acquisitions yields the best results and are easiest to implement, they are also the most time-consuming. The continuous unidirectional method can be implemented with little effort and gives results comparable to step-and-shoot. Finally, back-and-forth scans can be implemented easily and provide similar results, although there appears to be a small loss in spatial resolution. We present a comprehensive guide on using cycloidal sampling in practice.**

***Index Terms*— computed tomography, sampling, spatial resolution, structured illumination, x-rays**

I. Introduction

X-ray micro computed tomography (CT) is a robust imaging technique that produces high-resolution images (a few to tens of μm) of a sample. The images can then be combined to visualize a slice or 3D rendering of the sample with very fine detail. [1, 2]

In order to obtain high-resolution images specialized hardware is generally necessary, specifically a source with a sufficiently small focal spot and a detector with small pixels that suffer negligible crosstalk. These sources and detectors exist, but they are not the norm in several applications and impose a restricted field of view, respectively [3].

Higher spatial frequencies can be accessed when the x-ray beam is appropriately structured, even when the source focal spot and detector pixels are relatively large. We have developed an approach that consists of splitting the x-ray beam into an array of narrow beamlets (a few μm to tens of μm), the use of which allows achieving higher resolution, largely decoupled from the source and detector properties. However, in principle our approach requires a dense sampling scheme, by which the sample is scanned laterally in multiple steps at every rotation angle (a process called "dithering") in order to ensure that each part of the sample has been illuminated by the beamlets [4]. This is an inefficient procedure that introduces dead times in the acquisition process, lengthening scan times significantly. Importantly, it is incompatible with continuous acquisitions.

Cycloidal computed tomography overcomes this problem by under-sampling in an efficient manner and compensating the data incompleteness via a mathematical recovery method (explained in detail below).

We would also like to highlight that cycloidal computed tomography is a highly flexible concept, in the sense that it can be implemented in both attenuation and phase contrast modes (again see below). This is particularly powerful because different samples might benefit from using different contrast channels.

However, the cycloidal technique is relatively new and there has been little work on how to optimally implement it in practice. We have explored three different acquisition strategies for cycloidal computed tomography and compared them.

II. Background

A. Cycloidal computed tomography [5]

Cycloidal computed tomography is enabled by combining the beamlet-based setup described above with an innovative, efficient sampling scheme, and an appropriate data recovery method.

As mentioned above, before hitting the sample the x-ray beam is split into an array of beamlets with the help of an absorbing mask ("sample mask") with periodically spaced slit-shaped apertures (Fig. 1). There are two specific criteria for the mask: first of all, the aperture width (w) must be smaller than

This abstract was first submitted on the 17th of January 2020. Marco Endrizzi and Charlotte K. Hagen are supported by the Royal Academy of Engineering under the Research Fellowship scheme. Alessandro Olivo is supported by the Royal Academy of Engineering under the Chair in Emerging Technologies scheme.

Oriol Roche i Morgo, Marco Endrizzi, Alessandro Olivo and Charlotte K. Hagen are with the Department of Medical Physics and Biomedical Engineering at University College London, Gower St., WC1E 6BT, London, UK (contact: oriol.roche.15@ucl.ac.uk).

Fabio Vittoria is with ENEA – Radiation Protection Institute, Bologna, Italy.

the system blur in order to access spatial frequencies higher than those usually available with the same source and detector. Secondly, the mask period (p) must be large enough to prevent overlap between adjacent beamlets. Usually p matches the pixel size of the detector (a) when scaled to the isocenter of the sample. However, in order to ensure the second criterion, sometimes a "line-skipping" mask is employed, which has a larger period, so that some pixels are covered and only every second pixel (or third, or fourth…) is illuminated.

Under these conditions, the cycloidal sampling scheme can be employed. The sample is translated laterally through the setup simultaneously to being rotated around the vertical (y-) axis. At each rotation angle, only a single frame is acquired. The sample is then slightly shifted as it is rotated to the next angle, another frame is acquired, and so on. This sampling scheme differs from dithering (where the sample is scanned laterally in multiple steps at each angle, see above) because at every angle only a single frame, with the sample in a slightly different lateral position, is acquired. The distance d by which the sample is translated per rotation angle is a fraction of the mask period.

With this "roto-translation" motion, the acquired data correspond to a well-spread distribution on the sinogram sampling grid (Fig. 2c). The exact data distribution depends on the value of d, as well as the angular sampling step. Although the sinogram is incomplete (only the filled dots are available, all others are not acquired), this spread helps recover the missing entries, because the data points complement each other better than in the rotation-only case (Fig. 2a), where they are densely packed in the angular direction but far apart in the lateral direction. With an appropriate data recovery method (which can range from bicubic interpolation to machine learning approaches), cycloidal sampling can access spatial resolutions comparable to dithering (which leads to complete sinograms, Fig. 2b) even though fewer frames are acquired.

B. Edge-Illumination X-ray Phase Contrast

The edge-illumination (EI) x-ray phase contrast technique is a type of x-ray phase contrast imaging, developed originally with synchrotron radiation but subsequently extended for use in laboratory settings with conventional x-ray tubes [6]. Phase contrast imaging takes advantage of the phase effects introduced by a sample, in addition to the attenuation in conventional x-ray imaging. In the case of weakly attenuating materials, such as biological soft tissue, the exploitation of phase effects is known to increase image contrast. Therefore, phase contrast techniques, and the EI technique among them, present great potential in the field of biomedical imaging [7].

The beamlet-based scanner setup, which is the basis for performing cycloidal computed tomography, can easily be converted into an EI x-ray phase contrast imaging system. To do so, a second absorbing mask ("detector mask") is placed in front of the detector (see Fig. 1). This mask also has periodically spaced apertures which match, adjusted for magnification, the apertures of the first mask. The two masks are positioned with respect to each other so that, without a sample, approximately half of each beamlet falls on each side of an edge created by the detector mask apertures, giving a specific intensity signal.

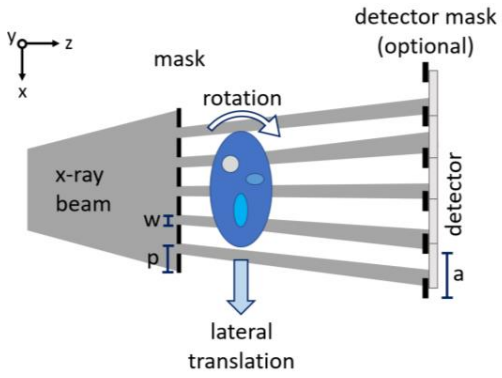


Figure 1. A schematic representation of our setup. The first mask splits the beam into beamlets. The second (optional) mask transforms the set-up into an edge-illumination system. Introducing the sample causes the beamlets to shift (due to refraction) which provides the phase contrast. By simultaneously rotating and translating the sample laterally, we can obtain cycloidally sampled datasets.

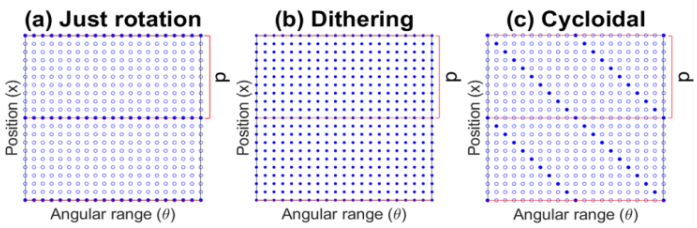


Figure 2. Sketch of the sinogram sampling grid for different sampling strategies, shown for two mask periods and a subset of rotation angles. In each panel, the filled dots represent data points that are acquired during a scan, while the empty dots represent the data that are not acquired. Panel (a) shows the grid if the sample just rotates. The data points are densely packed in one direction only. Panel (b) shows the grid when dithering is applied. The grid is fully sampled. Panel (c) shows the grid when applying a cycloidal scheme. The number of data points in (c) is the same as in (a), but they are more evenly distributed.

When a sample is introduced, refraction slightly shifts the beamlets according to the sample's composition and geometry, and the intensity on the detector plane changes, which gives rise to phase contrast [8, 9].

In order to accurately position the masks, usually an illumination curve (IC) is acquired by shifting the sample mask by several steps over one period and recording the intensity changes in the detector. The IC can be approximated as a Gaussian curve and used to determine the best position of the sample mask to maximize the phase signal.

When reconstructing CT images, a process of "phase retrieval" is required because the intensity in a single image is the combination of attenuation and refraction effects, and a specialized algorithm is necessary to isolate the phase signal. When the sample is homogeneous, the images are acquired on one of the slopes of the IC and the beam shift is sufficiently small, only one image per projection angle is required for the retrieval. This is referred to as a "single-shot" approach [10].

III. Practical Implementation Strategies

Cycloidal computed tomography can be implemented in a flexible manner. We have investigated three different options to do this: step-and-shoot, continuous unidirectional and continuous back-and-forth scanning.

The sample was a phantom composed of polyethylene

spheres between 425 and 500 μm in diameter, contained in a 3 mm diameter plastic straw. As this sample is weakly attenuating, all data were acquired using the scanner in x-ray phase contrast mode, by implementing the EI technique. The detector was a CMOS-based flat panel C9732DK-11 with 50 μm pixels from Hamamatsu (Japan). A "line-skipping" detector mask (with an aperture width of 17 μm and a period of 98 μm) was employed to reduce the adverse effects of pixel cross talk; this has resulted in an effective pixel size of 80 μm (at the sample plane). The sample mask had a 79 μm period and 10 μm aperture. The source was a MicroMax 007 HF x-ray tube with a Molybdenum target from Rigaku (Japan), operating at 40 kVp and 20 mA. The processing of all data included flat field and dark field corrections, and single-shot phase retrieval. The missing data points in the sinogram (see Fig. 2c) were recovered using bicubic interpolation.

A. *Step-and-shoot scanning*

In step-and-shoot mode, all frames are acquired by moving the sample to position, exposing it to x-rays and saving the detector's response. The sample is then moved to the next (lateral and angular) position, exposed again, and so on. In our case, we only translated the sample by up to one mask period, then shifted it back to its initial position, repeated the stepped, lateral movement by up to one mask period, and so on (see schematic in Fig. 3a). Due to the periodicity of the pixel array, this is equivalent to translating the sample along one direction.

In a step-and-shoot scan, the distance d by which the sample is translated per angular step does not vary throughout the scan, which opens the question as to whether there is an optimal value of d. We have run tests for different values, and $d = 0.25p$ has been found to lead to the best results in terms of spatial resolution [11].

900 frames were acquired over 180°, with an angular step of 0.2° and an exposure time of 1.25 s per frame.

B. *Continuous unidirectional scanning*

In continuous unidirectional mode, the sample starts rotating and moving laterally at the same time and does not stop until all projections have been acquired. The lateral movement is performed in just one direction, so the sample starts at one side of the field of view and moves towards the other side (see schematic in Fig. 3c). Because of this, the sinogram is "slanted" and the data points need to be re-positioned into what would be their pixel position if a step-and-shoot scan were acquired, in a process that we call regridding.

The continuous unidirectional data were acquired with $d = 0.25p$. 900 frames were acquired over 180°, with an angular step of 0.2° and an exposure time of 1.25 s per frame. This corresponds to a total translation of 17.775 mm. A few extra projections were acquired before and after and then discarded, to account for the acceleration and deceleration of the translation and rotation stages.

C. *Continuous back-and-forth scanning*

This acquisition strategy consists of moving the sample laterally between two points within the field of view, back and forth until all projections are acquired. In a similar way as for the continuous case above, the sinogram is "slanted" (Fig. 3e), so it also requires regridding.

The continuous back-and-forth data were acquired with $d = 0.25p$, an angular step of 0.2° and an exposure time of 1.25 s per frame. The two endpoints of the back-and-forth motion were selected so that 500 projections could be acquired during a single continuous motion (which corresponds to a distance of 9.875 mm). Again, a few extra projections at the beginning and end of the scan were included to account for motor acceleration and deceleration, which were omitted in the reconstruction. The motor acceleration and deceleration around the point where the sample changes direction were considered negligible.

D. *Dithered scanning*

For comparison, we have also acquired a dithered, i.e. complete dataset. The dithered data were acquired with 8 dithering steps, each with 1.25 s of exposure time. A total of 7200 frames were acquired over 180°, with an angular step of 0.2°. The sinogram data were already complete (see Fig. 2b), so no bicubic interpolation or regridding was used in this case. As for the cycloidal cases, single-shot retrieval was applied.

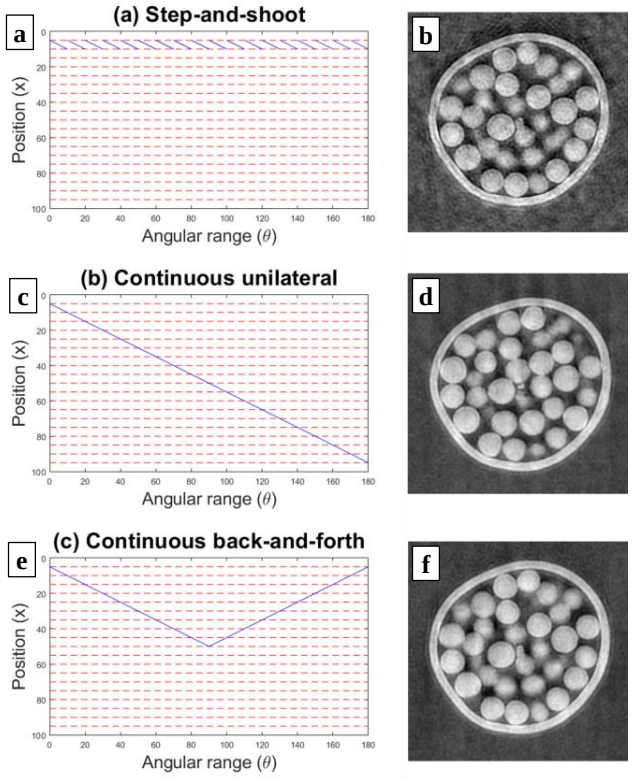


Figure 3. For the three cycloidal computed tomography implementation strategies, (a), (c) and (e) schematically show the trajectory of the sample during scans. Panels (b), (d) and (f) show the reconstructed images of the sample obtained for each implementation strategy.

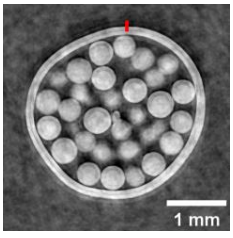


Figure 4. Reconstructed tomographic image from a fully dithered dataset with 8 dithering steps. The red line indicates where the spatial resolution was measured for the dithered and cycloidal images.

IV. Results and Discussion

The three strategies presented above are all realistic options for implementing cycloidal computed tomography. The results obtained through step-and-shoot and continuous acquisitions are qualitatively similar to each other, as well as to the dithered images. The dithered scan and the step-and-shoot cycloidal scan have the best spatial resolution (20 μm and 18 μm respectively). The spatial resolution of the continuous unidirectional image is 24 μm, significantly better than for the continuous back-and-forth image (30 μm). This shows the potential of the method: despite acquiring a fraction of the frames compared to dithering (eight times fewer frames), we obtain comparable images, especially in terms of spatial resolution.

Each of the presented strategies exhibits advantages and disadvantages. Data acquired in a step-and-shoot fashion do not require any regridding of the sinogram before reconstruction, like in the continuous cases. This is an advantage because the regridding process is error prone (as the speed of the translation motor has to be known precisely to achieve an accurate result). On the downside, step-and-shoot acquisitions require longer scan times because the motors are stopped and started again at every step, introducing dead times.

For the continuous unidirectional scan, the continuous movement does not seem to affect the image. The regridding is relatively straight-forward as long as the translation speed of the motor is accurately known. The key advantage of continuous scans is that they can be fast (no dead times are required for stop-starting the motors), which makes unidirectional continuous cycloidal CT a powerful alternative to step-and-shoot scanning. With our current setup, a step-and-shoot scan can typically take around 2.5 hours, where a continuous scan only takes 25 minutes.

A drawback of the unidirectional method lies in its field of view restrictions. The field of view needs to be wide enough for the sample to complete a 180º rotation while it moves from one end to the other and include all lateral positions in that movement. For small samples (a few millimeters) this might be feasible, depending on the detector and size of the masks (typically around 6 cm although larger ones have been built), but for larger samples it might be difficult to find a setup with a sufficiently large field of view.

The back-and-forth strategy offers a possible solution to this: for cases with a limited field of view, once the sample reaches the end it can turn around and continue in the other direction; this is then repeated until all frames are acquired. However, as mentioned above the back-and-forth case gives worse spatial resolution. This might be due to additional complications in the data processing. In the back-and-forth movement, when the sample changes direction, the value of d is not constant, due to the acceleration and deceleration of the translation motor, making the regridding process more difficult. In our example, there was only one change in direction, in which we have assumed the acceleration and deceleration to be negligible (and assumed that the position of the sample strictly follows from the d values). We suspect, though, that the effect of this erroneous assumption will increase with the number of times motion is reversed. A more accurate hardware may exist that allows for the exact sample position to be known with much higher accuracy, which would allow for the data to be regridded more accurately, thus improving the result.

We would like to highlight that, in the extreme case, the back-and-forth method could be implemented by only moving the sample back-and-forth within one (de-magnified) pixel. In that case, there would be no need for regridding and the field of view requirements would be even more relaxed, thus increasing the number of samples that could take advantage of the cycloidal method. This strategy requires quick acceleration and deceleration of the motors and very accurate knowledge of the sample position, so it will be the focus of future work.

V. Conclusion

In summary, we have presented three different practical implementation strategies for cycloidal computed tomography.

Both step-and-shoot and continuous cycloidal scans are relatively straight-forward to implement and provide images of comparable quality to fully dithered scans, despite the acquisition of substantially fewer projections. A key strength of the continuous approaches (unidirectional and back-and-forth scanning) is that these are much faster in terms of overall acquisition duration as they eliminate all dead times. Not only do these strategies lead to short scan times, they also have relaxed field of view requirements and provide good image quality. For these reasons, we envision the continuous acquisition strategies to become the methods of choice for applications of cycloidal computed tomography. More work to improve the image quality delivered by the presented continuous back-and-forth acquisition strategy will be performed, and variations of these strategies (such as moving the sample within one pixel) will be investigated.

References

[1] A. Kak and M. Slaney, *Principles of computerised tomographic imaging*, IEEE Press, New York, NY, USA, 1988.

[2] E. Baird and G. Taylor, "X-ray micro computed-tomography," *Current Biology*, vol. 27, no. 8, pp. R289-R291, 2017.

[3] J. Wan and D. Fleischmann, "Improving Spatial Resolution at CT: Development, Benefits, and Pitfalls," *Radiology*, vol. 289, no. 1, pp. 261-262, 2018.

[4] P. C. Diemoz, F. A. Vittoria and A. Olivo, "Spatial resolution of edge illumination X-ray phase-contrast imaging", *Optics Express*, vol 22, no. 13, pp. 15514-15529, 2014.

[5] Cycloidal Computed Tomography, by C. K. Hagen, F. Vittoria, M. Endrizzi and A. Olivo (2018, Dec 13). UK Patent Application 1820362.0

[6] A. Olivo and E. Castelli, "X-ray phase contrast imaging: from synchrotrons to conventional sources", *La rivista del nuovo cimento*, vol 37, no. 9, 2014.

[7] A. Bravin, P. Coan and P. Suortti, "X-ray phase-contrast imaging: from pre-clinical applications towards clinics", *Physics in Medicine & Biology*, vol 58, no. 1, 2012.

[8] A. Olivo et al, "An innovative digital imaging set-up allowing a low-dose approach to phase contrast applications in the medical field", *Medical Physics*, vol 28, no. 8, 2001.

[9] A. Olivo and R. D. Speller, "A coded-aperture technique allowing x-ray phase contrast imaging with low-brilliance x-ray sources", *Appl. Phys. Lett.* vol 91, no. 7, 2007.

[10] P. Diemoz et al, "Single-shot x-ray phase-contrast computed tomography with non-microfocal laboratory sources", *Phys. Rev. Appl.* vol 7, no. 4, 2017

[11] O. Roche, manuscript in preparation.

Fast Task-Driven Acquisition Optimization for Anisotropic X-ray Dark-field Tomography

Theodor Cheslerean-Boghiu*, Franz Pfeiffer†, Tobias Lasser*

***Abstract*—Anisotropic X-ray Dark-field Tomography aims at reconstructing spherical scattering functions from directional dark-field contrast measurements acquired using a Talbot-Lau X-ray grating interferometer. Compared to conventional X-ray Computed Tomography, this approach requires extensive and elaborate acquisition protocols due to the anisotropic nature of the dark-field signal and the unidirectional sensitivity of the interferometer setup. In recent years, new algorithms have emerged that create task-driven optimized acquisition trajectories, in an attempt to discard measurements which are "less relevant" to the reliable reconstruction of a user-defined task and to only keep the "relevant" ones, thus lowering both scan time and dose. Such algorithms, however, are computationally quite expensive. In this work, we introduce a new simplified method of computing such a task-driven acquisition geometry, and we empirically show its performance against state-of-the-art procedures.**

I. Introduction

In Anisotropic X-ray Dark-field Tomography (AXDT) the recovery of sub-voxel microstructure orientations is tied to the scattering profile of the structure, which is approximated by reconstructing a field of spherical functions from a set of dark-field (DF) measurements [8]. The X-ray DF contrast is measured using a Talbot-Lau grating interferometer with three gratings, like the one in Fig. 1.

The process of reconstructing the local scattering profile poses significant hurdles which must be dealt with before the actual reconstruction. The DF signal is produced by the ultra-small angle scattering of X-ray photons from structures smaller than the size of the voxels [6]. Moreover, the grating interferometer is limited by the unidirectional sensitivity, which has its highest DF measurement sensitivity when the DF signals originates from structures parallel to the grating orientation. The obtained DF signal is thus anisotropic in nature and encodes information about the three-dimensional scattering profile for each individual volume element inside the measured sample.

*T. Cheslerean-Boghiu and T. Lasser are with the Department of Informatics, Technical University of Munich, Germany.

†F. Pfeiffer is with the Chair of Biomedical Physics, Department of Physics and Munich School of BioEngineering, Technical University of Munich, Germany, and with the Department of Diagnostic and Interventional Radiology, School of Medicine and Klinikum rechts der Isar, Technical University of Munich, Germany.

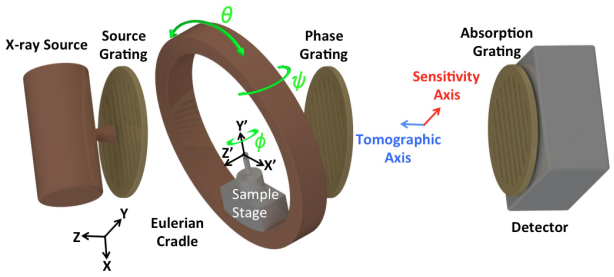


Fig. 1: Schematic of a Talbot-Lau X-ray grating interferometer setup using an Eulerian cradle to rotate the sample around its three axes, y (ψ), z' (θ), and y' (ϕ) [6].

Unfortunately, this leads to complex and time-consuming acquisition geometries for AXDT, since the DF signal has to be measured from all possible angles, for example using the generalized approach introduced in [6].

State-of-the-art

In a wide variety of industrial tomography applications prior knowledge of the imaged sample is present. Moreover, in most cases, the reconstruction process can be reduced to the reconstruction of a specific region-of-interest (ROI): e.g. inspecting a welding seam in an industrial sample [3]. It has been shown that prior information about the structure of the imaged sample and about the user-defined reconstruction task can be used to design better acquisition trajectories that increase the quality of the reconstruction of the ROI. Currently, acquisition trajectories for AXDT are created based on general rules, like t-designs, in order to efficiently measure all the possible scattering orientations spread over the unit sphere [6]. While this ensures a sufficiently sampled measurement of each scattering profile in each voxel, it does not take into account prior knowledge about the sample and the expected signal in some ROI. Moreover, an acquisition protocol fully covering the unit sphere has one major drawback: scan time. Currently, there is a lot of research going on to develop task-driven acquisition optimization for conventional X-ray computed tomography (CT) [7, 4] which show great promise for industrial applications [3].

Furthermore, these algorithms have also been previously derived for AXDT [2] and the methods showed a great reduction in the acquisition geometry size for the reconstruction of a user-defined task. However, there is still the need for faster methods that yield task-based optimized trajectories since the current state-of-the-art algorithms [3] are based on a greedy search approach that are very time-consuming. In this paper, we propose a new method for faster task-driven trajectory optimization, based on a simple sorting algorithm, which performs as good as the greedy approach from [3] but only computes the detectability index (DI) once for each pose.

II. Methods

In order to discuss the method for finding optimal task-based trajectories, we first need to lay out the mathematical model for AXDT.

System model

The goal of AXDT is to obtain the intra-voxel scattering orientation by reconstructing a field of spherical functions $\eta : \mathbb{S}^2 \times \mathbb{R}^3 \to \mathbb{R}$ from a set of DF measurements d obtained using the X-ray grating setup described in Section I. To model the relationship between each spherical function and DF contrast measurement, we use a set of weighting functions $h : \mathbb{S}^2 \times \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$ which encode additional information about the grating sensitivity direction, $s(x)$, the X-ray direction $t(x)$, and the ray x. We then simplify the problem of directly reconstructing a spherical function to a problem of estimating the spherical harmonics coefficients that represent the desired spherical function.

We discretize the spherical functions η and h using the basis of real-valued spherical harmonics truncated to a maximum degree of 4, and we use the Beer-Lambert law to formulate the forward model [8] as:

$$d(x) \approx \exp\left(-\frac{1}{4\pi}\sum_{k=0}^{4}\sum_{m=-k}^{k} h_k^m\big(s(x),t(x)\big)\int_{T(x)}\eta_k^m(r)dr\right), \tag{1}$$

where x is the corresponding X-ray, k is the degree and m the order of the spherical harmonics. The line integration is discretized over the X-ray path $T(x)$ yielding a matrix P, and the weighting functions $h_k^m\big(s(x),t(x)\big)$ are all concatenated into a diagonal matrix W_k^m, leading to the following matrix-based notation:

$$\begin{aligned} d &= \exp(-\sum_{k=0}^{K}\sum_{m=-k}^{m=k} W_k^m P\eta_k^m) \\ &= \exp(-\boldsymbol{\mathcal{B}}\eta), \end{aligned} \tag{2}$$

In the following equations we will use $\hat{d}$ to describe the actual dark-field measurements and $\epsilon(\eta) = \exp(-\boldsymbol{\mathcal{B}}\eta)$ to denote the forward projection of an arbitrary volume η. Using the previous notation, we obtain the energy function

$$L(\eta, d) := \left\| \exp(-\boldsymbol{\mathcal{B}}\eta) - d\right\|_2^2 + \beta R(\eta). \tag{3}$$

where $\beta > 0$, and $R(\eta)$ represents a quadratic roughness penalty, $R(\eta) = \sum_j \frac{1}{2}\sum_k w_{jk}\phi(\eta_j - \eta_k)$. Finally, we define the reconstruction of $\hat{\eta}$ as the minimizer of the likelihood function,

$$\hat{\eta} = \arg\min_{\eta} L(\eta, \hat{d}). \tag{4}$$

Detectability Index

The DI is a metric that encodes information about the visibility of a certain feature for a specific viewing angle. For CT this translates to differentiating between an occluded view and an occlusion-free view of the ROI. Fischer et al. [3] demonstrated DI maps given specific angle intervals, where they investigated the reconstruction of a screw in an industrial highly attenuating sample. For AXDT, the visibility of a specific location inside the sample is controlled by two factors: attenuation by a highly absorbing section of the imaged sample, and the relative position of the grating bars and the DF wavefront. This will be reflected in the DI values.

We compute the DI for a fixed geometry using a Non-Prewhitening Human Model Observer as in Gang et al. [4]:

$$\delta_j^2 = \frac{\left[\int_{\mathbb{R}^4}(MTF_j\mathcal{W}_{ROI})^2\ df^4\right]^2}{\int_{\mathbb{R}^4} NPS_j(MTF_j\mathcal{W}_{ROI})^2\ df^4} \tag{5}$$

For the computation of the Modulation Transfer Function MTF and the Noise Power Spectrum NPS we use a formulation introduced in an earlier paper [1]. The two metrics MTF and NPS are approximated by the Fourier transform of the linear impulse response and the local covariance [7]:

$$MTF_j(\eta) \approx \mathcal{F}\{\lambda_j(\eta)\} = \frac{\mathcal{F}\{-2\boldsymbol{\mathcal{B}}^T\mathrm{diag}(\varepsilon(\boldsymbol{\eta}))^2\boldsymbol{\mathcal{B}}e_i^*\}}{\mathcal{F}\{-2\boldsymbol{\mathcal{B}}^T\mathrm{diag}(\varepsilon(\boldsymbol{\eta}))\big(2\varepsilon(\boldsymbol{\eta}) - \boldsymbol{d}\big)\boldsymbol{\mathcal{B}}e_i^*\}}, \tag{6}$$

$$NPS_j(\eta) \approx \mathcal{F}\{\sigma(\eta_j)\} = \frac{\mathcal{F}\{4\boldsymbol{\mathcal{B}}^T\mathrm{diag}\big(\varepsilon(\boldsymbol{\eta})\boldsymbol{d}\varepsilon(\boldsymbol{\eta})\big)\boldsymbol{\mathcal{B}}e_i^*\}}{\left|\mathcal{F}\{-2\boldsymbol{\mathcal{B}}^T\mathrm{diag}(\varepsilon(\boldsymbol{\eta}))\big(2\varepsilon(\boldsymbol{\eta}) - \boldsymbol{d}\big)\boldsymbol{\mathcal{B}}e_i^*\}\right|^2} \tag{7}$$

where the division is element-by-element. e_i^* is a special impulse which represents the specific scattering profile we are interested in creating an acquisition protocol for.

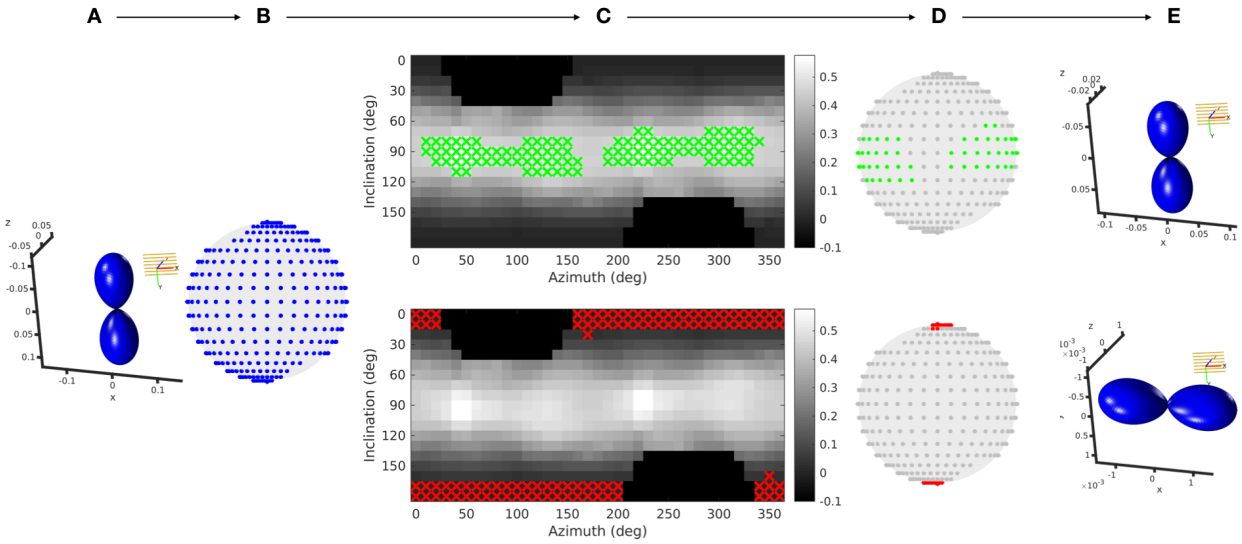


Fig. 2: (A) reference scattering profile reconstructed with the full geometry; (B) full geometry (please notice the two locations in the vicinity of the poles which our physical and mathematical models cannot cover; we will ignore those locations completely); (C top) chosen poses in the order of the highest DI using SBA; (C bottom) chosen poses in the order of the lowest DI using SBA; (D top and bottom) corresponding poses highlighted on the sphere; (E top and bottom) reconstructed scattering profiles using only the poses from the corresponding geometry.

In Fig. 2-C (top and bottom) we computed DI maps of a single scattering profile (see Fig. 2-A) for a general AXDT acquisition scheme with acquisition poses all around the unit sphere.

Given the physical properties of the Talbot-Lau interferometer to measure the DF signal, scattering of the X-ray photons happening parallel to the grating bars will produce little to no measurable DF contrast, while scattering produced perpendicular to the grating bars will generate the highest DF contrast. The corresponding DI maps are shown in Fig. 2-C (top). It is clear that as we measure the given impulse from a shallow angle ($\approx 90°$ inclination) we obtain the highest values for the detectability (highest visibility is at $0°$ inclination angle). Measuring towards more extreme inclination angles $0°$ and $180°$ the loss of dark-field contrast is indeed reflected in the values of the DI which start to shrink, getting close to 0 at $0°$ respectively $180°$.

The computation of the NPWM at the two contiguous black areas on the maps was discarded since our acquisition setup is unable to achieve those angles thanks to the Euler cradle, and the implemented physical model reflects that by ignoring those locations.

Task-driven Trajectory Optimization

The DI can now be used as a metric to differentiate between "valuable" or "less valuable" poses, given a user-specified reconstruction task. Currently, task-based acquisition trajectories are generated using the detectability index as a metric for an iterative greedy algorithm (IGA) [3]. We propose a faster method, called the sorting-based algorithm (SBA) shown in Algorithm 1, where the overall runtime is comparable to the runtime of one iteration of IGA. The main approach of SBA is to only compute the DI values once for every pose, basically running the first iteration of IGA, and then sort the poses from highest to lowest DI and choose the first (best) n poses which are then used to create an optimized geometry.

Algorithm 1 Sorting-based path optimization algorithm

1: Let $\mathcal{P}_{all}$ be the set of all possible poses
2: Let n be the number of required poses
3: Compute $D = \{\delta^2(p, \mathcal{W}_{Task}) \,|\, p \in P_{all}\}$
4: Sort $D \rightarrow D_{sorted}$
5: $P_{selected} = D_{sorted}[1:n]$
6: Use $P_{selected}$ to reconstruct volume

III. Simulations and Discussion

We seek to confirm that the DI can be used to select "valuable" poses on simulated data and to empirically confirm that SBA is a strong candidate for task-driven geometry optimization. Our dataset comprises of an almost empty volume with a scattering function parallel to the y-axis in the center. For our volume, we define a geometry that fully samples the unit sphere, composed of 589 angles.

For this geometry we apply the algorithm SBA from 1 to obtain a geometry composed of only "valuable" poses, chosen by their highest corresponding DI values, and one composed of "least valuable" ones, chosen based on the lowest corresponding DI values, in order to have a "bad" reconstruction to compare, see Fig. 2.

We implemented SBA in our open-source reconstruction framework *elsa* [5]. In our simulations, we empirically found out that setting $\beta = 10^3$ we obtain smooth enough DI maps. The analysis of the regularization term and parameter will be left for future investigation. Given the anisotropic nature of the dark-field signal and thus the scattering function, the two reconstructions with a "good" geometry (green) and a "bad" geometry (red) are completely different (notice the scale of the two scattering functions). In Fig. 2-E (bottom) it is quite clear that the reconstructed scattering profile is actually very small compared to the reference reconstruction in 2-A and can be considered residual. To qualitatively analyze the performance of the proposed algorithm we generate a geometry for an increasing number of angles, performed the reconstructions matching to the geometries, and plotted the absolute difference between the reference reconstruction and the reconstruction done with the corresponding sized geometry in Fig. 3. We observe a decreasing trend for both the proposed algorithm (SBA) and the existing one (IGA), with the former one performing slightly better. Furthermore, we show the results for the reconstruction done with the "bad" geometry in red. We can clearly see that the quality of the reconstruction using the proposed SBA performs as good as IGA.

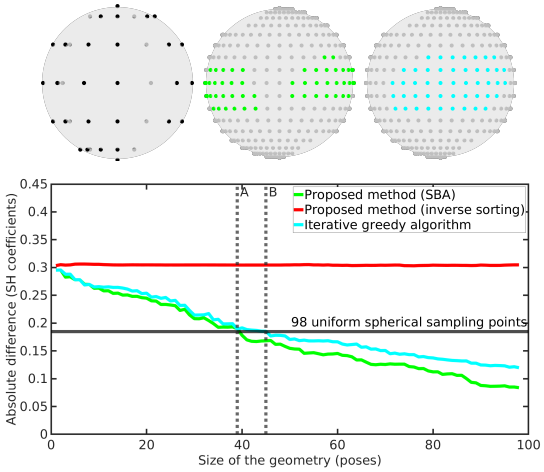


Fig. 3: (top left) Acquisition poses with uniform sampling with 98 angles; (top center) Geometry generated with SBA; (top right) Geometry generated with IGA; (bottom) Absolute difference in the ROI between the tomographic reconstruction using the full geometry (589 poses) and the reconstruction from a subset of chosen poses by the following criteria: (green) first n poses with the highest respective DI (in green in top-center), (red) first n poses with the lowest respective DI (in red in Figure 2); the two vertical dotted lines: one at $x = 39$ (A) and the other at $x = 45$ (B). These lines represent the number of angles in the geometry obtained using SBA high-low (green) and IGA (cyan) where the quality of the reconstruction surpasses the one done using a uniformly sampled geometry of 98 angles.)

Discussion

We propose a new fast method of computing task-based optimized acquisition trajectories (SBA). Our simulation results show that it performs just as well as the computationally intensive greedy approach (IGA) for a well-defined scattering function task. Both approaches enable significantly better reconstruction results than uniform sampling, and allow approaching the image quality of a reference reconstruction with only $\sim 17\%$ of the acquisition poses.

In future work we intend to combine the both algorithms (SBA and IGA) in a well-chosen trade-off between accuracy and computation speed, coupled with newly developed observer models for AXDT taking into account the spherical function nature of the scattering profiles.

References

[1] Theodor Cheslerean Boghiu, Franz Pfeiffer, and Tobias Lasser. "Task-driven acquisition in anisotropic x-ray dark-field tomography". In: *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. Ed. by Samuel Matej and Scott D. Metzler. Vol. 11072. International Society for Optics and Photonics. SPIE, 2019, pp. 205–209. DOI: 10.1117/12.2534831. URL: https://doi.org/10.1117/12.2534831.

[2] Theodor Cheslerean Boghiu et al. "Detectability Indices in Anisotropic X-ray Dark-Field Tomography". In: *Proc. CT Meeting*. May 2018.

[3] Andreas Fischer et al. "Object Specific Trajectory Optimization for Industrial X-ray Computed Tomography". In: *Scientific Reports* 6.4 (Jan. 2016). DOI: 10.1038/srep19135.

[4] Grace J. Gang et al. "Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance". In: *Medical Physics* 38.4 (2011), pp. 1754–1768. ISSN: 2473-4209. DOI: 10.1118/1.3560428.

[5] Tobias Lasser, Maximilian Hornung, and David Frank. "elsa - an elegant framework for tomographic reconstruction". In: *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. Ed. by Samuel Matej and Scott D. Metzler. Vol. 11072. International Society for Optics and Photonics. SPIE, 2019, pp. 570–573. DOI: 10.1117/12.2534833. URL: https://doi.org/10.1117/12.2534833.

[6] Y. Sharma et al. "Design of Acquisition Schemes and Setup Geometry for Anisotropic X-ray Dark-Field Tomography (AXDT)". In: *Scientific Reports* 7.1 (2017), p. 3195. ISSN: 2045-2322. DOI: 10.1038/s41598-017-03329-0.

[7] J. W. Stayman and J. A. Fessler. "Efficient calculation of resolution and covariance for penalized-likelihood reconstruction in fully 3-D SPECT". In: *IEEE Transactions on Medical Imaging* 23.12 (Dec. 2004), pp. 1543–1556. ISSN: 0278-0062. DOI: 10.1109/TMI.2004.837790.

[8] M. Wieczorek et al. "Anisotropic X-Ray Dark-Field Tomography: A Continuous Model and its Discretization". In: *Phys. Rev. Lett.* 117 (15 Oct. 2016), p. 158101. DOI: 10.1103/PhysRevLett.117.158101.

Quantification of Lipid-shelled Xenon Microbubbles with Dual-Energy CT

Nadav Shapira*[1], Rajarshi Chattaraj*[1,2], Leening P Liu[1], Michael Geagan[1], Daeyeon Lee[2], Chandra Sehgal[1], Peter B Noël[1]
[1]*Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA*
[2]*Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, USA*

***Abstract* - Xenon is an inert gas that is used for diagnostics and therapy. Recent technological advancements enable the production of lipid-stabilized microbubbles that contain xenon which are capable of both producing contrast under insonation and delivering xenon gas to precise locations via activation by clinical ultrasound. Dual-energy CT (DECT) is an advanced imaging method that uses x-ray attenuation measurements at two different energy regimes to provide access to elemental composition information of various materials. Here we use DECT to image and quantify different lipid-shelled xenon microbubbles densities located inside a dedicated 3D-printed phantom. By taking advantage of the similarity in atomic numbers between xenon (z=54) and iodine (z=53), we use commercially available tools to quantify xenon concentrations within different microbubble densities. Our measurements obtained xenon concentrations of 0.54 ± 0.09 mg/ml, 0.30 ± 0.05 mg/ml and 0.20 ± 0.06 mg/ml that correlated linearly with the expected concentration ratios. This demonstrated ability may serve as a basis for future clinical applications of xenon. For example, one could utilize activation capabilities of ultrasound in combination with the contrast and quantification abilities of DECT.**

***Index Terms*— Ultrasound, Xenon, Microbubbles, Computed tomography, Dual-energy CT, Spectral-CT**

I. Introduction

XENON (Xe), the fifth noble gas, has been of growing interest in recent years due to its antiapoptotic effects, especially in treating hypoxic-ischemic injury in the brain [1]–[3]. While xenon therapy through prolonged inhalation is in clinical trials, we have made progress in developing lipid-stabilized, xenon-containing microbubbles (MBs) which, unlike inhalation, can deliver the gas at precise locations via activation by clinical ultrasound [4]. A distinctive feature and advantage of Xe microbubbles is their ability to produce specific non-linear contrast under insonation due to bubble oscillations. At the same time, Xe itself has been known to produce contrast for thoracic imaging in conventional radiography and CT [5].

Spectral Computed Tomography (Spectral-CT) utilizes measurements of the energy-dependent material-specific x-ray attenuation at two different energy regimes during a single CT scan to provide access to elemental composition information [6]. The advantages of Spectral-CT over conventional CT include improved differentiation between various tissues [7]–[9], reduced metal artifacts [10], and quantification of certain endogenous and exogenous materials such as calcium [11] and iodine [9]. Dual-Energy Computed Tomography (DECT), the first realization of Spectral-CT, has been investigated since the 1970's [6] and became a commercially available clinical tool more than a decade ago. Current commercial approaches to DECT include the rapid kVp-switching method [12], the spectral detector method [13], the split-beam method [14], the spin-spin method [6] and the dual-source (DSCT) method [15], which we utilized in our study. In DSCT spectral separation is achieved using two orthogonal x-ray tubes opposing two corresponding detector systems with each tube at a different tube potential (kVp).

DECT is known for both its iodine visualization and quantification capabilities [16], [17]. Compared to conventional CT, DECT technology has a much higher sensitivity for iodine and is able to provide iodine concentration estimations by applying material decomposition methods [9]. In fact, since iodine quantification is considered a key capability of clinical DECT technology [7], the design of many of the commercial systems is optimized towards iodine quantification. In addition, using sufficiently accurate iodine estimations, the contribution of iodine to the total attenuation can be suppressed and Virtual-Non-Contrast (VNC) images, which mimic non-contrast images, can be calculated [18].

Due to the proximity in atomic numbers of iodine (I, z=53) and xenon (Xe, z=54), the mass attenuation coefficients (μ/ρ_m) of the two elements are similar [19]. This implies that the ratio between the cross-section for Compton scattering and the cross-section for the photoelectric effect are similar between the two elements. Consequently, the ratio between the attenuation at

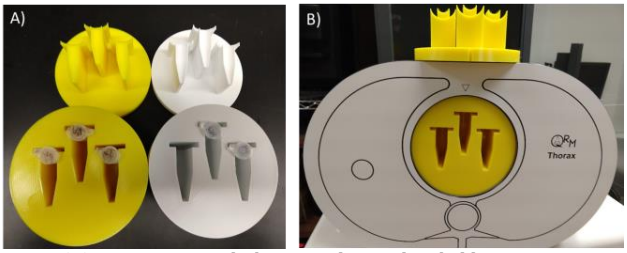


Fig. 1. (A) Two 3D-printed phantoms designed to hold up to six 1.5 mL micro-centrifuge tubes within the bore hole of a QRM Thorax phantom (B). The 3D-printed phantom design ensures minimal scatter artifacts that may result from sharp transitions in attenuation along the patient axis (z).

* Equal contribution to this work.

Corresponding author: Nadav Shapira, Department of Radiology, University of Pennsylvania, 3400 Spruce St., One Silverstein, Philadelphia, PA 19104, USA (e-mail: Nadav.Shapira@Pennmedicine.upenn.edu).

high x-ray energies and the attenuation at low x-ray energies is also similar for the two materials. This implies that, theoretically, DECT scanners are already equipped with the necessary imaging and visualization capabilities to identify and quantify xenon concentrations in a clinical scan. Indeed, DECT has been used in the past to quantify Xe gas down to concentrations of 5 mg/ml [20]. The purpose of this study is to utilize DECT for the quantification of Xe at even lower concentrations and that is contained within US-sensitive MBs.

II. Methods

A. Phantoms

In order to enable positioning of up to three 1.5 mL micro-centrifuge tubes vertically inside a semi-anthropomorphic phantom during a CT acquisition, a dedicated 3D-printed phantom was designed and two units were produced (Fusion3 F400, Greensboro, NC; 0.4 mm nozzle and 0.2 mm layer height). The phantom material Polylactic Acid (PLA) and the infill ratio were selected to achieve specific gravity similar to that of water. The design of the phantom ensures minimal scatter artifacts that may result from sharp transitions in attenuation along the patient axis (Fig. 1A).

The 3D-printed phantom was scanned inside a semi-anthropomorphic 300 mm x 200 mm thorax phantom (QRM-Thorax, QRM GmbH, Moehrendorf, Germany). The thorax phantom (Fig. 1B) includes artificial lung lopes, a spine insert, and a shell of soft tissue equivalent material, all composed of tissue mimicking plastics with respect to the x-ray attenuation characteristics at the clinical energy range (20-150 keV).

B. Xenon microbubbles production

To formulate Xe-containing microbubbles, a commercially available phospholipid (DBPC) and lipopolymer (DSPE-PEG5000) (Avanti Polar Lipids) were mixed in a 9:1 molar ratio (at 10 mg/mL) in phosphate buffered saline (PBS) at 90°C, followed by cooling to room temperature. Xenon (Praxair) gas was bubbled into the lipid-PEG solution. MBs were generated by probe-sonication (Branson), and size-isolated by differential centrifugation. MBs were imaged and counted using a hemocytometer to yield a size distribution largely between 2-8 µm and approximately $5x10^9$ MBs/ml for the most concentrated Xe MBs suspension ($\rho_{high}, 1x$) in our experiment (Fig. 2).

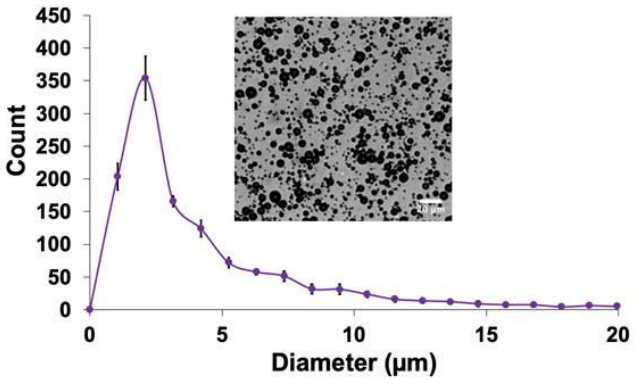


Fig. 2. Xenon MBs were imaged and counted using a hemocytometer to yield approximately $5x10^9$ MBs/ml for the most concentrated Xe MBs suspension ($\rho_{high}, 1x$) in our experiment and a distribution of MB diameters largely between 2-8 µm.

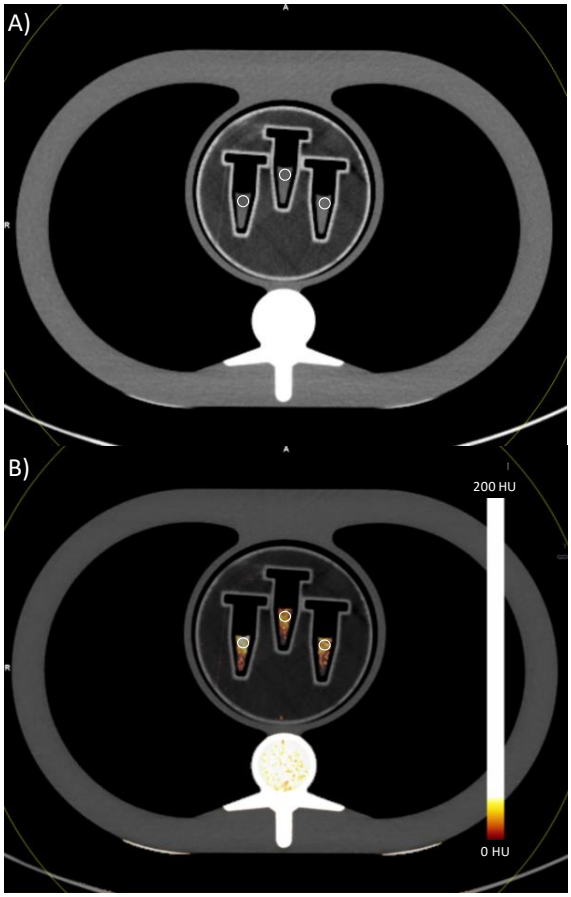


Fig. 3. Representative images of a Dual-Energy CT (DECT) acquisition of tubes filled with xenon microbubbles at different densities imaged within a dedicated 3D-printed phantom. (A) A Mixed-0.6 image composed of a linear blend of the high-kVp (150Sn) and low kVp (70) images. (B) Liver-VNC spectral application showing a Mixed-0.6 image overlaid with a Contrast Media (CM) result. ROIs were manually placed on the top-most region of observable CM values. (C/W = 40/400 for Mixed series, 25/25 for CM series)

For control bubbles with no expected CT contrast, a similar procedure was followed with perfluorobutane as the core gas instead of xenon. For negative control, a tube of the lipid-PEG stock was used without any bubbles.

C. Dual-Energy CT imaging and measurements

DECT imaging was performed with a third-generation dual-source scanner (SOMATOM Definition FORCE, Siemens Healthineers, Forchheim, Germany). A kVp pair of 70/150Sn was selected for maximal spectral sensitivity while dose values were maintained at their clinical settings (13.5 mGy) without applying dose modulation. Table I presents a complete list of scanning and reconstruction settings that were used for acquiring six repetitions of DECT datasets for the analysis. In addition to the Mixed-0.6 image (Fig. 3A), which is composed of a linear blend of the high and low kVp images, quantitative Contrast Media (CM) spectral results were generated using the Liver-VNC application provided by the dedicated spectral software (Syngo.via, VB 30; Siemens Healthineers, Forchheim, Germany). For each scan, five Regions-Of-Interest (ROIs),

each with an area of 300 mm^3, were manually placed on each of the tubes at the top-most region of observable CM values (Fig. 3B). Special attention was devoted to avoid the inclusion of voxels with partial volume of air or tube material within the ROIs.

D. Iodine Concentration Calibration

The Liver-VNC application provides, per each ROI measurement, the average and standard deviation (StdDev) values for Virtual-Non-Contrast (VNC), Contrast Material (CM) and a Mixed series (Fig. 3B). In order to obtain iodine concentration values that are (re)calibrated for the values given by the CM result of the specific scanner we used, a dedicated calibration scan was performed using the semi-anthropomorphic Thorax phantom with an insert containing 7 different concentrations of iodine (0.5 mg/ml; 0.75 mg/ml; 1 mg/ml; 2 mg/ml; 5 mg/ml; 10 mg/ml and 15 mg/ml CTIodine®; QRM, GmbH ,Forchheim Germany), together with a water-equivalent rod with an expected 0 mg/ml iodine concentration. The calibration scan was performed on the same scanner with the same acquisition parameters as in the scans for xenon quantification (Table 1).

III. Results

A. Iodine Calibration Formula

Five different slices from the iodine calibration scan were averaged to obtain a linear relation between the measured HU values of the CM result HU values and iodine density values. The resulting calibration curve equation, retrieved using a least squares linear fit with an R^2 value of 0.9998, is:

$$Iodine\left[\frac{mg}{ml}\right] = 0.037\left[\frac{\frac{mg}{ml}}{HU}\right] \times CM\,[HU] - 0.443\,\left[\frac{mg}{ml}\right] \quad \text{(Eq. 1)}$$

TABLE I
SCAN AND RECONSTRUCTION PARAMETERS

Parameter	Setting
Voltage pair (kVp)	70/150Sn
Pitch	0.5
$CTDI_{vol}$ (mGy)	13.5
Exposure (mAs)	580/145
Rotation time (s)	0.5
Collimation	19.2 (32 x 0.6)
Exposure Modulation	None
Recon. Kernel	Qr40
Slice Thickness (mm)	1
Slice Increment (mm)	1
Recon. Diameter (mm)	340
Matrix Size	512 x 512
Pixel Volume (mm^3)	0.66 x 0.66 x 1
Exposure Modulation	None
Reconstruction	FBP

Table 1. Scan and reconstruction parameters used for dual-source DECT acquisitions for xenon quantification.

B. Contrast Media Measurements

The measured CM average ROI values are presented in Fig. 4 where the dotted lines represent the average measurements between scans. The CM average values for the xenon microbubble tubes ranged between 34.0 ± 1.43 HU (31.6, 36.1) for the high concentration Xe MBs density ($\rho_{high}, 1x$) and 24.6 ± 0.9 HU (22.9, 25.4) for the low concentration Xe MBs density ($\rho_{low}, 1/3\,x$), while the CM average values for both the control bubbles and the negative (offset) control lipid-PEG were below 20 HU: 19.7 ± 1.3 HU (17.6, 21.5) and 18.8 ± 1.3 HU (16.9, 20.4), respectively. Applying the iodine calibration formula from Eq. (1) and subtracting the resulting average iodine concentration values obtained for the lipid-PEG to account for the substrate contribution to the iodine signal, resulted in iodine density values of 0.56 ± 0.09 mg/ml (0.41, 0.71) for Xe MBs ρ_{high} $(1\,x)$, 0.31 ± 0.05 mg/ml (0.23, 0.39) for Xe MBs ρ_{med} $(2/3\,x)$, and 0.21 ± 0.06 mg/ml (0.11, 0.29) for Xe MBs ρ_{low} $(1/3\,x)$.

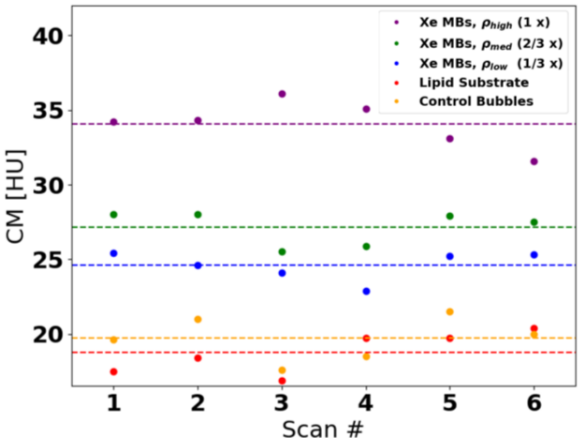


Fig. 4. CM measurement values from six scans. Average ROI values for each tube are represented by dotted lines. These HU values were transformed into iodine density values, in mg/ml, with the use of the iodine calibration formula (Eq. 1) for further analysis.

To determine the linearity of the acquired iodine densities, a second linear least squares linear fit was performed between the estimated Xenon MBs concentration ratios ($1/3\,x$, $2/3\,x$ and $1\,x$) and the retrieved iodine quantification values. The result of the linear fit is presented in Fig. 5.

IV. DISCUSSION

In this study we have used commercially available DECT technology to measure different xenon concentrations trapped within lipid-shelled microbubbles. For this purpose, we have utilized commercially available iodine visualization techniques and quantification algorithms while taking advantage of the similar attenuation properties of iodine and xenon that rises from their adjacent atomic numbers. To estimate difference between ideal iodine quantification and ideal xenon quantification using dual-energy information we examined the slopes and relative amplitudes of the expected attenuation values at low and high photon energies in the energy regime of medical CT. According to the NIST X-COM database [19], the

attenuation in CT^{num} (which is given by $[\mu/\rho_m(E)]/[\mu_{water}/\rho_m(E)] \times 1000 = HU + 1000$) of a single mg/ml iodine at 60 and 120 keV is $CT_I^{num}\,(60\,keV) = 36.81$ and $CT_I^{num}\,(120\,keV) = 7.50$ respectively while the attenuation of a single mg/ml xenon is $CT_{Xe}^{num}\,(60\,keV) = 38.01$ and $CT_{Xe}^{num}\,(120\,keV) = 7.76$. The very similar slopes in CT^{num} values at the low and high photon energies, $CT_{Xe}^{num}\,(60\,keV)/CT_{Xe}^{num}\,(120\,keV) = 4.89$ compared to $CT_I^{num}\,(60\,keV)/CT_I^{num}\,(120\,keV) = 4.90$, confirm that, at least theoretically, the response of a dual-energy CT system should indeed be similar for iodine and xenon signals.

Due to the close proximity in low-high attenuation slopes, the quantification correction factor for xenon for using DECT iodine concentration measurements can be calculated as the ratio between the vector magnitudes pointing from the origin, i.e. {0,0}, to $\{CT_i^{num}\,(120\,keV), CT_i^{num}\,(60\,keV)\}$ for the two materials (i.e. $i = I\ or\ Xe$). Based on this calculation, the correction factor is $\rho_I^{Xe} = 0.97$, meaning that every 1 mg/ml of measured iodine represents 0.97 mg/ml xenon. Using the calculated correction factor, we find measured Xe concentrations of 0.54 ± 0.09 mg/ml, 0.30 ± 0.05 mg/ml and 0.20 ± 0.06 mg/ml for the $1x$-tube, $2/3\,x$-tube and $1/3\,x$-tube, respectively.

Our study did have limitations. Firstly, we did not assess the effect of a reduced radiation dose and the associated increase in noise on the quantification of xenon. In addition, we did not asses the effect of different kVp-pairs on the quantification of xenon. Finally, production-estimated values exhibited large uncertainties that compelled us to correlate signal ratios of Spectral-CT xenon quantification accuracy at low concentrations rather than direct accuracy evaluations.

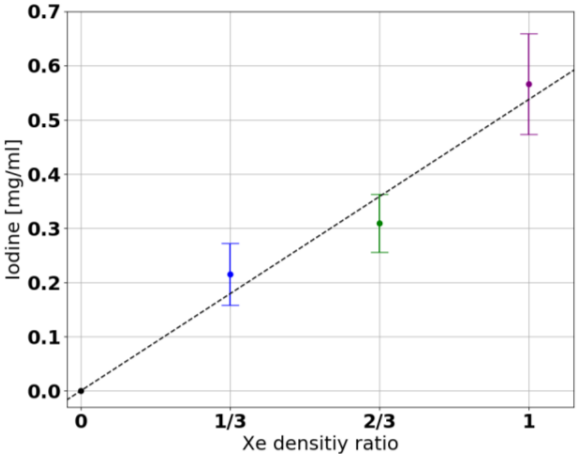


Fig. 5 Linear correlation between the quantified iodine concentrations and the production-estimated xenon MBs concentration ratios. Linear least square fit resulted with an R^2 value of 0.973.

V. Conclusion

In our experiment, we have successfully measured xenon concentrations in lipid-shelled microbubbles down to concentrations of approximately 2 mg/ml. This ability may serve as a basis for future clinical applications for xenon that take advantage of the contrast and quantification capabilities of DECT with the activation capabilities of ultrasound.

Acknowledgment

We acknowledge support through the University of Pennsylvania Research Foundation (URF).

References

[1] G. L. Britton *et al.*, "In vivo therapeutic gas delivery for neuroprotection with echogenic liposomes.," *Circulation*, vol. 122, no. 16, Oct. 2010.

[2] Y. F. Miao *et al.*, "Delivery of xenon-containing echogenic liposomes inhibits early brain injury following subarachnoid hemorrhage," *Sci. Rep.*, vol. 8, no. 1, Dec. 2018.

[3] E. Esencan, S. Yuksel, Y. B. Tosun, A. Robinot, I. Solaroglu, and J. H. Zhang, "XENON in medical area: emphasis on neuroprotection in hypoxia and anesthesia," *Med. Gas Res.*, vol. 3, no. 1, 2013.

[4] D. Chattaraj R., Hwang, M., Zemeroz, S., Dmochowski, I., Hammer, D.A., Sehgal, C.M., & Lee, "Ultrasound Responsive Noble Gas Microbubbles," *submitted*, 2019.

[5] W. D. Foley, V. M. Haughton, J. Schmidt, and C. R. Wilson, "Xenon contrast enhancement in computed body tomography," *Radiology*, vol. 129, no. 1, 1978.

[6] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography," *Phys. Med. Biol.*, vol. 21, no. 5, Sep. 1976.

[7] C. H. McCollough, S. Leng, L. Yu, and J. G. Fletcher, "Dual- and multi-energy CT: Principles, technical approaches, and clinical applications," *Radiology*, vol. 276, no. 3, Sep. 2015.

[8] D. C. Rotzinger, S. A. Si-Mohamed, N. Shapira, P. C. Douek, R. A. Meuli, and L. Boussel, "'Dark-blood' dual-energy computed tomography angiography for thoracic aortic wall imaging," *Eur. Radiol.*, 2019.

[9] C. H. Hua, N. Shapira, T. E. Merchant, P. Klahr, and Y. Yagil, "Accuracy of electron density, effective atomic number, and iodine concentration determination with a dual-layer dual-energy computed tomography system," *Med. Phys.*, vol. 45, no. 6, 2018.

[10] Y. Wang *et al.*, "Metal artifacts reduction using monochromatic images from spectral CT: Evaluation of pedicle screws in patients with scoliosis," *Eur. J. Radiol.*, vol. 82, no. 8, Aug. 2013.

[11] K. Mei *et al.*, "Bone mineral density measurements in vertebral specimens and phantoms using dual-layer spectral computed tomography," *Sci. Rep.*, vol. 7, no. 1, 2017.

[12] M. M. Goodsitt, E. G. Christodoulou, and S. C. Larson, "Accuracies of the synthesized monochromatic CT numbers and effective atomic numbers obtained with a rapid kVp switching dual energy CT scanner," *Med. Phys.*, vol. 38, no. 4, 2011.

[13] E. Shefer *et al.*, "State of the Art of CT Detectors and Sources: A Literature Review," *Curr. Radiol. Rep.*, vol. 1, no. 1, 2013.

[14] A. Euler *et al.*, "Initial Results of a Single-Source Dual-Energy Computed Tomography Technique Using a Split-Filter: Assessment of Image Quality, Radiation Dose, and Accuracy of Dual-Energy Applications in an in Vitro and in Vivo Study," *Invest. Radiol.*, vol. 51, no. 8, Aug. 2016.

[15] T. G. Flohr *et al.*, "First performance evaluation of a dual-source CT (DSCT) system," *Eur. Radiol.*, vol. 16, no. 2, Feb. 2006.

[16] G. J. Pelgrim *et al.*, "Accuracy of iodine quantification using dual energy CT in latest generation dual source and dual layer CT," *Eur. Radiol.*, vol. 27, no. 9, Sep. 2017.

[17] A. P. Sauter *et al.*, "Accuracy of iodine quantification in dual-layer spectral CT: Influence of iterative reconstruction, patient habitus and tube parameters," *Eur. J. Radiol.*, vol. 102, May 2018.

[18] L. Ananthakrishnan *et al.*, "Spectral detector CT-derived virtual non-contrast images: comparison of attenuation values with unenhanced CT," *Abdom. Radiol.*, vol. 42, no. 3, 2017.

[19] M.J. Berger; J.H. Hubbell; S.M. Seltzer; J. Chang; J.S. Coursey; R. Sukumar; D.S. Zucker; and K. Olsen, "XCOM: Photon Cross Sections Database," *NIST, PML, Radiation Physics Division*. [Online]. Available: https://www.nist.gov/pml/xcom-photon-cross-sections-database. [Accessed: 12-Oct-2019].

[20] A. P. Sauter *et al.*, "Perfusion-ventilation CT via three-material differentiation in dual-layer CT: a feasibility study," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019.

Low Dose CT Denoising via Joint Bilateral Filtering and Intelligent Parameter Optimization

Mayank Patwari, Ralf Gutjahr, Rainer Raupach, and Andreas Maier

***Abstract*—Denoising of clinical CT images is an active area for deep learning research. Current clinically approved methods use iterative reconstruction methods to reduce the noise in CT images. Iterative reconstruction techniques require multiple forward and backward projections, which are time-consuming and computationally expensive. Recently, deep learning methods have been successfully used to denoise CT images. However, conventional deep learning methods suffer from the 'black box' problem. They have low accountability, which is necessary for use in clinical imaging situations. In this paper, we use a Joint Bilateral Filter (JBF) to denoise our CT images. The guidance image of the JBF is estimated using a deep residual convolutional neural network (CNN). The range smoothing and spatial smoothing parameters of the JBF are tuned by a deep reinforcement learning task. Our actor first chooses a parameter, and subsequently chooses an action to tune the value of the parameter. A reward network is designed to direct the reinforcement learning task. Our denoising method demonstrates good denoising performance, while retaining structural information. Our method significantly outperforms state of the art deep neural networks. Moreover, our method has only two parameters, which makes it significantly more interpretable and reduces the 'black box' problem. We experimentally measure the impact of our intelligent parameter optimization and our reward network. Our studies show that our current setup yields the best results in terms of structural preservation.**

I. Introduction

The radiation dose in clinical CT needs to be as low as reasonably achievable to warrant patient's safety. Reducing radiation dose creates noise in the reconstructed images, which may obscure critical details and low contrast objects [1]. The non-stationary nature of CT noise means that simple filtering approaches are not always successful in denoising. Moreover, filtering also removes the edge information present in the image, particularly the small low-contrast features . Therefore, denoising the reconstructed images requires a more intelligent solution, and is currently an active area of research.

Non linear edge-preserving filters such as the bilateral filter have been used to denoise clinical CT data [2], [3]. Most clinically approved methods currently use iterative reconstruction methods [4] to denoise images [5], [6]. Deep learning has also been successfully applied to the clinical CT denoising problem [7], [8], [9]. However, iterative reconstruction methods are time-consuming and computationally expensive. Deep learning methods, while outperforming iterative reconstruction, suffer from a lack of accountability, the so-called 'black box' problem. It is absolutely crucial for any denoising approach to be completely interpretable for possible clinical approval. A middle ground between the underperforming known methods and deep learning would be to use deep learning to tune or optimize the performance of known techniques, such as by optimizing the parameters [10] or altering the regularization strength [11].

M. Patwari, and A. Maier are with the Pattern Recognition Lab at the Friedrich-Alexander University of Erlangen-Nuremberg, 91058, Erlangen, Germany e-mail: mayank.patwari@fau.de.

M. Patwari, R. Gutjahr, and R. Raupach are with Siemens Healthcare GmbH, 91301, Forchheim, Germany.

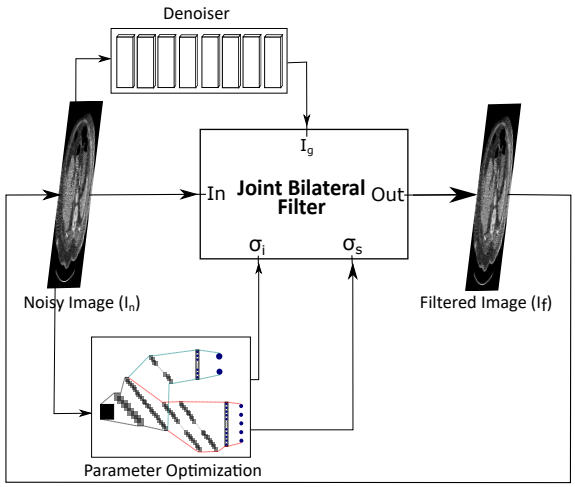


Fig. 1. Our denoising scheme with the JBF, parameter optimization scheme, and guidance image estimating denoiser.

In this paper, we aim to combine the strengths of both known operators and deep learning. We introduce a Joint Bilateral Filter (JBF) [2] with a deep learned guidance image to denoise the image. The JBF has two parameters, which are tuned by reinforcement learning [12]. The reinforcement learning scheme tunes our parameters pixelwise to obtain the best quality image overall. The JBF is applied iteratively for a given number of iterations. The parameters are optimized for every iteration. We compare our methods to state of the art deep learning networks for the task of low dose CT denoising. Our proposed method shows good denoising performance and preserves structural information. Furthermore, our method outperforms deep learning methods, while being significantly more interpretable.

II. Materials and Methods

A. Joint Bilateral Filtering concept

We revisit the concept of the JBF for better explaining our methods. The JBF is a variant of the standard bilateral filter, which uses a separate guiding image to smooth the difference

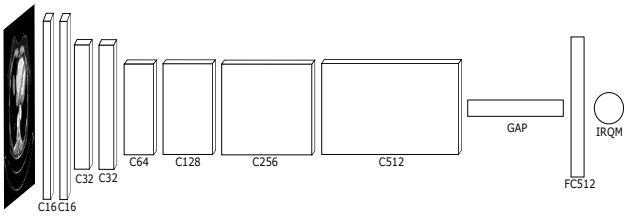


Fig. 2. The quality network used to measure image quality and noise. C represents the number of convolutional filters.

in intensities. The operation of the JBF is defined by the following equation:

$$I_f(x) = \frac{\sum_{o \epsilon N(x)} I_n(o) G_{\sigma_s}(x-o) G_{\sigma_i}(I_g(x) - I_g(o))}{\sum_{o \epsilon N(x)} G_{\sigma_s}(x-o) G_{\sigma_i}(I_g(x) - I_g(o))} \quad (1)$$

where I_n is the noisy image, I_f is the filtered image, x is the spatial coordinate, $N(x)$ is the neighborhood of x taken into account for the filtering operation, I_g is the guidance image estimated by the method described in Section II-C and G_σ is a Gaussian operator. The first Gaussian operator in Equation 1 is a spatial filter, while the second operator is the range filter. The Gaussian operator is defined by

$$G_\sigma(x) = \frac{e^{-x^2/2\sigma^2}}{2\sigma^2} \quad (2)$$

where σ is a parameter controlling the level of smoothing. σ is usually a hand tuned parameter. Therefore, the JBF at any point has three inputs, the guidance image I_g, and two σ parameters, σ_i for the range operator, and σ_s for the spatial operator. We used a Gaussian operator of size $9 \times 9 \times 5$ across the entire image.

B. Quality Analysis

We train a CNN to estimate the Image Reconstruction Quality Metric (IRQM). We use the architecture and training hyperparameters described in Patwari et al. [13], with two differences **(1)** we leave out the batch normalisation layers **(2)** the output of the final neuron is passed through a sigmoid layer to constrain the output to [0, 1]. The network is trained using 4 CT volumes. Each volume was reconstructed with standard clinical dose, and with doses reduced by 50% and 75%. The gradient structural similarity score between the standard dose image and the reduced dose image is used as a proxy quality score for training. The heatmap loss is included as a secondary loss function. A network diagram is present in Fig. 2.

C. Estimating a Guidance Image

The guidance image is estimated to be a noise free image. We can solve the following optimization problem to find a noise removal function F:

$$\min_F ||I_{nf} - F(I)||^2 \quad (3)$$

where I_{nf} is the noise free image and I is the noisy image. Solving equation 3 will give us the ideal noise removal function F. We parameterize F using a residual CNN (Fig. 3) [8]. The residual CNN has 4 convolutional layers and 4 deconvolutional layers, each with 32 filters. The convolutional layers have filters of size $3 \times 3 \times 3$. No padding is applied, shrinking the input image in each dimension. The deconvolutional filters have filters of size 3×3. This results in a receptive field of $17 \times 17 \times 9$. A leaky ReLU activation follows each layer, except the final layer.

The network is trained using CT volumes of 10 patients. Each volume is reconstructed with the standard clinical dose and with simulated dose reductions. The standard dose images act as ground truth volumes I_{nf}. The reduced dose volumes are reconstructed at doses of 5%, 10%, 25% and 50% of the standard clinical dose, and are used as training volumes. The network is trained using non-overlapping image patches of size $64 \times 64 \times 9$. This is shrunk to $56 \times 56 \times 1$ by the convolutional layers, and expanded to 64×64 by the deconvolutional layers. The training took place over 70,000 iterations with 48 patches per iteration.

D. Tuning parameters via deep reinforcement learning

We introduce a convolutional network NET as our agent, which would choose a parameter and subsequently choose the action that affects this parameter. We limit the network to choose between the two σ parameters, and between the following five actions **(1)** increase by 50% **(2)** increase by 10% **(3)** do nothing **(4)** decrease by 10% **(5)** decrease by 50%.

The network is used to maximize a reward function r. r in this case is the output of our quality network Q_n, described in section II-B.

$$r = Q_n(s') - Q_n(s) \quad (4)$$

where s' is the image after the parameter tuning step and s is the original image. We create a loss which punishes deviations from the Bellman equation. The Bellman equation is given by

$$Q^*(s,a) = r + \gamma \max_{a'} Q^*(s', a') \quad (5)$$

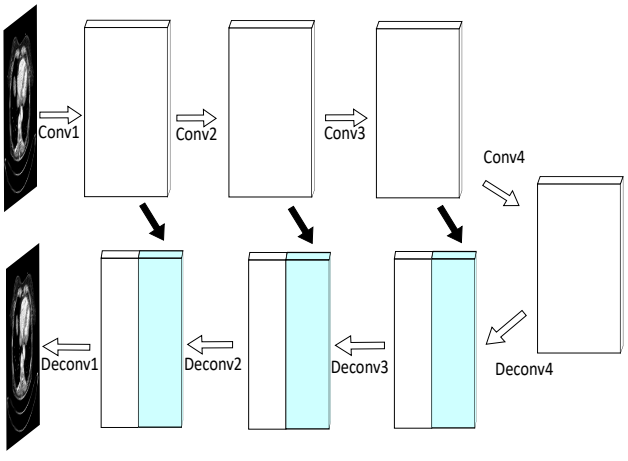


Fig. 3. The residual network used to create the guidance image. Black arrows represent the residual connections.

where a is the action which changes s to s' and γ is the discount factor for future rewards.

We parameterize the Q^* function with our CNN NET. NET has three sets of parameters, θ_p which influence the parameter choosing, θ_t which control the parameter tuning, and θ_c, which are shared parameters. We split our target term into two target terms. Following standard reinforcement learning techniques [12], we use an older version of our policy network (NET') as a target.

$$\begin{aligned} y_1 &= r + \gamma \max_{p'} Q^*(s', p'; NET'(\theta_c, \theta_p)) \\ y_2 &= r + \gamma \max_{t'} Q^*(s', t'; NET'(\theta_c, \theta_t)) \end{aligned} \tag{6}$$

The network is trained by solving the following minimization problem:

$$\begin{aligned} L_1 &= [y_1 - Q(s, p; NET(\theta_c, \theta_p))]^2 \\ L_2 &= [y_2 - Q(s, t; NET(\theta_c, \theta_t))]^2 \\ \min_{NET} L &= L_1 + L_2 \end{aligned} \tag{7}$$

The network chosen (Fig. 4) has two paths, one which chooses the parameter to be tuned, and the other which tunes the parameters. The first path has 3 convolutional layers, one fully connected layer, and an output layer. The convolutional layers have 32, 64, and 64 layers. The fully connected layer has 128 neurons. There are two output neurons. The second path has 4 convolutional layers, one fully connected layer, and one output layer. The convolutional layers have 32, 64, 128 and 128 layers. The fully connected layer has 256 neurons. There are 5 output neurons. The first two convolutional layers are shared between the paths, and have $3 \times 3 \times 3$ convolutions. All other convolutional layers have 3×3 convolutions. All layers except the output layers are followed by leaky ReLU activations.

The network is trained for 120,000 iterations. γ is fixed at 0.99. NET' is updated to NET every 300 steps. We allow 20 tuning steps before application of the bilateral filter. The input to the network is an image patch of size $9 \times 9 \times 5$.

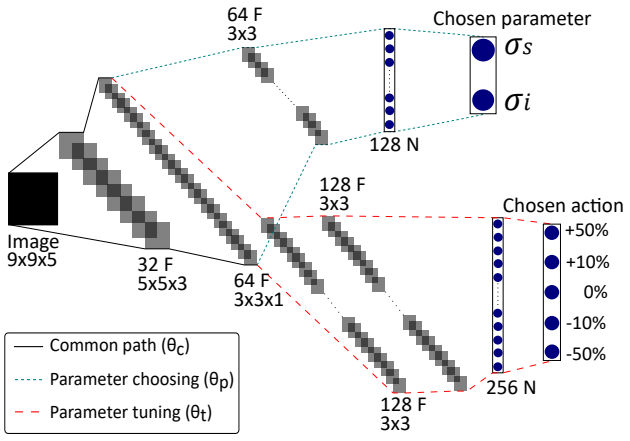


Fig. 4. Our RL network architecture with separate parameter choosing and parameter tuning paths. F represents the number of convolutional filters, N the number of neurons, and the dimensions the size of the outputs.

TABLE I
MEAN AND STANDARD DEVIATIONS OF THE QUALITY METRIC SCORES OF OUR METHOD AND VARIANTS. PLAIN JBF APPLICATION YIELDS THE HIGHEST PSNR, WHILE INCLUDING NETWORK BASED PARAMETER TUNING YIELDS THE HIGHEST SSIM.

	PSNR	**SSIM**
Low Dose CT [14]	43.12 ± 1.021	0.9636 ± 0.0073
Joint Bilateral Filter (JBF)	$\mathbf{46.79 \pm 0.8685}$	0.9770 ± 0.0046
JBF + PT-GT	43.40 ± 1.025	0.9656 ± 0.0072
CPCE3D [8]	45.43 ± 0.6914	0.9817 ± 0.0029
GAN[7]	41.87 ± 0.7079	0.9398 ± 0.0079
Our method	45.02 ± 0.9709	$\mathbf{0.9856 \pm 0.0027}$

The training set consists of 20 slabs of size $256 \times 256 \times 16$ reconstructed at 25% dose. The parameter chosen, action chosen, slab, filtered slab, and increased quality score are calculated for each patch. 3200 patches are added at random into the training pool. 512 patches are chosen at random for training at each optimization step.

III. EXPERIMENTS AND RESULTS

A. Data and Evaluation Metrics

We used the data of the Grand Low Dose CT challenge [14] as our testing data. This dataset consists of body CT scans of 10 different patients, each reconstructed with both standard dose, and with 25% of the standard dose. We use the images reconstructed with 1mm slice thickness with a medium smooth B30 kernel. The reduced dose images are our input images, while the standard dose images are treated as reference images.

We use the PSNR and SSIM [15] as our metrics for noise removal and structural preservation respectively. The SSIM is a significantly more important predictor of the diagnostic quality of the image, as the PSNR promotes smoothing of images. Due to the limited number of patients, we use the Wilcoxon's signed rank test to assess statistical significance. A p-value of less than 0.05 is considered as the threshold for statistical significance.

B. Denoising and comparison to state of the art

In our denoising experiments, we denoise the image by iterative application of the JBF. For each iteration, we allow five optimization steps to tune our parameters, before applying the bilateral filters. The parameters are reset to their original guesses and the optimization process starts anew for each iteration. We restrict the number of JBF iterations to four in our testing phase.

Our method significantly improved PSNR and SSIM compared to the noisy image ($w = 0.000$, $p = 0.0051$). This shows that our method reduces image noise while preserving the structural information present in the images. An example can be seen in Fig. 5. Our method outperforms state of the art deep learning networks such as CPCE3D [8] and GAN [7] in terms of SSIM. However, CPCE3D has a higher PSNR than our method. This could indicate superior denoising at the cost of feature preservation. The results can be viewed in Table I.

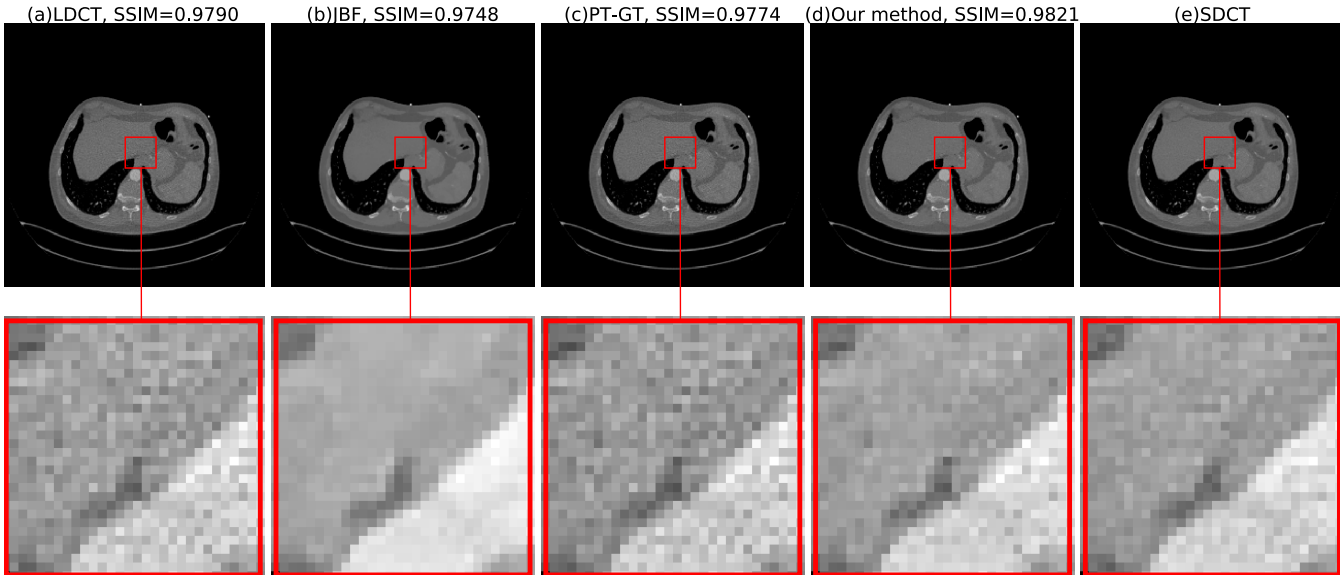


Fig. 5. An image showing the (a) reduced dose CT image (b) image denoised by JBF (c) image denoised by using the reinforcement learning technique which optimizes based on the ground truth image (PT-GT) (d) our method and (e) the standard dose reference image. Our method yields the closest looking image to the reference image (SSIM = 0.9821). The images are reconstructed with a window of [-800, 1200].

C. Ablation study

Impact of parameter tuning: We observe the impact of the JBF with and without parameter tuning. Use of parameter tuning improves the SSIM (w = 0.000, p = 0.0051), but significantly decreases the PSNR (w = 0.000, p = 0.0051). Therefore, the use of parameter tuning reduces the amount of denoising, but improves the amount of structural information preserved.

Using our reward network: We compare the use of our reward network (Section II-B) with the mean-squared-error based reward function used in Shen et. al. [11]. We call this method parameter tuning - ground truth (PT-GT). PT-GT reduces both the PSNR and the SSIM significantly (w = 0.000, p = 0.0051). Therefore, the use of our reward network results in better CT image quality compared to PT-GT.

IV. Conclusion

In this study, we introduce a method which tunes JBF parameters to thoroughly denoise low dose CT images. Our method provides good denoising performance while retaining the structural information. The use of the JBF reduces noise, while the parameter optimization helps in retaining structures. Our method is outperformed several existing state of the art solutions for low dose CT denoising. Additionally, our method is significantly more interpretable, and reduces the 'black box' problem that exists in deep learning, which improves applicability in clinical imaging.

References

[1] A. Oppelt, "Noise in Computed Tomography," in *Imaging Systems for Medical Diagnostics*, 2nd ed., S. Aktiengesselschaft, Ed. Publicis Corporate Publishing, 2005, ch. 13.1.4.2, p. 996.

[2] M. Manhart, R. Fahrig, J. Hornegger, A. Doerfler, and A. Maier, "Guided Noise Reduction for Spectral CT with Energy-Selective Photon Counting Detectors," in *Proceedings of the Third CT Meeting*, no. 1, 2014, pp. 91–94.

[3] A. Manduca, L. Yu, J. D. Trzasko, N. Khaylova, J. M. Kofler, C. M. McCollough, and J. G. Fletcher, "Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT," *Medical Physics*, vol. 36, no. 11, pp. 4911–4919, 2009.

[4] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *Journal of Theoretical Biology*, vol. 36, no. 1, pp. 105–117, 1972.

[5] J. C. Ramirez-Giraldo, K. L. Grant, and R. Raupach, "ADMIRE : Advanced Modeled Iterative Reconstruction," 2015. [Online]. Available: https://www.siemens-healthineers.com/computed-tomography/technologies-innovations/admire

[6] E. Angel, "AIDR 3D Iterative Reconstruction : Integrated, Automated and Adaptive Dose Reduction," 2012. [Online]. Available: https://us.medical.canon/download/aidr-3d-wp-aidr-3d

[7] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Generative Adversarial Networks for Noise Reduction in Low-Dose CT," *IEEE Trancations on Medical Imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[8] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3D Convolutional Encoder-Decoder Network for Low-Dose CT via Transfer Learning from a 2D Trained Network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1522 – 1534, 2018.

[9] A. Maier and R. Fahrig, "GPU Denoising for Computed Tomography," in *Graphics Processing Unit-Based High Performance Computing in Radiation Therapy*, 1st ed., J. Xun and S. Jiang, Eds. Boca Raton, Florida, USA: CRC Press, 2015, pp. 113–128.

[10] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3116–3132, 2010.

[11] C. Shen, Y. Gonzalez, L. Chen, S. B. Jiang, and X. Jia, "Intelligent Parameter Tuning in Optimization-Based Iterative CT Reconstruction via Deep Reinforcement Learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1430–1439, 2018.

[12] R. S. Sutton and A. G. Barto, *Reinforcement Learning*, 2nd ed. Cambridge, Massachussets: MIT Press, 2014.

[13] M. Patwari, R. Gutjahr, R. Raupach, and A. Maier, "Measuring CT Reconstruction Quality with Deep Convolutional Neural Networks," in *Proceedings of the Second International Workshop on Machine Learning in Medical Image Reconstruction*, F. Knoll, A. Maier, D. Rueckert, and J. Ye, Eds. Shenzhen: Springer, 2019, pp. 113–124.

[14] C. H. Mccollough, "Low Dose CT Grand Challenge," 2016.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, S. Member, E. P. Simoncelli, and S. Member, "Image Quality Assessment : From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

GPU-accelerated Monte Carlo simulations of x-rays through CAD objects

Tobias Funk, Daniel S. Badali, and Oleg Konings

***Abstract*—Although Monte Carlo simulations are an indispensable tool for studying a wide range of applications, their usefulness is often limited by their long computation times. In fact, it is not uncommon for simulations to take hours or even days, depending on the complexity of the objects being studying. To this end, we herein present a new x-ray Monte Carlo library that makes use of the massively parallel architecture of Graphics Processing Units (GPUs) to significantly improve the computation times. The library includes the ability to propagate particles through objects defined by triangulated surfaces, which allows for the support of CAD-modelled objects. To benchmark this library, we have incorporated it into PRISM, which is a tested software platform which manages the simulations of computed tomography projection images. Using GEANT4 as the benchmark, we found that the GPU-accelerated library was able to achieve statistically similar results (difference on the order of a few percent), even though it was restricted to a limit set of particle types and interactions. A speedup factor of about 640x was achieved.**

***Index Terms*—Graphical Processing Unit, Monte Carlo, Simulation, Computed Tomography**

I. Introduction

Monte Carlo simulations are the gold-standard technique for studying the propagation of x-rays through matter. Because of this, they are an indispensable part of the workflow of many fields. For instance, in the field of radiotherapy, Monte Carlo simulations are used to calculate the dose distribution produced by the radiation source**Error! Reference source not found.**. As another example, such simulations are the workhorse of modern high-energy physics research[2].

Although x-ray Monte Carlo simulations serve a wide range of applications, the requirements for such simulations to be useful is shared across all fields: most importantly, the simulations must be accurate. In this case, accuracy refers to both the implementation of the physics of the x-ray-matter interactions and the modeling of the objects being simulated. Of more practical importance is the speed at which the simulations can be performed, a requirement which becomes more challenging to meet as the complexity of the object and system increases.

There are a number of Monte Carlo software libraries that have met these requirements with mixed success. Some notable libraries are GEANT4[3] and MCNP[4]. All of these libraries have favored accuracy and have strived to incorporate the up-to-date models of the physics of x-ray-matter interactions. However, this accuracy comes at the cost of speed, and it can often take inordinate amounts of time (e.g. hours or days) to achieve sufficient statistics when running on typical Central Processing Units (CPUs). The most computationally expensive part of the simulation is the navigation of the x-ray particles through the object's geometry[5].

Fortunately, Monte Carlo simulations are "embarrassingly parallel", since the propagation of each x-ray is entirely independent of all other x-rays. Because of this, they map well to the massively parallel architecture of Graphical Processing Units (GPUs), which can dramatically improve the simulation time. Although preliminary progress has been made to port existing CPU-based Monte Carlo libraries to GPUs[5], there still lacks a library that meets both the speed and accuracy requirements.

We propose to develop a GPU-accelerated x-ray Monte Carlo library that is both accurate and fast. To test this library, we have made use of our PRISM software tool[6], which offers a flexible framework to interchange Monte Carlo libraries while maintaining a common interface. Because of its design, PRISM allows for a direct comparison between the new GPU-accelerated library and existing libraries such as GEANT4.

II. PRISM

The Particle/Ray Interaction Simulation Manager (PRISM) is an internal software platform developed by Triple Ring Technologies to perform simulations of Computed Tomography (CT) projection images. Originally designed for security applications, we have since used successfully used PRISM in a wide range of CT applications.

PRISM allows for the configuration of all aspects of the system to be simulated, including the spectrum, focal spot shape, and system geometry. There is also a highly configurable detector module, which can inject noise and artifacts due to sources such as veiling glare, electronic noise, quantum efficiency, quantum noise, detector lag, and crosstalk.

PRISM was architected in a way to facilitate the easy incorporation of different Monte Carlo libraries while

Tobias Funk, Daniel S. Badali, and Oleg Konings are with Triple Ring Technologies, 39655 Eureka Dr, Newark CA 94560. (tfunk@tripleringtech.com, dbadali@tripleringtech.com, okonings@tripleringtech.com)

maintaining a common interface for the user. This hides the unique details of each library's application programming interface from the user, which often has a steep learning curve. This design allows us to do a direct comparison between the newly developed GPU-based library and an established library. For the latter, we chose to use GEANT4.

With an emphasis on accuracy, PRISM represents objects in the CAD format of STL files[7], which are triangulated, watertight surfaces. Many Monte Carlo libraries (such as GEANT4) support such triangulated surfaces.

Fig. 1 shows a simulated projection image produced by PRISM using GEANT4. The packed suitcase was assembled using common items in a CAD program.

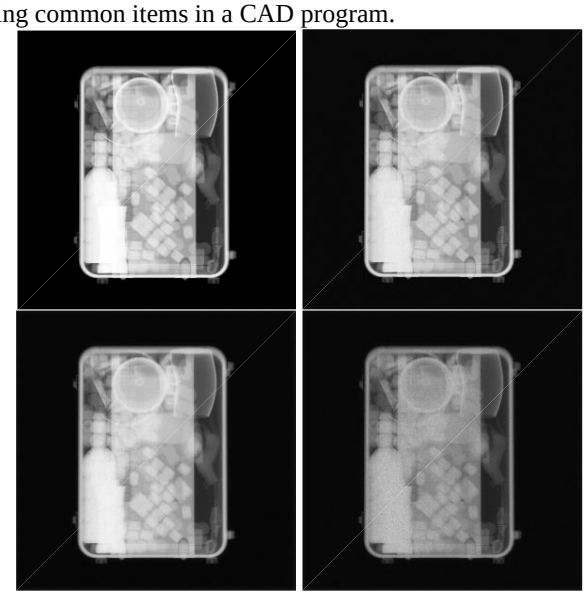


Fig. 1 Simulated projection images demonstrating the capabilities of PRISM with GEANT4. (Top left) Ray tracing simulation with scatter. (Top right) With Poisson noise and limited detected quantum efficiency. (Bottom left) Also with veiling glare. The magnitude of the veiling glare is exaggerated in this figure. (Bottom right) Also with crosstalk and exaggerated electronic noise.

III. GPU-Monte Carlo Library

In designing our GPU-accelerated Monte Carlo library, we strove to retain upon the accuracy of existing Monte Carlo libraries. However, mature libraries often have extensively detailed physics, typically including a wide number of particle types and particle-matter interactions. Our library contains a pared down subset of these particles and interactions; in particular, we start with only supporting x-ray photons and limit the interactions to Compton scattering, coherent scattering, and absorption. Table 1 lists all the current and planned features and their implementation status.

As the photons propagate through a triangulated object, it is very computationally expensive to test each triangle for intersection with the photon's trajectory. In order to avoid this, we use the standard technique of constructing a Bounding Volume Hierarchy (BVH) tree. A BVH is a tree structure, where the bounding volume of each parent node contains the bounding volume of each of its children. Sibling volumes are allowed to overlap and do not necessarily need to fully partition the volume of the parent. The leaf nodes each contain a single triangle, which minimizes the number of triangle intersections that need to be calculated. Our implementation of the BVH tree was inspired by the GPU-accelerated optical Monte Carlo library Chroma[8], and we used the Watertight ray-triangle algorithm[9].

Table 1: Current and planned featured for the GPU-accelerated x-ray Monte Carlo library

Feature	**Status**
Ray tracing	Fully implemented
Compton scattering	Fully implemented
Absorption	Fully implemented
Coherent scattering	In progress
Secondary particles	Planned
Electrons	Planned

IV. Benchmarking

Because of the flexibility of the PRISM architecture, switching between GEANT4 and the GPU library can be achieved with a single flag in the configuration. The pre- and post-processing of the simulation (such as the loading of the STL files, the generating of the initial photon phase space, the creating of the output projection images, etc.) is independent of the choice of Monte Carlo library. That way, all that is being compared is the navigation and photon-matter interactions.

Figs. 3 and 4 compare the results of the GEANT4 and GPU Monte Carlo libraries with a disk-based phantom. As seen in Fig. 3, the GPU library is able to reproduce the GEANT4 results almost identically, in spite of its limited particle and interaction types (see Table 1). A histogram of the difference between the two images is shown in Fig. 4, and it can be seen that the difference is on the order of a few percent, which can likely be attributed to the lack of coherent scattering in the GPU-library.

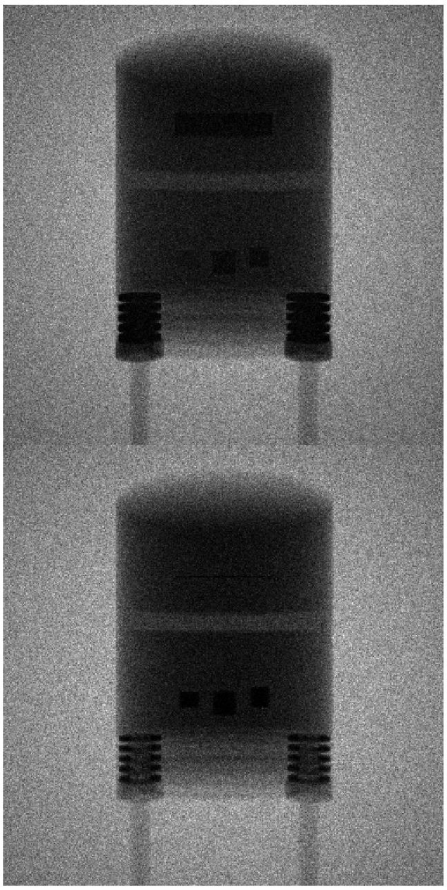


Fig 3. Comparison between the simulated projection images of a phantom produced by GEANT4 (top) and the new GPU-accelerated Monte Carlo library (bottom)

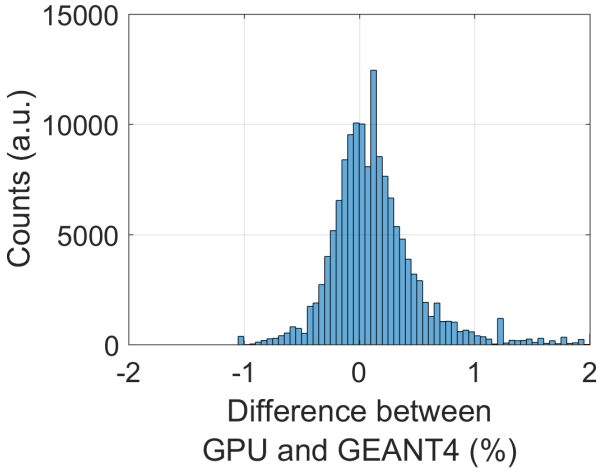


Fig. 4 Histogram of the pixel-by-pixel differences between the projection images (shown in Fig. 3) produced by GEANT4 and the GPU-accelerated Monte Carlo library

The simulation time per photon for the two libraries can be found in Table 2. It can be seen that the GPU-accelerated library is significantly faster than GEANT4, with a speed up factor of about 640x.

Table 2: Simulation time per photon for the results shown in Fig. 3. The GEANT4 simulations were performed on a single-core CPU

GEANT4	GPU
450 µs	0.7 µs

V. Outlook

We are strongly encouraged by the quality of the results produced by the GPU-accelerated Monte Carlo library. As we continue to develop the library, more particle types and interactions will be included to allow for more realistic simulations. Using the PRISM framework, we will continue to be able to benchmark against established libraries.

VI. References

[1] S. Guatelli and S. Incerti, "Monte Carlo simulations for medical physics: From fundamental physics to cancer treatment", *Phys. Med.*, vol. 33, pp. 179 – 181, 2017

[2] A. Buckley, *et al.*, "Monte Carlo event generators for high energy particle physics event simulation", arXiv: 1902.01674 [hep-ph], Feb, 2019

[3] S. Agostinelli *et al.*, "Geant4 – A simulation toolkit", *Nucl. Instr. Meth. Phys. Res. A*, vol. 506, pp. 250 – 303, 2003

[4] T. Goorley, *et al.*, "Initial MCNP release overview", *Nucl. Tech.*, vol. 180, pp. 298 – 315, 2012

[5] O. Seiskari, J. Kommeri, and T. Niemi, "GPU in physics computation: Case Geant4 navigation", arXiv:1209.5235 [physics.comp-ph], Sep 2012

[6] T. Funk, D. Badali, and T. Gilat-Schmidt, "PRISM: A new software tool for simulating realistic CT data with CAD model based objects", presented at the *CT Meeting*, Regensburg, Germany, 2016

[7] 3D Systems, Inc., "Stereolithography Interface Specification", 1988

[8] S. Siebert and A. LaTorre, "Fast optical Monte Carlo simulation with surface-based geometries using Chroma", 2011

[9] S. Woop, C. Benthin, and I. Wald, "Watertight ray/triangle intersection", *J. Comp. Graph. Tech.*, vol. 2, pp. 65 – 82, 2013

[10] W. P. Risk, G. S. Kino, and H. J. Shaw, "Fiber-optic frequency shifter using a surface acoustic wave incident at an oblique angle," *Opt. Lett.*, vol. 11, no. 2, pp. 115–117, Feb. 1986.

Deep Learning Based Distortion Correction and Material Decomposition for Photon Counting CT: A Simulation Study

Matt D. Holbrook, Darin P. Clark, and Cristian T. Badea

***Abstract*—We are now at the cusp of a major advancement in CT due to photon-counting detectors (PCD). While the benefits of photon counting (PC) CT are significant, the performance of current PCDs is limited by physical effects distorting the spectral response. Here, we propose deep learning (DL) solutions for spectral distortion corrections and material decompositions for PCCT. The spectral measurements of a nuclear source, Ba133 acquired with a real PCD, were used to fit a distortion model. This distortion model was used to degrade the ideal x-ray spectra in our realistic simulations. Virtual cone beam acquisition and CT reconstructions of two digital phantoms were used for training and testing of two CNNs (U-net configurations) that performed corrections and image-based decompositions. Our preliminary investigation showed promising results, lowering image noise, error, and increasing the accuracy and precision of CNN material maps compared to results provided by inversion of non-corrected PCCT data. This performance came with some loss in resolution. Our next step will be to adapt this method to real experimental data acquired in 3D printed phantoms using transfer learning.**

I. Introduction

New exciting applications can be facilitated by energy-resolving photon-counting detectors (PCDs) which can quantitatively differentiate multiple CT probes with a single CT scan. Our group has advanced preclinical photon-counting (PC) CT by building prototype systems and demonstrating their value in cancer [1] and cardiac [2] studies. However, the full potential of PCCT has not yet been realized. PCDs suffer from spectral distortions (charge sharing, pulse pile-up). Moreover, their spectral and spatial resolution are inversely related, and K-edge imaging of iodine is challenging with current CdTe-based PCDs.

Data-driven deep learning (DL) methods promise to enhance many aspects of X-ray CT including for PCCT [3]. Examples of successful DL applications include metal artifact reduction [4], synthesizing missing sinogram data [5], generating monochromatic images from single-spectrum CT scans [6], and reducing noise in low-dose CT [7].

Spectral distortions inherent in PCD technology such as charge sharing, K-escape, and pulse pileup [8] degrade the recorded spectral response of the detector and the accuracy of these decompositions. Several methods have been proposed to correct spectral distortions. For example, by modeling the imaging process, including sources of spectral distortion, maximum likelihood reconstruction methods [9] or Monte Carlo methods [10] can be employed to improve imaging. Unfortunately, because PCD spectral distortion is non-linear, these inversions can be computationally expensive, numerically unstable, and sensitive to errors in the forward model. These drawbacks have motivated a number of data-driven approaches to compensate for spectral distortions including lookup table methods and methods based on artificial neural networks (ANNs) [11].

Here, we propose new deep learning (DL) solutions for PCCT. Given spatial correlations inherent in spectral distortions, we hypothesize that CNNs can be trained to correct spectral distortions and provide image-based decompositions. We have previously demonstrated a novel spectral distortion correction method [12] that worked by training a shallow ANN to map distorted projections acquired with a real system and a 3D printed phantom to ideal projections digitally acquired from the digital model of the same phantom. The trained model could then be applied to correct any real, distorted projections prior to reconstruction and material decomposition. This approach is very attractive because it does not require extensive knowledge of the imaging system. PCCT data is typically decomposed into physical and/or material basis functions. We have also tested material decomposition based on a CNN [13]. In this paper we build on the success of our previous attempts and use two CNNs that will perform distortion corrections and material decompositions. Realistic simulations are used to verify performance.

II. Methods

Conducting realistic digital phantom experiments is crucial to the development of DL solutions for distortions and material decomposition for PCCT data. We next introduce the spectral model, phantoms, CNN structures and quality metrics.

A. Spectral model

The validity of work performed on simulated data depends on the accuracy of the simulations. We have created a spectral model to mimic our CdTe PCD with four energy thresholds and 150 μm pixel pitch (Dectris AG, Switzerland). The spectral performance of our PCD was measured using a nuclear source, Ba133. Thresholds were swept from 25 to 85 keV, exposing each set of thresholds for 10 seconds. Differentiation of the raw signal gave our

M. Holbrook, D. Clark, and C. Badea are with the Center for In Vivo Microscopy, Department of Radiology, Duke University, Durham, NC 27710. Contact: cristian.badea@duke.edu

All work was performed at the Duke Center for In Vivo Microscopy. This work was also supported by the NIH National Cancer Institute (R01 CA196667, U24 CA220245).

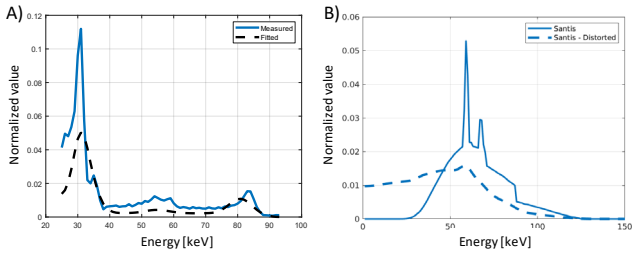


Fig. 1. Spectral modeling of a CdTe PCD imaging system. A) Recorded spectrum of Ba133 nuclear source and the fitted spectrum. B) Spectrum used in simulations (125 kVp, 0.1 mm Pb filtration) along with the distorted spectrum.

spectral response. This response was used to fit a distortion model as presented by Schlomka et al. [9]. Once fit, the model was used to degrade an ideal 125 kVp spectrum used in our simulations (Fig. 1). The ideal spectral response that included the x-ray source (tungsten-based anode), Pb filter (0.1 mm), and the CdTe detector was created in MATLAB using the Spektr package [14].

B. *Digital phantoms*

Digital 3D phantoms were used for both training and testing our models. For training, a series of virtual phantoms (440^3 voxels, 0.125 µm voxel size) containing random tetrahedra were generated (Fig. 2). The tetrahedra were assigned random material concentrations of aqueous I, Gd, or Au up to 15 mg/mL. Reference vials for each material were also added. Half of the tetrahedra only contained water. Six of these phantoms, along with their respective projections (360 per phantom), were created. Energy-dependent cone beam 3D projections were created using the ideal and distorted spectral response functions and our in-house forward projector [15]. Poisson noise was added to the projections to match noise level in experimental reconstructions. The labels used in training were noise free. The PCD thresholds were set to 25, 34, 50 and 80 keV. The last three thresholds match the K-edges of I, Gd, and Au.

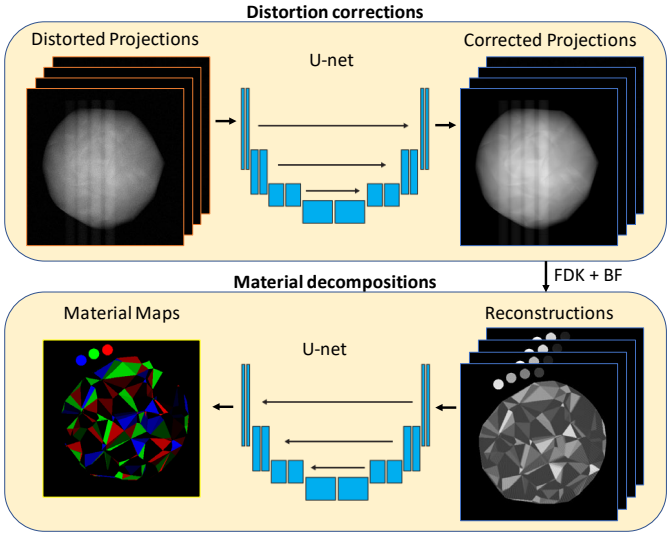


Fig. 2. Data processing pipeline for spectral distortion correction and material decompositions. Distorted, normalized, projections are passed through a residual U-net, resulting in corrected projections. Once reconstructed, these are passed through a second CNN to create material maps.

The networks were tested on a separate digital phantom that we have used previously [16]. This phantom (440x440x344) contains the same aqueous K-edge contrast materials as the training data in concentrations up to 15 mg/mL. Each material has a corresponding set of slices with spherical and bar patterns. This phantom allows for measurement of image quality metrics related to the image noise, image resolution, and material map accuracy.

C. *Network setup and processing*

Two CNNs were used in this work, one for distortion correction and one for material decomposition (Fig. 2). Both networks were similar to the conventional U-net [17]. Generally, pairs of convolutional layers followed by ReLU activation functions and pooling/unpooling operations compose an encoder/decoder structure. Skip connections re-introduce high frequency information via concatenation from the encoder to the decoder. Each network was trained on simulated data, after which the material decomposition finished training on outputs of the distortion correction network.

The cost functions in both networks are similar, only differentiated by a weighting factor. The cost function was set to value data fidelity with a high emphasis on preserving high frequencies. Thus, mean squared error was used to ensure that low frequency contrast was treated well, and a term comparing image gradients was added to benefit higher frequencies. The cost $L(\Theta)$ is given as follows for labels y and predictions p:

$$\mathcal{L}(\Theta) = \sum_{n=1}^{N}\left[\frac{\left(y_n - p_n(\Theta)\right)^2}{N}\right] - \lambda \log\left(\frac{\rho(y, p(\Theta)) + 1}{2}\right),$$

where Θ is the network weights, N is the number of pixels in the image, ρ the correlation of the image gradients, and λ is a weighting term. Optimal values for λ were found to be 0.1 for the distortion correction network and 2.0 for the decomposition network.

D. *Reconstruction and material decomposition*

Reconstruction was performed via filtered backprojection using the Feldkamp algorithm [18]. To reduce image noise, reconstruction was followed by joint bilateral filtration (BF) in which energy channels are jointly processed [19]. The reconstructions were converted to HU numbers. To decompose spectral PCCT data, we used a three materials post-reconstruction decomposition, subject to a non-negativity constraint similar to what has been previously described [1]. After decomposition the material maps were color coded with I in red, Au in green, and Gd in blue.

E. *Image quality metrics*

The bar phantom provides opportunity to measure resolution in reconstructions and material maps and compare to results from noiseless, distortion-free projections. The bar patterns were used to approximate the modulation transfer function (MTF) using the following equation:

$$MTF(a,b) = \frac{|a-b|}{a+b}$$

where a is the mean value from contrast-enhanced lines and b is the mean value between the lines for a given pattern. Noise was measured in slices of the phantom without contrast material.

III. Results

A. Distortion correction

Reconstructions via FDK from the spectrally distorted and CNN-corrected projections, each followed by bilateral filtration, are shown in Fig. 3. The distorted images have differences in contrast from the label image, particularly in the 25 and 34 keV energy bins. The higher energy thresholds are most affected by noise. For the corrected images, high frequency structures show the most residual error, however these images also show significantly reduced noise.

Noise and root mean squared error (RMSE) for the label, distorted, and corrected reconstructions are given in Table 1. The noise in the label images comes from FDK reconstruction with an ideal ramp filter. The projection-based distortion correction greatly reduces noise and errors in the reconstructed images. The resolution of all reconstructions is represented by the MTF (Fig. 4). The correction method appears to reduce the spatial resolution (from 3.37 to 2.85 lp/mm) in the reconstructed images.

B. Material decomposition

Material decompositions were created via inversion using the sensitivity matrix for both spectrally distorted and corrected reconstructions. These were compared to CNN decompositions (Fig. 5). There are marked differences in noise and artifacts between the maps generated from distorted data. The material maps from the distorted data contains high Au contamination in the water phantom. The CNN shows visual improvements in both noise and recovery of high frequency content.

TABLE I
IMAGE NOISE AND ERRORS

Noise [HU]	25 keV	34 keV	50 keV	80 keV
Label	5.9	6.2	7.2	7.7
Corrected	10.0	10.2	11.0	11.5
Distorted	30.8	35.2	45.9	116.8
RMSE [HU]	**25 keV**	**34 keV**	**50 keV**	**80 keV**
Corrected	14.4	14.4	14.9	16.8
Distorted	28.9	31.4	37.6	93.5

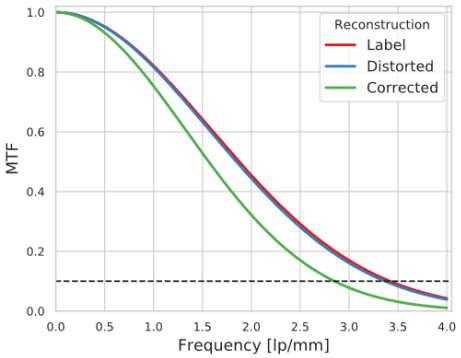


Fig. 4. MTF for the label, distorted, and distortion-corrected reconstructions.

Error metrics and material values showing the accuracy of each decomposition are shown in Fig. 6. Both spectral distortion correction and CNN-based material decompositions are shown to reduce the errors and improve the accuracy and precision of the resulting material decompositions. Paricularily, the CNN-based decomposition method reduces heterogeneity within homogenous structures (Fig. 6B).

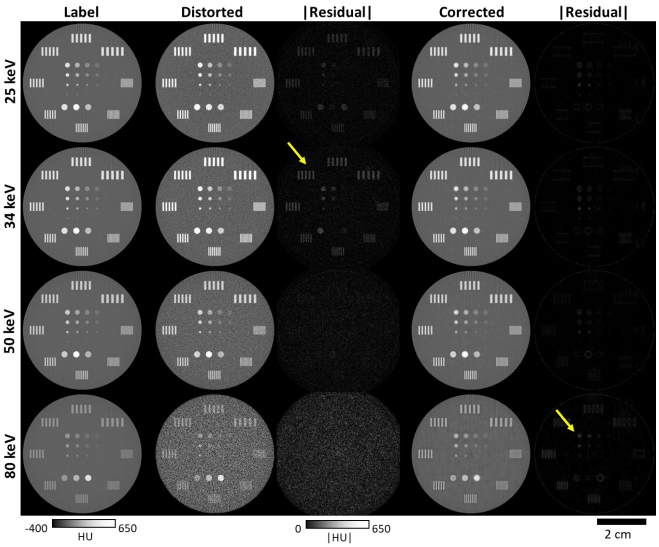


Fig. 3. Bar phantom reconstructions for all 4 energies with distorted and CNN-corrected projections for the iodine slice. Residual images show the difference with the label. The distorted reconstruction shows contrast differences in K-edge materials and the corrected images show errors in high frequencies and high energies (yellow errors).

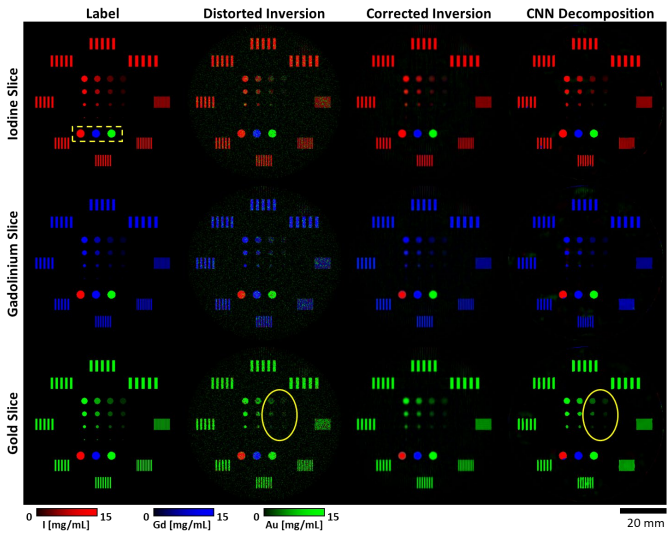


Fig. 5. Material maps for I (red), Gd (blue), and Au (green) as created by inversion before and after spectral distortion correction alongside the decomposition output by the trained CNN.

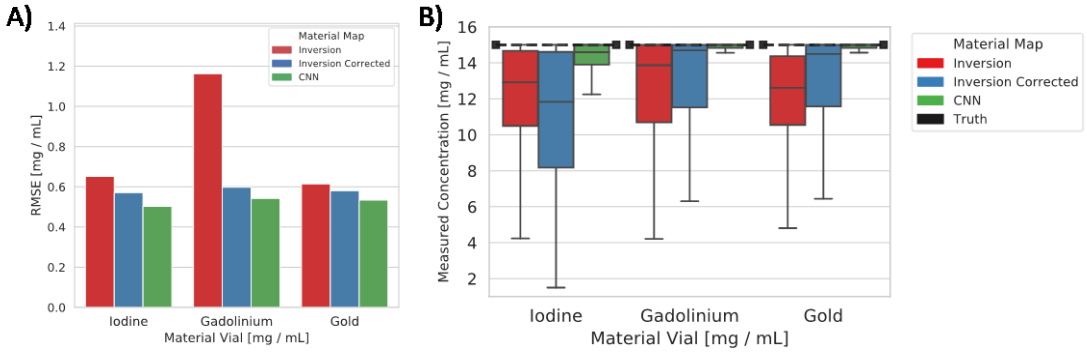


Fig. 6. Accuracy metrics for material decomposition. A) RMSE shows that distortion correction and CNN-based decomposition both decrease errors in recovering the concentrations of materials. B) Box plots of voxel values in material vials (Fig 5. yellow box) show the CNN best recovers the homogeneous nature of the vials, and thus is the most precise.

IV. Discussion

Spectral distortions are a significant problem for the widespread adoption of PCD technology. In this work, we have used simulations to examine a post-processing DL based correction method. We have also introduced a CNN-based method for a 3-material decomposition. This preliminary investigation showed promising results, lowering image noise, error, and increasing the accuracy and precision of resulting material maps. Unfortunately, this performance came with a small loss in resolution. We hypothesize that with slight alterations to the processing pipeline (e.g., processing distorted and corrected reconstructions jointly) resolution can be preserved. To become practical, this work must be replicated with experimental data. To obtain experimental results new training data would be required. This data could be generated using scanned and digital models of 3D printed phantoms with known contrast materials. Transfer learning from the simulated networks could result in models which function on experimental data.

References

[1] C. T. Badea, D. P. Clark, M. Holbrook, M. Srivastava, Y. Mowery, and K. B. Ghaghada, "Functional imaging of tumor vasculature using iodine and gadolinium-based nanoparticle contrast agents: a comparison of spectral micro-CT using energy integrating and photon counting detectors," *Phys Med Biol*, Feb. 2019, doi: 10.1088/1361-6560/ab03e2.

[2] D. P. Clark, M. Holbrook, C.-L. Lee, and C. T. Badea, "Photon-counting cine-cardiac CT in the mouse," *PLOS ONE*, vol. 14, no. 9, p. e0218417, Sep. 2019, doi: 10.1371/journal.pone.0218417.

[3] G. Wang, "A Perspective on Deep Imaging," *IEEE Access*, vol. 4, pp. 8914–8924, 2016, doi: 10.1109/ACCESS.2016.2624938.

[4] Y. Zhang and H. Yu, "Convolutional Neural Network Based Metal Artifact Reduction in X-ray Computed Tomography," *arXiv:1709.01581 [physics]*, Sep. 2017.

[5] J. Lee, H. Lee, and S. Cho, "Sinogram synthesis using convolutional-neural-network for sparsely view-sampled CT," in *Medical Imaging 2018: Image Processing*, 2018, vol. 10574, p. 105742A, doi: 10.1117/12.2293244.

[6] W. Cong and G. Wang, "Monochromatic CT Image Reconstruction from Current-Integrating Data via Deep Learning," *arXiv:1710.03784 [physics]*, Jan. 2018.

[7] H. Chen *et al.*, "Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017, doi: 10.1109/TMI.2017.2715284.

[8] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Med Phys*, vol. 40, no. 10, Oct. 2013, doi: 10.1118/1.4820371.

[9] J. P. Schlomka *et al.*, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys. Med. Biol.*, vol. 53, no. 15, p. 4031, 2008, doi: 10/dwrn84.

[10] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep Scatter Estimation (DSE): Accurate Real-Time Scatter Estimation for X-Ray CT Using a Deep Convolutional Neural Network," *J Nondestruct Eval*, vol. 37, no. 3, p. 57, Jul. 2018, doi: 10.1007/s10921-018-0507-z.

[11] K. C. Zimmerman and T. G. Schmidt, "Experimental comparison of empirical material decomposition methods for spectral CT," *Phys. Med. Biol.*, vol. 60, no. 8, pp. 3175–3191, Mar. 2015, doi: 10.1088/0031-9155/60/8/3175.

[12] M. Touch, D. P. Clark, W. Barber, and C. T. Badea, "A Neural Network-based Method for Spectral Distortion Correction in Photon Counting X-ray CT," *Phys Med Biol*, vol. 61, no. 16, pp. 6132–6153, Aug. 2016, doi: 10.1088/0031-9155/61/16/6132.

[13] D. P. Clark, M. Holbrook, and C. T. Badea, "Multi-energy CT decomposition using convolutional neural networks," in *Medical Imaging 2018: Physics of Medical Imaging*, 2018, vol. 10573, p. 105731O, doi: 10.1117/12.2293728.

[14] Siewerdsen J. H., Waese A. M., Moseley D. J., Richard S., and Jaffray D. A., "Spektr: A computational tool for x-ray spectral analysis and imaging system optimization," *Medical Physics*, vol. 31, no. 11, pp. 3057–3067, Oct. 2004, doi: 10/df8ftg.

[15] D. Clark and C. Badea, "GPU-Based Tools for Multi-Channel X-ray CT Reconstruction," 5th International Conference on Image Formation in X-Ray Computed Tomography; 2018; Salt Lake City.

[16] D. P. Clark and C. T. Badea, "Hybrid spectral CT reconstruction," *PLOS ONE*, vol. 12, no. 7, p. e0180324, Jul. 2017, doi: 10.1371/journal.pone.0180324.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv:1505.04597 [cs]*, May 2015.

[18] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A, JOSAA*, vol. 1, no. 6, pp. 612–619, Jun. 1984, doi: 10.1364/JOSAA.1.000612.

[19] D. P. Clark, C.-L. Lee, D. G. Kirsch, and C. T. Badea, "Spectrotemporal CT data acquisition and reconstruction at low dose," *Medical Physics*, vol. 42, no. 11, pp. 6317–6336, Nov. 2015, doi: 10.1118/1.4931407.

Reduction of Artifacts in Multi-Energy Imaging for a Stationary Gantry CT Scanner

Seongjin Yoon, Alexander Katsevich, Michael Frenkel, Ed Morton, and William Thompson

***Abstract*—Multi-Energy Stationary Gantry CT (ME-SGCT) is a powerful tool for security screening. However, the SGCT geometry leads to an anisotropic voxel illumination pattern, which varies significantly across the field of view (FOV) and results in irregular-view streaks. Multi-Energy CT (MECT) exacerbates this artifact compared with single energy CT. In this paper, we use MECT reconstruction based on the Photoelectric (PE) and Compton Scattering (CS) decomposition. We propose two algorithms to address the ME-SGCT streaks: (1) modified bilateral filtering, which uses the more stable CS coefficient to filter the less stable PE coefficient, and (2) data-filling based on interpolation. We test the methods using simulated data for complex phantoms. The results demonstrate significantly reduced irregular-view streaks, which allows for more accurate estimation of the effective atomic number Z_{eff} and density ρ that are important for security screening.**

***Index Terms*—Multi-Energy CT imaging, Stationary Gantry CT, image reconstruction, modified bilateral filter, nonuniform view sampling, streak artifact reduction**

I. Introduction

At present, carry-on baggage screening is done on the basis of projection X-ray imaging, and the process relies heavily on the operator performance. In an effort to improve screening, there are several developments underway to use X-ray CT for these purposes. To meet the processing time requirement for the security check point, a stationary gantry CT (SGCT), which uses a distributed source and stationary detector, has been sought to provide the fastest possible scanning time.

While SGCT enables faster scanning, the design complicates the development of analytic reconstruction algorithms: (1) view-direction derivative cannot be used due to the X-ray source not following a well-defined trajectory, and (2) the illumination pattern of any voxel is highly anisotropic and varies significantly across the field of view. To address those issues, we previously developed a No View-Differentiation (NVD) type algorithm that incorporates a weight function. The latter is selected based on the illumination pattern to reduce streaks that arise due to irregular view sampling [1].

S. Yoon, A. Katsevich, and M. Frenkel are with iTomography Corp., Texas Medical Center Innovation Institute, 2450 Holcombe Blvd, Houston, TX 77021, USA (e-mail: syoon@iTomography.com, alexander@iTomography.com, michael@iTomography.com).

A. Katsevich is also with the University of Central Florida, Mathematics Department, Orlando, FL 32816, USA.

A. Katsevich is a shareholder and chief technology officer of iTomography Corporation and, as such, may benefit financially as a result of the outcomes of the work reported in this publication.

E. Morton and W. Thompson are with Rapiscan Systems, 2805 Columbia St, Torrance, CA 90503, USA (e-mail: emorton@rapiscansystems.com, withompson@rapiscansystems.com).

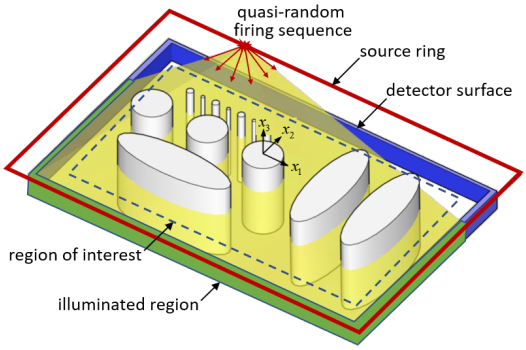


Fig. 1: Conceptual drawing of the SGCT scanner.

In an effort to improve threat detection even further, the SGCT system can incorporate a photon-counting detector with multiple energy bins to enable the scanner to capture spectral information. The Photoelectric (PE) and Compton Scattering (CS) decomposition method has been widely used as a foundation of multi-energy CT (MECT) reconstruction [2], [3]. In principle, PE-CS decomposition scheme can be directly used for MECT reconstruction from the spectral data collected by the SGCT scanner. However, the high dynamic range of the PE coefficient combined with inherent irregular-view streak artifacts of SGCT make it harder to use the reconstructed images for computing the physical properties of the scanned objects such as Z_{eff} and ρ. While the previously proposed weighted NVD reconstruction algorithm reduces irregular-view streak artifacts significantly, the PE-CS decomposition process can amplify the remaining streaks. The amplitude of the streaks in the PE coefficient can even exceed the amplitude of the original image, especially when metals are present in the field of view (FOV).

Here, we propose two methods to reduce the streak artifact in ME-SGCT: (1) modified bilateral filtering, which uses the more stable (smaller dynamic range) CS coefficient to filter the less stable (greater dynamic range) PE coefficient, and (2) data-filling based on interpolation. We test the proposed methods using simulated data. The tests show that the methods reduce the streaks significantly and enable one to more accurately estimate Z_{eff} and ρ pointwise in the FOV.

II. Basis Decomposition

Basis decomposition models for MECT have been widely used to simplify the reconstruction. The PE-CS decomposition was developed by Alvarez and Macovski [2] based on two major physical processes involved in photon attenuation at the

range of medium energies (30-200 keV). The PE-CS model is given by the equation

$$\mu(E) = c_{\mathrm{p}} f_{\mathrm{PE}}(E) + c_{\mathrm{s}} f_{\mathrm{KN}}(E), \tag{1}$$

where E is the energy level, $f_{\mathrm{PE}} = 1/E^m$, f_{KN} is the Klein-Nishina function, and c_{p} and c_{s} are the PE and CS coefficients, respectively. The benefit of the PE-CS decomposition is that there exist semi-empirical formulas relating ρ, Z_{eff} with $c_{\mathrm{p}}, c_{\mathrm{s}}$:

$$c_{\mathrm{p}} = K_1 \frac{Z}{A} \rho Z^{n-1}, c_{\mathrm{s}} = K_2 \frac{Z}{A} \rho. \tag{2}$$

where K_1, K_2, and n need to be obtained empirically, assuming they are constant within the given energy range. The ratio Z/A is assumed to be 1/2. For this work, we use the parameters from Paziresh *et al.* [3]: $K_1 = 13.96$, $K_2 = 0.30$, and $n = 4.2$.

Since the noise in photon counting data is known to be Poisson distributed, we use Poisson likelihood to compute the estimates of C_{p} and C_{s}, which are the line integrals of c_{p} and c_{s}, respectively. The corresponding cost functional is given by:

$$\left[\widehat{C}_{\mathrm{p}}, \widehat{C}_{\mathrm{s}}\right] = \underset{C_{\mathrm{p}}, C_{\mathrm{s}}}{\arg\min} \sum_{l=1}^{N_l} h_l(\bar{y}_l(C_{\mathrm{p}}, C_{\mathrm{s}})), \tag{3}$$

where N_l is the number of detector energy bins. The Poisson likelihood function h_l and the estimated mean count $\bar{y}_l$ are defined as follows:

$$h_l(t) = t - y_l \log t, \tag{4}$$

$$\bar{y}_l(C_{\mathrm{p}}, C_{\mathrm{s}}) = \int_{E_l}^{E_{l+1}} I_l(E) \exp(-C_{\mathrm{p}} f_{\mathrm{PE}}(E) - C_{\mathrm{s}} f_{\mathrm{KN}}(E)) \mathrm{d}E, \tag{5}$$

where I_l is the energy spectrum within the l-th energy bin.

Once we get $\widehat{C}_p$ and $\widehat{C}_s$ from (3), we apply the view-density weighted NVD algorithm described in our previous work [1], to separately reconstruct c_p and c_s.

III. Image Domain Sparse-View/Noise Streak Reduction Filtering

Due to irregular view sampling, sparse-view streaks in ME-SGCT reconstructions are distributed irregularly across the FOV. Any conventional filter with localized support cannot effectively deal with such high amplitude irregularly distributed artifacts. To suppress the artifact while keeping the algorithm numerically simple and efficient, we introduce a *modified bilateral (MBL) filter guided by a reference image*:

$$\mathrm{MBL}[f(i)] = \sum_{j \in \Omega_{(i,n_{\mathrm{B}})}} \exp\left(\frac{-(g(i) - g(j))^2}{\sigma \max(\mathbf{g})^2 + \delta}\right) f(j), \tag{6}$$

where f is the image being filtered, $\Omega_{(i,n_{\mathrm{B}})}$ is the set of neighboring pixels in the square region centered at the i-th pixel with size $(2n_{\mathrm{B}} + 1) \times (2n_{\mathrm{B}} + 1)$, and σ and δ are tunable parameters. The latter also prevents dividing by zero. The spatial weight of the filter is constant within the square region. The Gaussian kernel is normalized by the maximum value of the reference image for the following two reasons: (1) to account for the dynamic range difference between c_{p} and c_{s}, (2) the most significant sparse-view streak component is proportional to the maximum value of each image. g is the reference image which contains information about the same object, but has a smaller dynamic range and less sparse-view streaks. This approach belongs in a group of multi-channel image denosing, see e.g. [4] and references therein.

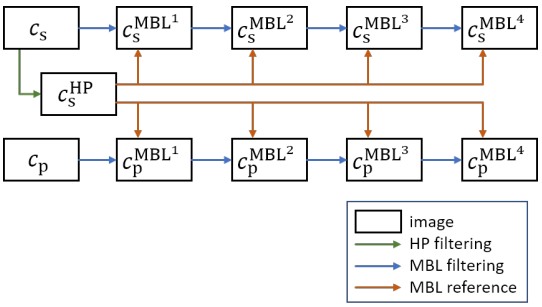


Fig. 2: Illustration of the image filtering workflow.

To obtain the reference image g, we first filter the reconstructed CS image c_{s} using edge-preserving hyperbolic potential (HP) regularization: $c_{\mathrm{s}} \rightarrow c_{\mathrm{s}}^{\mathrm{HP}}$. Since streak intensity is much lower in c_{s} than in c_{p}, the localized HP filtering can produce a good quality image $c_{\mathrm{s}}^{\mathrm{HP}}$ with most of the streaks removed. Instead of using $c_{\mathrm{s}}^{\mathrm{HP}}$ directly for subsequent processing, we use $g = c_{\mathrm{s}}^{\mathrm{HP}}$ as a reference image for MBL filtering of c_{s} to produce a new image $c_{\mathrm{s}}^{\mathrm{MBL}}$, then we apply the MBL filter to c_{p} using the same $c_{\mathrm{s}}^{\mathrm{HP}}$ as the reference image.

For both c_{p} and c_{s}, we apply the MBL filter four times with $c_{\mathrm{s}}^{\mathrm{HP}}$ as the reference image. See Fig. 2 for the diagram that illustrates the proposed image filtering workflow. Here, f^{MBL^n} represents the image f filtered by MBL n times. The reason why we use $c_{\mathrm{s}}^{\mathrm{HP}}$ for MBL filtering of c_{p} (even though $c_{\mathrm{s}}^{\mathrm{MBL}^4}$ is more accurate) is to have the same amount of blurring in $c_{\mathrm{s}}^{\mathrm{MBL}^4}$ and $c_{\mathrm{p}}^{\mathrm{MBL}^4}$. This leads to more accurate pointwise ratios $c_{\mathrm{p}}/c_{\mathrm{s}} \approx c_{\mathrm{p}}^{\mathrm{MBL}^4}/c_{\mathrm{s}}^{\mathrm{MBL}^4}$ when computing Z_{eff}.

While the MBL filter effectively removes high amplitude streaks, it comes with one caveat. For accurate reconstruction of a region with constant Z_{eff} and density, the size of the region should be large enough to contain sufficiently many streaks. In this case the streaks are reduced due to averaging. For voxels farther away from a high-Z material that generates streaks, it gets harder to remove sparse-view streaks because the frequency of the streaks is decreasing. To address this issue, we fill in some of the missing data by interpolating available data. This procedure is explained in the next section.

IV. Data Filling Using Interpolation

Adding to the MBL filtering described in Sec. III, we create synthetic data by interpolating two projection data with the same source-detector position (but the object is shifted) during the back-projection stage to reduce the streak artifact further.

Due to the rectangular geometry of the detector, we first project the data onto a virtual flat detector to use the conventional FDK-type backprojection. This backprojection process combines the horizontal axis (along the detector rows) of the virtual detector and the vertical axis (along the columns) of the actual detector to parametrize the data to be backprojected.

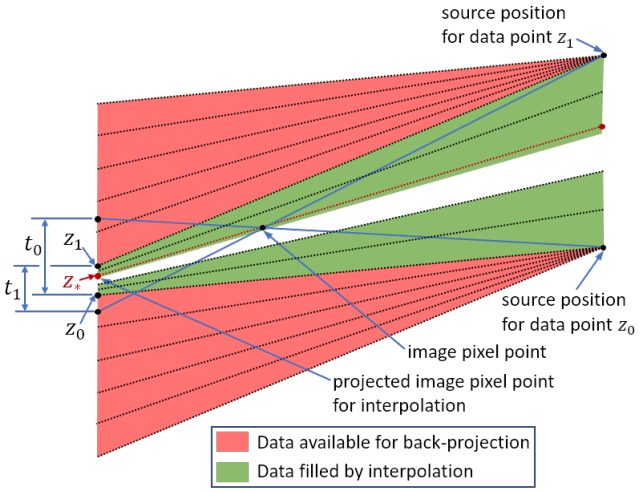


Fig. 3: Illustration of the vertical cross section of the areas illuminated by two vertically-shifted source positions. Image pixels lying on the same dotted line share the same z_* value. Red region shows where the projection data is originally available, and green region shows where the synthetic projection data is filled by interpolation.

For illustration purposes, let us assume that the object is stationary, and the source-detector gantry is shifting vertically. Since we interpolate two (filtered) cone beam projections with the vertically-shifted source-detector positions, the horizontal coordinate of the interpolated data is still the same in both cases. Only the vertical coordinate of one illuminated region is shifted relative to the other. The illuminated regions for the two projections do not overlap, and there is a gap between the two illuminated regions where the projection data are not available. Fig. 3 shows the schematic vertical cross section of the two illuminated regions and the gap between them.

Two data points (z_0 and z_1 in Fig. 3) to be used for interpolation do not share the same mapping onto the detector vertical coordinate, as the mapping depends on the vertical source positions. Thus, the interpolated projection data does not belong to any of those two mapping systems. To solve this issue, we define a set of virtual source positions between the two actual source locations, that enables the projected vertical coordinate of the data point to always be between the top of the bottom projection data region and the bottom of the top projection data region. Additionally, to avoid estimating values too far from the available data, we apply interpolation only to image pixels if their projected vertical coordinate is within a threshold distance of z_e from either the top source position or the bottom source position (see the green regions in Fig. 3). Consequently, the following value is used in the interpolation algorithm to identify the location of the interpolated data point:

$$z_* = \frac{z_0 t_1^{n_\mathrm{I}} + z_1 t_0^{n_\mathrm{I}}}{t_0^{n_\mathrm{I}} + t_1^{n_\mathrm{I}}},$$
$$z_0 < z_* \le z_0 + z_e \text{ or } z_1 - z_e \le z_* < z_1. \qquad (7)$$

where t_0 is the vertical distance between the reference data point z_0 and the shadow of the image point with the corresponding source position, t_1 is the equivalent of t_0 with the reference data point being z_1.

Boundary detector pixels are more vulnerable to high amplitude noise as they have fewer neighboring pixels that can be used for denoising. Nearest neighbor interpolation makes the image more crisp, but it can create discontinuity and amplify noise. Linear interpolation attenuates noise, but it reduces spatial resolution. The factor $n_\mathrm{I} > 0$ provides the flexibility to transform the interpolator from linear to nearest neighbor, i.e., linear interpolation if $n_\mathrm{I} = 1$, and nearest neighbor if $n_\mathrm{I} \gg 1$.

V. Numerical Experiments

We simulate noise-free sinogram data for a 3D phantom consisting of water, aluminum, and titanium ellipsoids (see Fig. 4), with poly-energetic sources fired in a periodic quasi-random sequence.

Fig. 5 shows the estimated cofficients c_p and c_s without any streak artifact removal applied. Note that the c_s image has significantly fewer streaks than the c_p image. Fig. 6 shows the corresponding coefficients after data-filling and MBL filter are applied. Without artifact correction, the intensity of streaks in the c_p image is much greater than the value of c_p for water, thus making the estimation of ρ and Z_eff almost impossible. The proposed streak removal workflow enables one to estimate ρ and Z_eff based on (2). Figs. 7 and 8 show the profiles of ρ and Z_eff in four different locations marked by colored lines in Fig. 4. While the estimated ρ deviates from the ground truth, the estimated Z_eff matches well with the ground truth. The error in the density estimation is mostly caused by two factors: (1) cone-beam error of the FDK-type reconstruction, (2) approximation error of $Z/A = 0.5$. Fortunately, estimation of the effective atomic number is less vulnerable to those two errors because they affect both the c_p and c_s equally. Dividing c_p by c_s compensates the error. Also, the error in density due to the Z/A approximation is systematic and should not affect material identification since this can be calibrated.

VI. Conclusions

We develop a workflow for robust MECT image reconstruction for a stationary gantry scanner. The presented numerical experiments demonstrate that the proposed algorithms allow one to efficiently reduce streak artifacts while preserving the accuracy of ME-SGCT reconstruction.

References

[1] A. Katsevich, S. Yoon, M. Frenkel, E. Morton, and W. Thompson, "Reduction of irregular view-sampling artifacts in a stationary gantry ct scanner," in 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, vol. 11072. International Society for Optics and Photonics, 2019, p. 110720G.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerized tomography." Physics in medicine and biology, vol. 21, no. 5, pp. 733–744, 1976.

[3] M. Paziresh, A. M. Kingston, S. J. Latham, W. K. Fullagar, and G. M. Myers, "Tomography of atomic number and density of materials using dual-energy imaging and the Alvarez and Macovski attenuation model," Journal of Applied Physics, vol. 119, no. 21, 2016.

[4] M. Salehjahromi, Y. Zhang, and H. Yu, "Iterative spectral CT reconstruction based on low rank and average-image-incorporated BM3D," Physics in Medicine and Biology, vol. 63, 2018, article id 155021.

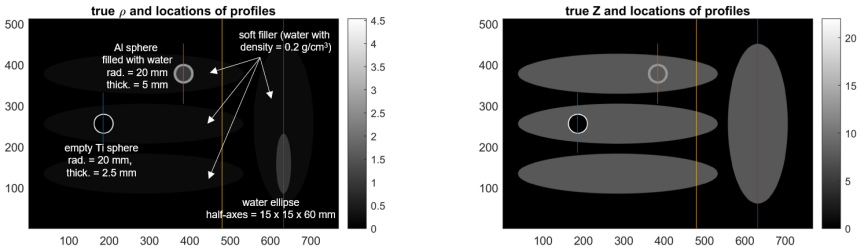


Fig. 4: Ground truth phantom data, density ρ (left) and effective atomic number Z_{eff} (right). Lines indicate the locations of profiles shown in Figs. 7 and 8.

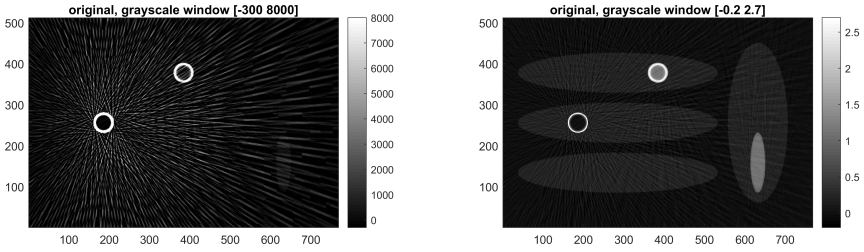


Fig. 5: Estimated c_{p} (left) and c_{s} (right), without data-filling or MBL filtering.

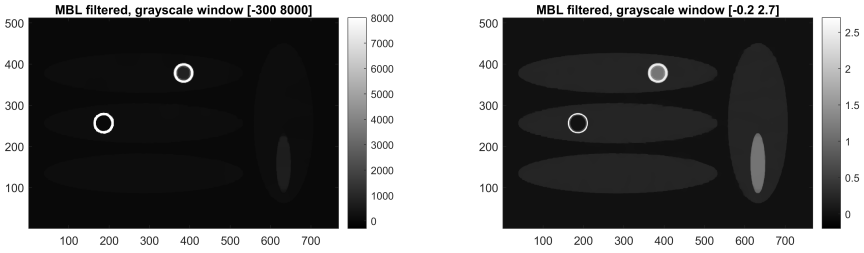


Fig. 6: Estimated c_{p} (left) and c_{s} (right), after data-filling and MBL filtering.

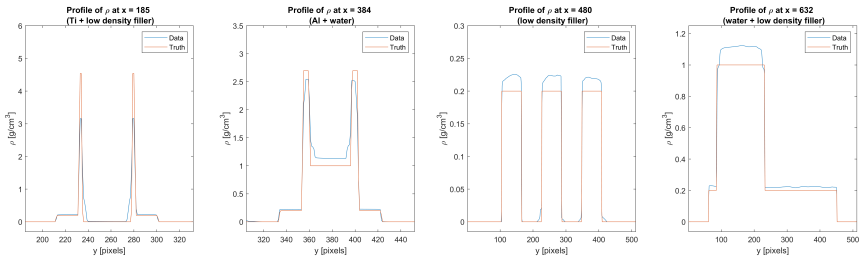


Fig. 7: Profiles of estimated density after data-filling and MBL filtering. Locations of the sampled profiles are shown in Fig. 4.

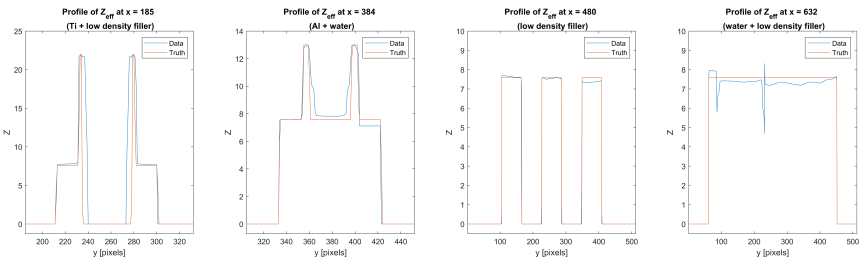


Fig. 8: Profiles of estimated effective atomic number after data-filling and MBL filtering. Locations of the sampled profiles are shown in Fig. 4.

Flux Adaptation for Task-based Dose Optimization in Grating-based X-ray Imaging

Chengpeng Wu, Hewei Gao, Yuxiang Xing, Zhiqiang Chen, Xiaohua Zhu, Li Zhang

***Abstract*—For grating-based x-ray imaging (GBXI), the phase-stepping strategy is still the most commonly used approach to retrieve multiple contrasts such as the differential phase contrast and the dark-field contrast, while it's also the main reason for the high dose problem of GBXI. With the technology developed from bench to bedside, there is increasing interest for dose reduction recently. In this work, we challenge the typical constant flux distribution in the phase-stepping process at the first time, and try to optimize adaptive flux distributions for dose reduction. After some tedious derivations, we propose a flux adaptation strategy for task-based dose optimization, which is applicable to several recently proposed analytic information retrieval algorithms and multiple contrasts. To validate our analyses, simulated and real experiments are conducted, and the results demonstrate the effectiveness of our proposed method.**

***Index Terms*—grating-based x-ray imaging; phase contrast; phase-stepping; noise.**

I. Introduction

GRATING-BASED x-ray imaging can obtain multiple complementary contrasts besides the traditional attenuation contrast (ATC), such as the differential phase contrast (DPC) and the dark-field contrast (DFC). Compared with the ATC, DPC has much higher sensitivity for light materials like soft tissues [1], and DFC can get access to the reciprocal space information about the sample, i.e. the small-angle x-ray scattering (SAXS) distribution [2]. At present, GBXI is going from bench to bedside, and there are many ongoing clinical trials based on GBXI worldwide, such as the first human lung dark-field imaging prototype [3] and some mammography prototypes [4], [5].

For clinical applications, the absorbed dose must be within the safety range. GBXI suffers from rickets by the commonly used phase-stepping (PS) approach, and it needs multiple exposures at a scanning angle. At first, the PS approach was used in visible-light interferometry [6], and it was adapted to the earliest GBXI system (i.e., the Talbot interferometry [7]) by Weitkamp et al. [8]. Using the PS approach, we can acquire a so-called phase-stepping curve (PSC) at each detector pixel. From both PSCs with and without the sample, multiple contrasts can be retrieved by an information retrieval algorithm.

In the PS approach, it's routine to adopt a constant flux at all steps. To the best of our knowledge, it's almost blank about studies on non-uniform flux effects in the PS approach. Inspired by one of our recently proposed information retrieval algorithms [9], we found that the raw data will be weighted across steps in the post-processing information retrieval, which means that we can also weight the exposure flux distribution directly. In this paper, we overturn the traditional flux paradigm in the PS process, and try to change the exposure flux distribution across steps. Among all non-uniform distributions, we propose a flux adaptation method for task-based dose optimization in GBXI, which can significantly reduce the dose while keeping almost the same image quality as that of the typical flux acquisition scheme. The "task-based" qualifier means that the proposed optimization should be designed according to the targeted contrast (DPC or DFC) and the pre-selected information retrieval algorithm.

This work was supported in part by the National Natural Science Foundation of China (No. 81771829). (Corresponding author: Li Zhang).

C. Wu, H. Gao, Y. Xing, Z. Chen, X. Zhu and L. Zhang are with the Department of Engineering Physics, Tsinghua University, Beijing, China, and also with the Key Laboratory of Particle and Radiation Imaging (Tsinghua University), Ministry of Education, China (e-mail: zli@mail.tsinghua.edu.cn).

II. Method

A. The Information Retrieval Algorithms

The information retrieval is one of the most important processes to obtain desired contrasts, and therefore, it's actively investigated in GBXI. So far, there are mainly two types of algorithms. The first one is the most classic Fourier component analysis (FCA), which retrieves typical contrasts by computing harmonic coefficients of both the flat PSC and the sample PSC. FCA is simple to be understood and efficient to be implemented, but has the strict cosinusoidal assumption for both PSCs and the inherent phase-wrapping problem. The other one is the recently developed multi-order moment analysis (MMA). Totally different from FCA, MMA regards typical contrasts as moments of the underlying SAXS distribution [10]. Based on the well-established convolution relationship between PSCs and SAXS distributions [11], MMA obtains the SAXS distribution by a deconvolution algorithm, and then retrieves multiple contrasts by computing moments of the SAXS distribution. Compared with the FCA, MMA can obtain additional higher moments useful in some applications [10], [12]. Simultaneously, MMA discards the cosinusoidal assumption and avoids the annoying phase-wrapping problem naturally. However, the critical deconvolution procedure in MMA is usually iterative like the Lucy-Richardson deconvolution and therefore, is much more time-consumed and difficult to determine a proper stopping criteria for the iteration.

To overcome these limitations of the original deconvolution-based MMA (DB-MMA), two analytic MMA methods are independently developed recently [13], [14]. One is referred as the direct MMA (D-MMA) [13] that is applicable to edge-illumination and analyzer-based imaging at first, and it's

extended to GBXI by our group with some pre-processing [15]. Then a generalized D-MMA (GD-MMA) is developed by using the trigonometric orthogonality of PSCs [9], which can be optimized to improve noise performances of the original D-MMA. Based on Fourier properties, the other one is proposed by our group [14], referred as the truncated analytic MMA (TA-MMA). Compared with DB-MMA, TA-MMA not only has higher computation efficiency similar with D-MMA and GD-MMA, but also establishes a quantitative connection between FCA and MMA, which means the first-order moment computed by MMA is essentially the product of DPC and DFC retrieved by FCA.

From the analytic expressions in GD-MMA, it's found that raw PSCs will be weighted for subsequent calculations, which inspires us to weight the flux distribution across steps to acquire pre-designed weighted PSCs directly. Thus in the following section, we try to explore the flux adaptation in the PS approach at the first time, and then optimize the adaptive flux distributions according to different contrasts and information retrieval algorithms.

B. Flux Adaptation for Task-based Dose Optimization

In D-MMA [13] or GD-MMA [9], the complementary contrasts besides ATC, such as DPC and DFC in the presence of noise, can be uniformly expressed as:

$$C = \frac{\int w(\phi)\left[\bar{s}(\phi) + n_s(\phi)\right] d\phi}{\int \left[\bar{s}(\phi) + n_s(\phi)\right] d\phi} - \frac{\int w(\phi)\bar{f}(\phi) d\phi}{\int \bar{f}(\phi) d\phi}, \quad (1)$$

where, C denotes a target contrast and ϕ denotes the phase term of PSCs; $\bar{s}(\phi)$ and $\bar{f}(\phi)$ denote the noise-free sample PSC and flat PSC correspondingly acquired with high fluxes; $n_s(\phi)$ denotes the statistical noises in the sample PSC; while the flat PSC can be considered as noise-free acquired under high fluxes without any actual doses for the scanning sample; $w(\phi)$ denotes the weighted function, for example, $w(\phi) = \phi$ for the DPC in D-MMA.

As described in Section I, the optimization goal is to reduce the total exposure dose in the data acquisition and keep the similar image quality of a target contrast, which will be evaluated by the noise performance in this work. In turn, the equivalent goal is to minimize the noise (i.e., the variances of the contrast in background areas) among various flux distributions with the same total exposure dose.

The acquired sample PSC is proportional to the normalized flux distribution, i.e.,

$$\bar{s}(\phi) = \bar{s}_0(\phi) \cdot t(\phi), \quad (2)$$

where, $\bar{s}_0(\phi)$ denotes the original sample PSC under the constant flux, and $t(\phi)$ denotes the normalized flux distribution. As known to us, the flux at every step can be adjusted by controlling the exposure time or the tube current, both of which are proportional to the flux.

In practice, the PS process is discrete, and the total number of steps can be assumed as N. Thus, the equivalent optimization problem can be written mathematically as:

$$\text{minimize}_t \text{Var}\left(\frac{\sum_{n=1}^{N} w'(n)\left[\bar{s}_0(n)t(n) + n_s(n)\right]}{\sum_{n=1}^{N}\left[\bar{s}_0(n)t(n) + n_s(n)\right]}\right) \quad (3)$$

$$\text{s.t.} \sum_{n=1}^{N} t(n) = N, \quad (4)$$

where, $w'(n) = w(n)/t(n)$ is the adjusted weighted function to keep the average value of the target contrast consistent with the original value C.

As demonstrated in numerous literatures [1], [7], the original PSC $s_0(n)$ acquired with the constant flux distribution can be expressed with a cosinusoidal model, i.e.,

$$\bar{s}_0(n) = a_0 + a_1 \cos\left(2\pi \frac{n}{N} + \phi_1\right). \quad (5)$$

For real GBXI systems, it's reasonable to assume that $a_0 \gg a_1$, and therefore, the denominator in Eq. (3) can be also regarded as a constant, i.e.,

$$\sum_{n=1}^{N}\left[\bar{s}_0(n)t(n) + n_s(n)\right] \approx N a_0. \quad (6)$$

Substituting Eq. (6) into Eq. (3), the object function in Eq. (3) can be simplified as:

$$L(t) = \text{Var}\left(\sum_{n=1}^{N} w'(n)\left[\bar{s}_0(n)t(n) + n_s(n)\right]\right). \quad (7)$$

From our definitions, the first term about $\bar{s}_0(n)$ in Eq. (7) is noise-free, which means its variance can be omitted. Moreover, it's generally recognized that the raw detector data $s(n)$ obeys the Poisson distribution, and therefore, the second term about the noise $n_s(n)$ can be calculated as:

$$\text{Var}\left(\sum_{n=1}^{N} w'(n) n_s(n)\right) = \sum_{n=1}^{N} w'^2(n)\bar{s}_0(n)t(n) \quad (8)$$

$$= \sum_{n=1}^{N} w^2(n)\frac{\bar{s}_0(n)}{t(n)}. \quad (9)$$

With these conditions, the original optimization problem defined in Eqs. (3)(4) can be expressed as:

$$\text{minimize}_t \sum_{n=1}^{N} w^2(n)\frac{\bar{s}_0(n)}{t(n)} \quad (10)$$

$$\text{s.t.} \sum_{n=1}^{N} t(n) = N, \quad (11)$$

Next, we will solve the problem with the classic Lagrange multiplier method. Firstly the Lagrange function can be written as:

$$L(t, \lambda) = \sum_{n=1}^{N} w^2(n)\frac{\bar{s}_0(n)}{t(n)} + \lambda\left(\sum_{n=1}^{N} t(n) - N\right), \quad (12)$$

where λ is a constant. Then let's set the partial derivatives of the Lagrange function to $t(n)$ and λ as zero, i.e.,

$$\frac{\partial L}{\partial t(n)} = -\frac{w^2(n)\bar{s}_0(n)}{t^2(n)} + \lambda = 0, \tag{13}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{n=1}^{N} t(n) - N = 0. \tag{14}$$

From Eq. (13), it can be seen that the optimal adaptive flux distribution $t^*(n)$ is proportional to the absolute value of weighted function $w(n)$ multiplied with the squared original sample PSC $\bar{s}_0(n)$, i.e.,

$$t^*(n) = \sqrt{\frac{w^2(n)\bar{s}_0(n)}{\lambda}} \propto |w(n)|\sqrt{\bar{s}_0(n)}. \tag{15}$$

Substituting Eq. (15) to Eq. (14), the optimal adaptive flux distribution can be expressed as:

$$t^*(n) = |w(n)|\sqrt{\bar{s}_0(n)} \cdot \frac{N}{\sum_{n=1}^{N} |w(n)|\sqrt{\bar{s}_0(n)}}. \tag{16}$$

Eq. (16) is the ideal optimal solution of the original problem in Eqs. (3)(4). However, the sample PSC $\bar{s}_0(n)$ with the constant flux can not be obtained in advance sometimes. Combined with the assumption $a_0 \gg a_1$ described above, it's reasonable to regard $s_0(n)$ as a constant a_0, and therefore, the optimal adaptive flux distribution can be approximated as:

$$t^*(n) \approx \frac{|w(n)| \cdot N}{\sum_{n=1}^{N} |w(n)|}. \tag{17}$$

According to Eq. (17), we can calculate the optimal adaptive flux distributions for D-MMA and GD-MMA, respectively, which are shown in Fig. 1. To check the effectiveness of Eq. (17), we substitute it to the object function in Eq. (10), and the calculated function values are listed in Table. I. From Table. I, it can be seen that the nominal noise variances can be reduced more than 20% by our adaptive flux distributions. From the derivations above, the proposed flux adaptation method can be applied to retrieve DPC and DFC by D-MMA, GD-MMA and TA-MMA directly.

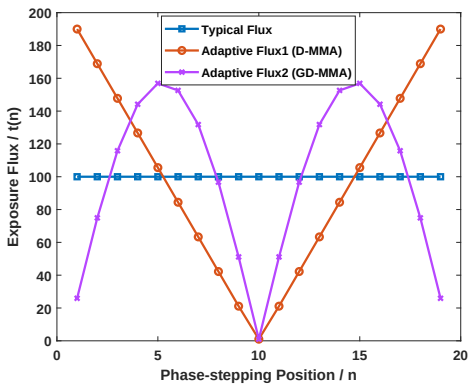


Fig. 1. The typical constant flux distribution and our proposed optimal adaptive ones in Eq. (17) for retrieving the DPC by D-MMA and GD-MMA correspondingly. The number of steps in phase-stepping is $N = 19$.

TABLE I
THE OBJECT FUNCTION VALUES IN EQ. (10) CALCULATED BY D-MMA AND GD-MMA UNDER THREE FLUX DISTRIBUTIONS IN FIG. 1.

PS Steps	Algorithm	Typical	Adaptive1	Adaptive2	Reduction
N=9	D-MMA	29.24	**21.66**	38.57	25.92%
	GD-MMA	18.00	20.93	**14.29**	20.61%
N=19	D-MMA	62.33	**46.62**	110.36	25.20%
	GD-MMA	38.00	45.84	**30.66**	19.32%

C. Validation System Setups

In order to validate our analyses, we compared the DPCs retrieved by D-MMA and GD-MMA on both simulated data and experimental data, whose detailed parameters can be found in Ref. [9]. For evaluating noise performances, we add various Poisson noises to the original simulated data. Due to the limitations of experimental conditions, we indirectly obtain the raw PSCs under adaptive flux distributions by multiplying original PSCs with our proposed adaptive flux distributions for experimental data. The noise levels in original PSCs are very low, and therefore, we also insert Poisson noise to these experimental data for evaluating noise performances.

III. RESULTS AND DISCUSSION

As shown in Fig. 2, both the simulated DPC profiles of a PMMA cylinder retrieved by D-MMA and GD-MMA under our proposed adaptive flux distributions have lower levels of noises compared with that under the typical constant flux distribution. As shown in Fig. 3, the noise reduction ratios by using the proposed adaptive flux distribution are almost constant for various levels of noises.

The retrieved DPC images of experimental data by D-MMA are shown in Fig. 4, and it's clear to found the flux adaptation strategy can reduce the noises in background areas. As shown in Table. II, the reduction ratios are around 20% (18.96% for D-MMA and 26.65% for GD-MMA), which are consistent with our theoretical expectations in Table. I.

IV. CONCLUSION

In this work, we propose a flux adaptation method for task-based dose optimization in GBXI. The simulated and experimental results validate our theory, and the noise can be reduced more than 20% under the same total dose, which means the proposed method can reduce more than 20% dose while keeping the same noise performance. This work explores a new direction for dose reduction of GBXI. Future work includes more studies on the GBXI system using the proposed flux adaptation approach and promoting the similar flux adaptation on other x-ray phase-contrast imaging systems.

TABLE II
THE EXPERIMENTAL NOISE VARIANCES (10^{-6}) RETRIEVED BY D-MMA AND GD-MMA UNDER THREE FLUX DISTRIBUTIONS IN FIG. 1.

Algorithm	Index	Typical	Adaptive1	Adaptive2	Reduction
D-MMA	Mean	90.17	**73.07**	106.05	18.96%
	Std	1.14	**0.75**	1.12	
GD-MMA	Mean	61.12	73.19	**44.83**	26.65%
	Std	0.74	0.69	**0.46**	

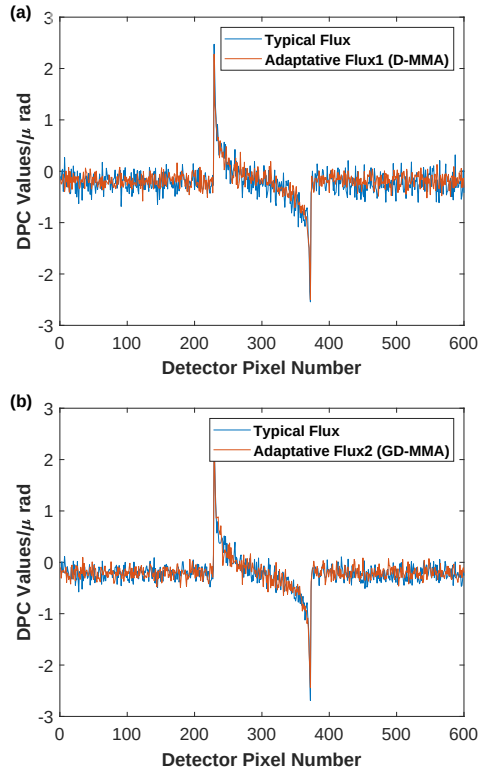


Fig. 2. The simulated DPC profiles of a PMMA cylinder retrieved by D-MMA (a) and GD-MMA (b) under the typical flux distribution and our proposed adaptive flux distributions.

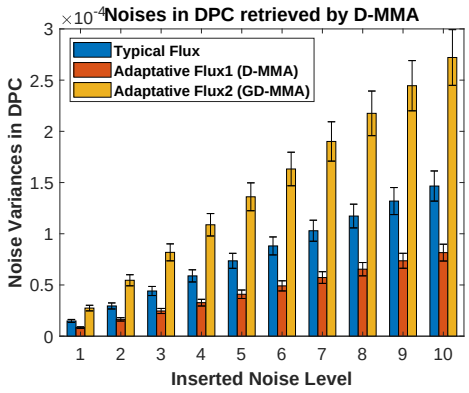


Fig. 3. The simulated noise variances in DPC retrieved by D-MMA under three mentioned flux distributions and various inserted Poisson noise levels.

REFERENCES

[1] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, "Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources," *Nature Physics*, vol. 2, no. 4, p. 258, 2006.

[2] F. Pfeiffer *et al.*, "Hard-x-ray dark-field imaging using a grating interferometer," *Nature Materials*, vol. 7, no. 2, p. 134, 2008.

[3] K. Willer *et al.*, "X-ray dark-field imaging of the human lunga feasibility study on a deceased body," *PloS One*, vol. 13, no. 9, p. e0204565, 2018.

[4] C. Arboleda *et al.*, "Towards clinical grating-interferometry mammography," *European Radiology*, pp. 1–7, 2019.

[5] R. Zhang, K. Li, J. W. Garrett, and G.-H. Chen, "Human-compatible multi-contrast mammographic prototype system," in *Medical Imaging 2019: Physics of Medical Imaging*, vol. 10948. International Society for Optics and Photonics, 2019, p. 109480X.

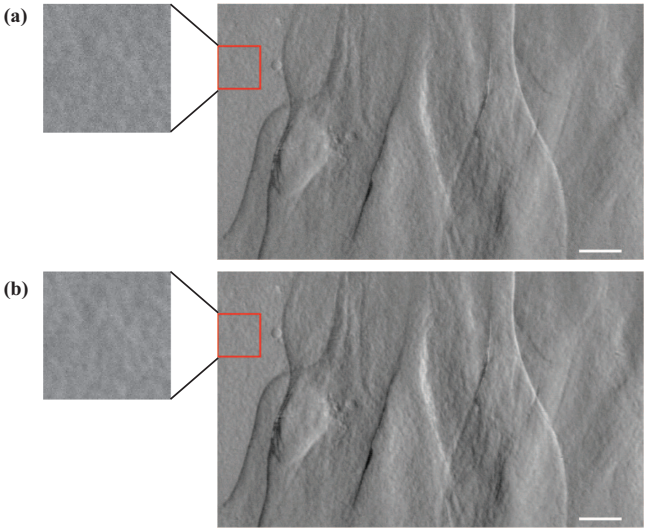


Fig. 4. The experimental DPC images of a breast specimen retrieved by D-MMA under the typical flux distribution (a) and our proposed adaptive flux distribution (b), and the zoomed background area as shown in the upper left corner. The scale bar is 0.65 mm.

[6] K. Creath, "V phase-measurement interferometry techniques," in *Progress in optics*. Elsevier, 1988, vol. 26, pp. 349–393.

[7] A. Momose, S. Kawamoto, I. Koyama, Y. Hamaishi, K. Takai, and Y. Suzuki, "Demonstration of x-ray talbot interferometry," *Japanese Journal of Applied Physics*, vol. 42, no. 7B, p. L866, 2003.

[8] T. Weitkamp *et al.*, "X-ray phase imaging with a grating interferometer," *Optics Express*, vol. 13, no. 16, pp. 6296–6304, 2005.

[9] C. Wu *et al.*, "The trigonometric orthogonality of phase-stepping curves in grating-based x-ray phase-contrast imaging: integral property and its implications for noise optimization," *Medical Physics*, 2019.

[10] P. Modregger, F. Scattarella, B. Pinzer, C. David, R. Bellotti, and M. Stampanoni, "Imaging the ultrasmall-angle x-ray scattering distribution with grating interferometry," *Physical Review Letters*, vol. 108, no. 4, p. 048101, 2012.

[11] Z. Wang, K. Kang, Z. Huang, and Z. Chen, "Quantitative grating-based x-ray dark-field computed tomography," *Applied Physics Letters*, vol. 95, no. 9, p. 094105, 2009.

[12] P. Modregger *et al.*, "Interpretation and utility of the moments of small-angle x-ray scattering distributions," *Physical Review Letters*, vol. 118, no. 26, p. 265501, 2017.

[13] P. Modregger, M. Endrizzi, and A. Olivo, "Direct access to the moments of scattering distributions in x-ray imaging," *Applied Physics Letters*, vol. 113, no. 25, p. 254101, 2018.

[14] C. Wu *et al.*, "Truncated analytic moment analysis and its hybrid-field contrast in grating-based x-ray phase contrast imaging," *arXiv preprint arXiv:1912.13196*, 2019.

[15] X. Zhu *et al.*, "Direct information retrieval methods based on moment analysis of grating-based x-ray interferometry: Analysis and feasible pre-processing approaches," in *Proceedings of the 5th International Conference on X-ray and Neutron Phase Imaging with Gratings (XNPIG)*, 2019.

A modular phantom for cone beam and whole-body CT image quality assessment

John M. Boone, Ali Uneri, Jeffrey H. Siewerdsen

***Abstract*— There is no shortage of phantoms designed for CT image quality and radiation dose assessment. However, the majority of these phantoms have expensive and complicated structures embodied within them (*e.g.* line pair structures for spatial resolution, low contrast inserts for contrast resolution, *etc.*) which enable subjective visual evaluation for quantifying image quality. A newly designed phantom, called the Corgi phantom, was designed to move the field away from qualitative assessment in favor of mathematically rigorous quantitative assessment. In addition to modules for assessment of the modulation transfer function and noise power spectrum, a cone beam test tool is also included to evaluate the magnitude of cone beam artifact at various levels along the phantom. The 13.5 cm diameter phantom is well-suited for evaluation of cone beam systems (*e.g.* orthopedic, head and neck, dental, breast, image guided radiation therapy), but with the inclusion of outer clamshell structures which increase the phantom diameter to 16.0 cm, whole body CT systems can be evaluated for head and pediatric abdomen imaging performance. This 16 cm combination phantom also fits into a 32 cm (outer) diameter phantom, for assessment of image quality for the adult body. Because image noise assessed using the noise power spectrum is of course dependent upon the radiation dose levels used, the phantom also includes central and peripheral holes to allow a thimble chamber to be inserted for the assessment of radiation dose levels. In this way, noise (*e.g.* NPS) can be reported as a function of radiation dose levels, allowing a straightforward comparison between the performance of different scanners, or the same scanner over time. Software has also been written to evaluate the DICOM images representing the volume data set of phantom images, which automatically analyzes MTF, NPS, cone beam artifact, cupping artifact, and HU accuracy.**

***Index Terms*—MTF: Modulation transfer function, NPS: noise power spectrum,**

I. Introduction

X ray Computed Tomography (CT) became a clinical modality in the 1970s, when virtually all display was analog film. In that era, access to the raw digital CT images was difficult to the point of impossible to the general scientist. CT scanners were connected directly to printers which produced film images of the CT series, typically 4 × 5 images on a 35 × 43 cm sheet of film. An entire CT scan required perhaps 10-15 of these sheets of film to accommodate the different slices, reconstruction filters, and window/level settings needed for complete interpretation by the radiologist.

Therefore, in the first decades of computed tomography, the only practical way of assessing image quality was by using purpose-built phantoms with line pair objects for qualitative assessment of spatial resolution, and rods of different contrast to allow qualitative assessment of contrast resolution. That era of qualitative image analysis in CT has been propagated to this day by continued use of legacy phantoms, including the American College of Radiology (ACR) accreditation phantom in the United States, which is still required for CT accreditation programs. In the current era where access to the digital CT images in the DICOM format is widely available, it is time to migrate from perfunctory qualitative visual image assessment to rigorous image assessment using quantitative mathematics. The Corgi phantom was designed specifically to promote this transition, not only for researchers in CT, but for vendors and clinical practitioners alike [1,2,3].

The Corgi phantom was designed initially for quality assurance and acceptance testing on cone beam CT systems for specialty applications such as orthopedic imaging, breast imaging, dental and maxillofacial imaging, cone beam systems on linear accelerators (image guided radiation therapy), and angiography systems which also have a cone beam CT mode. However, some thin clamshell polymethyl methacrylate fixtures can be added to the 13.5 cm diameter Corgi phantom to bring the total diameter to 16.0 cm, which is consistent with radiation dose phantoms used for routine dose assessment in head and pediatric body CT. Some commercially available CT dose phantoms also include a 32 cm diameter PMMA annulus with a 16 cm diameter inner cut-out for the insertion of the head CT dose phantom. These simple and widely available additions to the Corgi phantom allow it to be a viable tool to assess image quality and radiation dose in whole body CT systems. It should also be mentioned that with the advent of large collimated beam whole body CT systems (with 160 mm coverage along the *z* dimension at isocenter) , the assessment of cone beam artifacts in whole body CT systems is clearly needed.

II. Methods and Materials

The design of the modular corgi phantom is described in the figures below. The system was designed to be reconfigurable, given the different geometries of cone beam CT systems that

Submitted on January 17, 2020. Author JMB is with the Departments of Radiology and Biomedical Engineering at University of California Davis, Ellison ACC Building, 4860 Y Street, Sacramento, CA 95817, and authors AU and JHS are with the Department of Biomedical Engineering at the Johns Hopkins University. Traylor Building #622, 720 Rutland Ave., Baltimore, MD 21205 USA

are available on the market and the different clinical applications that they address. For example, most cone beam CT systems have the central ray located at the central location of the field-of-view along the z-axis; however, for breast CT systems the central ray is located as close to the posterior edge of the breast as possible, to allow maximal coverage of the chest wall. Given these differences, the cone beam module can be positioned at different locations along the length of the phantom (the z-axis of the CT scanner) to evaluate and quantify cone beam artifacts.

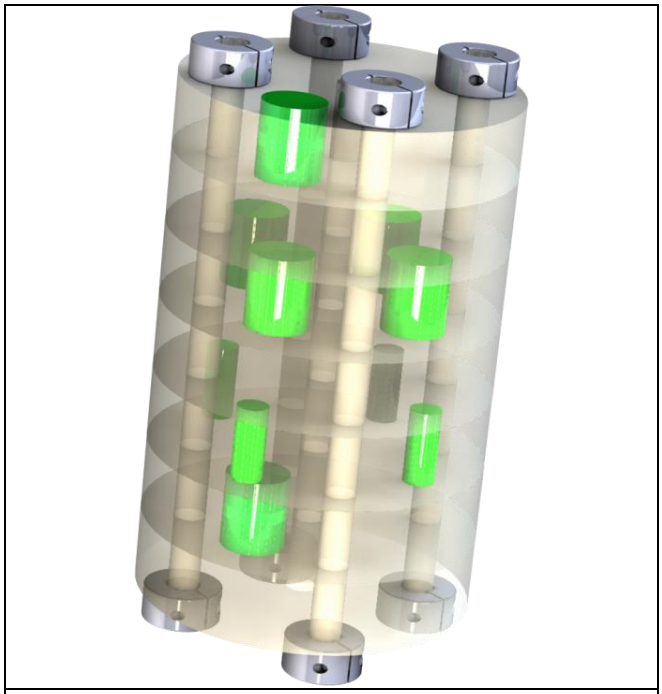


Figure 1: The initial phantom design was modular in nature and included different modules for measuring different metrics. The system was reconfigurable, which allowed appropriate placement for the various modules (e.g. cone beam artifact module) to be placed at appropriate levels along the length of the phantom, corresponding to different cone angles.

A homogeneous region of the phantom is designed to allow the computation of the 3D noise power spectrum, and the homogeneous region necessarily contains peripheral holes which allow long PMMA rods to hold the phantom together, along with the central through-whole which is designed to accommodate a thimble chamber for the assessment of radiation dose at the center of the phantom.

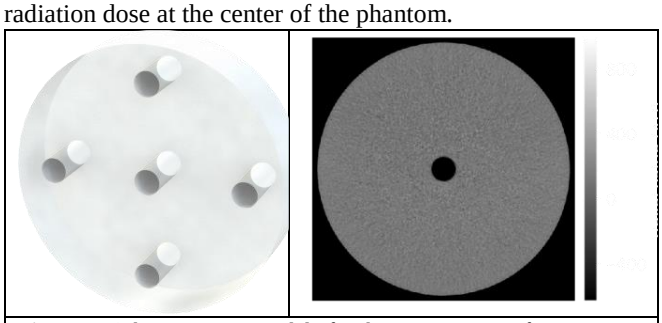


Figure 2: A homogeneous module for the measurement of noise power spectrum is included in the corgi system. The peripheral holes were initially designed to accept PMMA rods necessary for holding the module phantom together. The central hole, however, is designed to be the primary area where radiation dose can be measured.

As with many CT phantoms, it is also appropriate to include a module which contains various plastics which have different densities and atomic numbers, which are useful for quantifying the Hounsfield Unit accuracy of a given cone beam CT system. It is noted that some cone beam systems do not actually report their images using the Hounsfield Unit scale, however in these units the quantification of the linear attenuation coefficient is useful.

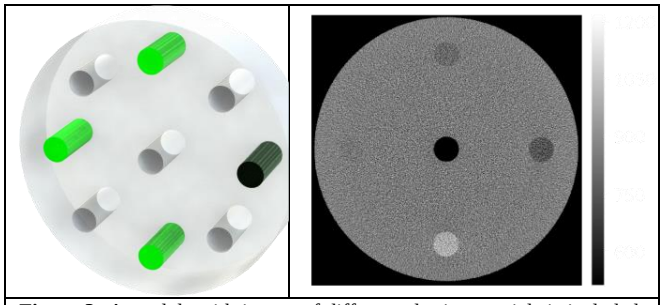


Figure 3: A module with inserts of different plastic materials is included, to allow the phantom to be used for HQ accuracy measurement.

The quantitative measurement of spatial resolution involves assessing the modulation transfer function, which is computed from slit images which are used to define the line spread function. Because the phantom is designed to accommodate CT systems with different resolution characteristics, slit thicknesses vary from very thin (0.10 mm) to thicker slit widths (0.25 mm).

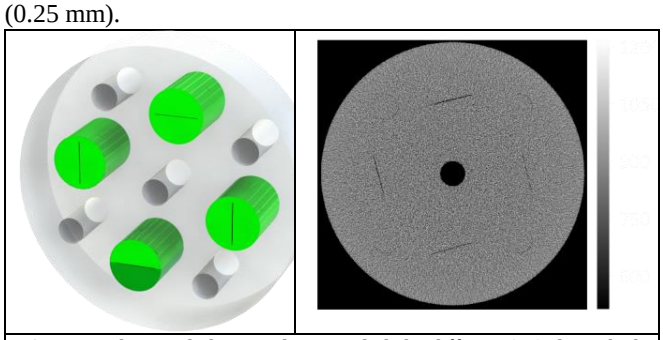


Figure 4: The initial phantom design included 4 different (air) slits, which can be used to measure the line spread function (and hence the modulation transfer function). Each slit was of a different width, to accommodate the different resolution performance of various cone beam CT systems.

The cone beam module was designed after considerable experimentation with different geometric designs on a cone beam system. The cone beam module demonstrated in Figure 5 shows the use of 2 thin Teflon disks, each 25 mm in diameter and 2 mm thick. These disks are separated by a disk of the matrix material (PMMA), which provides neutral contrast relative to the surrounding phantom.

Figures 1 through 5 demonstrate the original design of the Corgi phantom. After discussions with a commercial manufacturer, a more efficient (and less expensive) design was developed, which however leverages the individual modules illustrated in Figures 1-5. For the assessment of spatial resolution, however, the use of hemispherical rods allowing the assessment of the edge spread function in 3 dimensions has replaced the slits of Figure 4, which allow MTF measurements in the axial dimension only. The hemispherical rod component of Corgi-2 is illustrated in Figure 6.

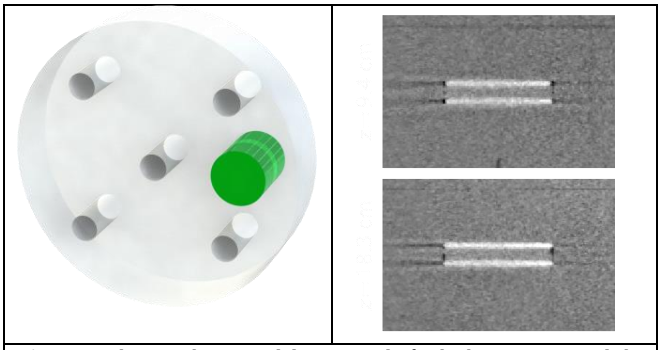


Figure 5: The cone beam module consisted of 2 higher attenuation disks (Teflon), with the spacer of the matrix material. This allows the quantification of a cone beam artifact which is discussed in the text.

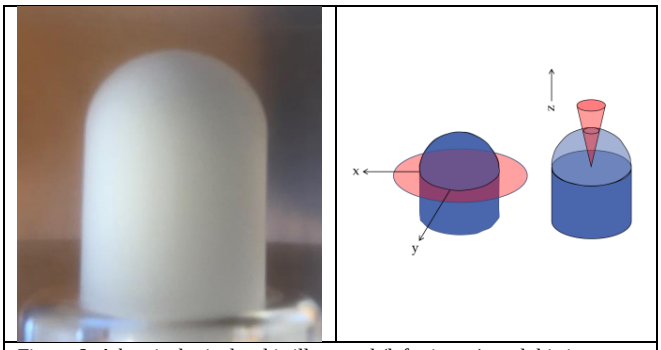


Figure 6: A hemispherical rod is illustrated (left picture), and this insert allows the assessment of spatial resolution (MTF) in both the horizontal (z, y) plane as well as along the z-axis (right figure).

III Results

The quantitative measurements which can be made on the Corgi phantom, as described above, are demonstrated below. All these metrics were automatically evaluated using software used to evaluate the DICOM images produced by the CT scanners.

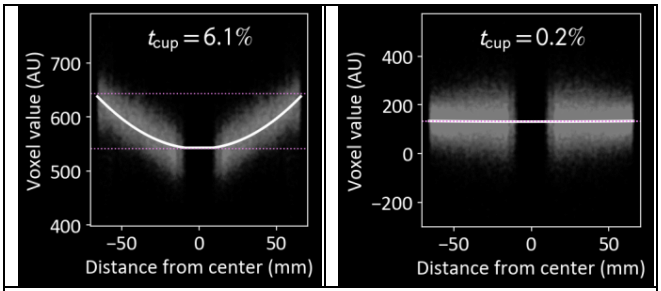


Figure 7: The degree of cupping artifact is illustrated between 2 cone beam imaging systems. The system on the left has experienced significant cupping, while the system on the right has made use of the cupping correction algorithm.

The quantitative assessment of cupping artifacts is illustrated in Figure 7, and these data were acquired from images (and the phantom module) shown in Figure 2.

Figure 8 illustrates Fourier-based metrics including the modulation transfer function (left side) and the noise power spectrum (right side). The MTF results [4] were computed from a module shown in Figure 4, while the NPS was computed from the module illustrated in Figure 3.

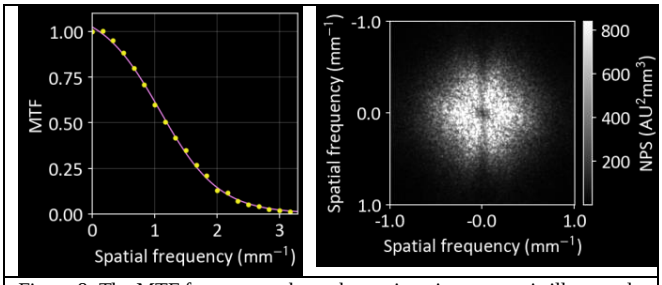


Figure 8: The MTF from a cone beam breast imaging system is illustrated on the left, and the noise power spectrum from a cone beam system designed for orthopedic imaging is shown on the right.

Figure 9 illustrates the analysis of the cone beam module (left side). The contrast to noise ratio produced from different plastic inserts (with different HU values) is shown in Figure 9 on the right side. The cone beam artifact is defined from the trace with the following mathematics:

$$Z_{\text{mod}} = \frac{\mu_{disk} - \mu_{middle}}{\mu_{disk} - \mu_{air}}$$

Where Z_{mod} is constrained between 0 and 1.

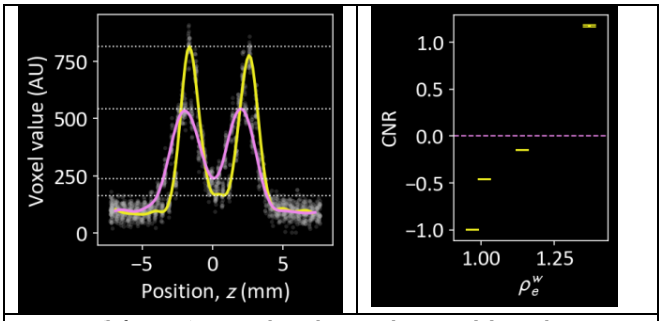


Figure 9: (left image) Traces through 2 cone beam modules is shown on the left, with the yellow trace corresponding to a cone beam module located near the central ray, and the fuchsia trace illustrating the data from a module located more distance from the central ray. The dramatic reduction in modulation is clearly visible. (right image) the HU accuracy (, cast as contrast to noise ratio) is illustrated on the pot on the right.

IV Discussion

The utility of the Corgi phantom is demonstrated in Figures 7, 8, and 9. It is noted that the quantitative metrics shown in these figures was derived from automated software which simply analyzed the CT volume data set which was produced by imaging the Corgi phantom on various CT systems. Therefore, these quantitative metrics can be produced quickly (less than 1 minute) and reproducibility. We feel that it is essential that the hardware phantom shown in Figures 1 through 6 be accompanied by analysis methods which are consistent, producing the results shown in Figures 7 through 9. In other words, the value of the physical phantom is enhanced by a uniform analysis package. Otherwise, different methods for computing (for example) the modulation transfer function may lead to different results from user to user, and this would degrade from the overall concept of a uniform, quantitative, and reproducible CT analysis package.

While not illustrated above, an important aspect of the Corgi phantom is that it is designed to quantify radiation dose as well as image quality. A typical experimental protocol for doing this is described below:

Imagine that for a given cone beam (or whole body) CT scanner, that we stipulate that a reasonable dose level would be 6 mGy measured at the central location on the Corgi phantom. The phantom is placed in the scanner with a thimble chamber located at the central hole in the phantom, and an initial scan with an appropriate tube potential, scan time and tube current (mA) is made. The reading on the thimble chamber is recorded. Without changing the tube potential or scan time, the tube current is adjusted such that the next CT scan will deliver approximately 6 mGy to the thimble chamber. That scan is then acquired, the 6 mGy dose to the center is quantified, and the image data produced from that CT scan is then evaluated using automated software. In this way, the noise power spectrum is associated with a consistent dose level, which allows comparison between different CT scanners for a given application, as well as for quantifying the performance of the same CT system over time.

The purpose of this presentation is to demonstrate the overall capabilities and utility of this new quantitative phantom for assessing image quality in both cone beam and whole-body CT systems [5]. While the results shown above are focused on cone beam CT systems, it is clear that similar results –with different quantitative endpoints – would be achieved when the Corgi phantom is used to evaluate a whole-body CT system.

V REFERENCES

[1] A. Uneri, X. Zhang, T. Yi, et al., “Image quality and dose characteristics for an O-arm intraoperative imaging system with model-based image reconstruction”, Med. Phys. 45(11), 4857–68 (2018).

[2] M.J. Daly, J.H. Siewerdsen, D.J. Moseley, D.A. Jaffray, and J.C. Irish, “Intraoperative cone-beam CT for guidance of head and neck surgery: Assessment of dose and image quality using a C-arm prototype”, Med. Phys. 33(10), 3767–80 (2006).

[3] R.L Dixon and J.M. Boone, “Cone beam CT dosimetry: A unified and self-consistent approach including all scan modalities-With or without phantom motion”, Med. Phys. 37(6), 2703–18 (2010).

[4] A.L.C. Kwan, J.M. Boone, K. Yang, and S.-Y. Huang, “Evaluation of the spatial resolution characteristics of a cone-beam breast CT scanner”, Med. Phys. 34(1), 275–81 (2006).

[5] J.H. Siewerdsen, A. Uneri, A.M. Hernandez, G.W. Burkett, J.M. Boone, “Cone-beam CT dose and imaging performance evaluation with a modular, multi-purpose phantom”, Medical Physics, (2020), *in press*

Deep learning-based dental segmentation and tooth identification in dental cone-beam CT

Taejun Jang, Sungmin Lee, Kangcheol Kim, Hyuncheol Cho, Hyesun Yun and Jin Keun Seo

***Abstract*—This paper proposes an automated segmentation and identification method of three-dimensional(3D) teeth and jaws from con-beam CT(CBCT) images, which can provide a basic platform for digital dentistry. Accurate 3D segmentation and identification of individual tooth by separating tooth and alveolar bone is difficult tasks due to their similar grayscale intensities and contact with adjacent teeth. To address these difficulties, the proposed method uses a deep learning based hierarchical model using a 2D panoramic images, that are generated by projecting CBCT images in the normal direction of the patient's dental arches of mandible and maxilla. The main preliminary step is to determine the region of interest (ROI) of the individual teeth that is obtained by accurate segmentation and identification in the 2D panoramic images. Finally, 3D teeth segmentation in the each 3D ROI is obtained through deep learning method.**

***Index Terms*—Cone-beam computerized tomography, tooth segmentation, deep learning**

I. Introduction

Recently, digital dentistry is evolving rapidly along with the rapid innovation of artificial intelligence and the development of CBCT, intra-oral and face scanner, and dental 3D printing skills. Digital dentistry can improve efficiency of streamlining the dentist's workflow and provides more accurate orthodontic diagnosis, treatment planning, and surgical guides. One of the fundamental components of digital dentistry is 3D segmentation of teeth, jaws, and skulls from CBCT images, because accurate digital models of the individual tooth geometry and jaws are important for simulating prosthetic evaluation, cephalometric analysis, computer-aided digital implant planning, predicting bite irregularities, and so on.

Automatic and accurate segmentation of individual teeth from 3D CBCT images is a challenging task due to contact with adjacent teeth and similar intensities between tooth and their neighboring alveolar bone. Since existing automatic segmentation methods [1] [2] [3] [4] [5] have not yet reached a satisfactory level of clinical use, dentists still prefer to use semi-automatic segmentation methods with their manual intervention but it is a time-consuming and fatigue-prone process.

To address these difficulties, the proposed method uses a deep learning based hierarchical model consisting of the following four steps:

Step 1 Generates 2D panoramic images by projecting in the normal direction of dental arches for maxilla and mandible in CBCT images.

Step 2 Perform segmentation and identification of individual teeth in the 2D panoramic images obtained in step 1.

Step 3 Determine 3D ROIs of individual teeth in 3D CBCT image by backprojecting segmented results in step 2.

Step 4 Get 3D tooth segmentation from the individual 3D ROI obtained in step 3 using the knowledge of the 2D tooth segmentation obtained in step 2.

The efficacy of the proposed method was experimentally validated using 96 CBCT data. Our method successfully achieved tooth segmentation with dice coefficient 90% and identification with accuracy 96%.

Taejun Jang was with the Department of Computational Science and Engineering, Yonsei University, Seoul, South Korea, e-mail: (tajjang10@gmail.com)

II. Method

The proposed hierarchical learning method aims for automatic and accurate tooth segmentation and identification in 3D CBCT images. Fig.1 shows the workflow of the proposed method that consist of four steps.

A. Step 1: Generating a panoramic image from 3D CBCT

We generate a panoramic image, a flattened image of the curved maxilla and mandible anatomical structures, to identify and localize each tooth in 3D CBCT images. To generate the 2D panoramic image, we first find a reference curve (medial axis) passing through middle of teeth and to project in the normal direction of the curve. Fig. 2) shows our approach which consist of the following process: (i) Get binarized skull data by thresholding. (ii) Separation of maxilla and mandible using the region growing technique [17] and acquire the gray-scaled maxilla and mandible data. (iii) Detect the region of a dental arch using maximum intensity projection(MIP) images and the morphological operator [18]. (iv) Obtain the reference curve using skeletonization method [19] from dental arches. (v) Get the panoramic image by projecting CBCT images in the normal direction of the reference curve.

B. Step 2: Segmentation and identification of individual teeth in the 2D panoramic image

Let $\mathbf{I}$ denote a panoramic image generated by Step 1. We denote the center point of $i\,th$ tooth in $\mathbf{I}$ by $\mathbf{p}_i$. To identify center points of all teeth, we adapt the deep learning-based pose estimation [11]–[13] and labeled data $\{\mathbf{H}^{(k)} : k = 1, \cdots, K\}$ given by

$$\mathbf{H}(\mathbf{x}) = \max_{i=1,2,\cdots,N} \left[\exp\left(-\frac{\|\mathbf{x} - \lfloor \frac{\mathbf{p}_i}{8} \rfloor\|^2}{\sigma^2} \right) \right]. \qquad (1)$$

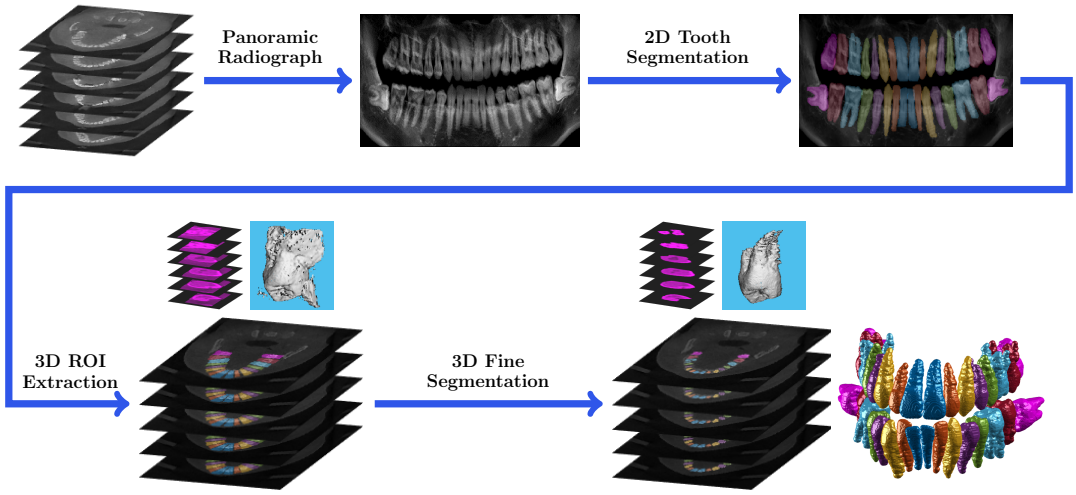


Fig. 1: The workflow of the proposed method. It consists of four steps: 1) generating a panoramic radiograph from a 3D CBCT image, 2) detecting ROIs including single teeth in the panoramic radiograph, 3) generating 3D ROIs including single teeth in the 3D CBCT image, and 4) 3D teeth segmentation in the each 3D ROI.

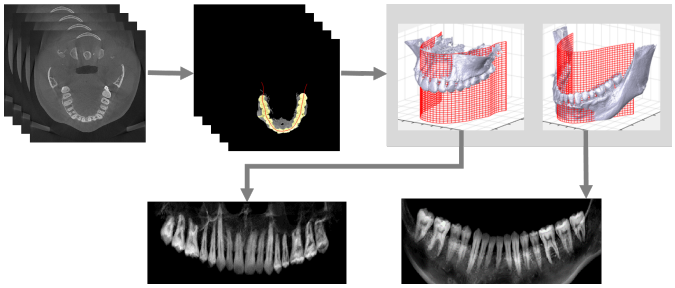
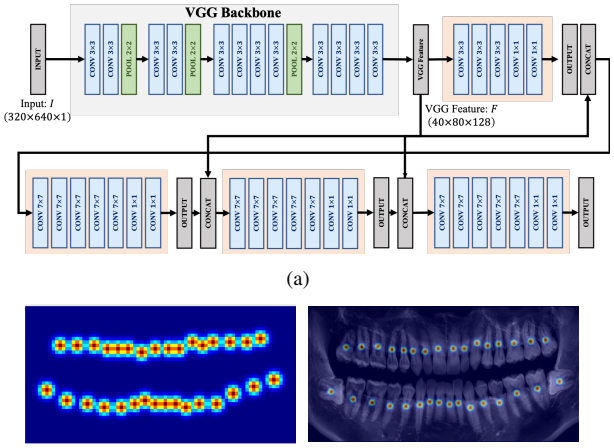


Fig. 2: The process of generating 2D panoramic radiograph.



(a)

(b)

Fig. 3: (a) The proposed network to detect center points of teeth. (b) The left is label of size 40×80 and the right is the image overlapping a panorama image and the interpolated label.

Our deep learning architecture for the teeth identification is shown in Fig. 3(a). Fig.3(b) shows the heat map $f_{pose} : \mathbf{I} \rightarrow \mathbf{H}$ that is obtained by minimizing the following loss function:

$$Loss = \sum_k \|f_{pose}(\mathbf{I}^{(k)}) - \mathbf{H}^{(k)}\|^2 \tag{2}$$

The output $f_{pose}(\mathbf{I})$ allows to detect center points of teeth, which will be denoted by $\mathbf{c}_{ij} = (x_{ij}, y_{ij})$. Here, the subindex ij is the dental notation of ISO system by WHO.

Next, for each tooth with $\mathbf{c}_{ij}$, we determine the corresponding bounding box($\mathbf{I}^{ij}_{roi}$) with the following height h_{ij} and width w_{ij}:

$$w_{ij} = \begin{cases} 2|x_{ij} - x_{i(j+1)}| & for\ j = 1, \cdots, 7 \\ 2|x_{i8} - x_{i7}| & if \text{ a wisdom teeth exists} \end{cases} \tag{3}$$

$$h_{ij} = \begin{cases} 2|y_{ij} - 1| & if\ i = 1, 2 \\ 2|y_{ij} - 320| & if\ i = 3, 4 \end{cases} \tag{4}$$

We use the U-Net [14] to learn the segmentation map $f_{seg} : \mathbf{I}_{roi} \rightarrow \mathbf{S}_{roi}$ as shown in Fig. 4.

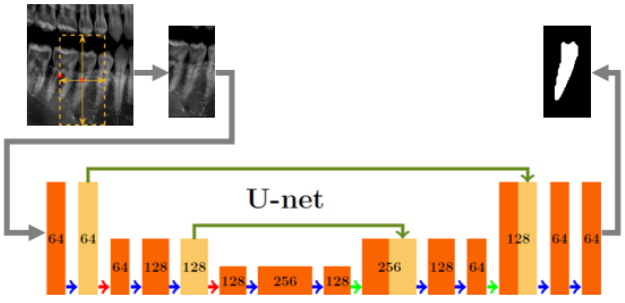


Fig. 4: U-Net for tooth segmentation in a panoramic image

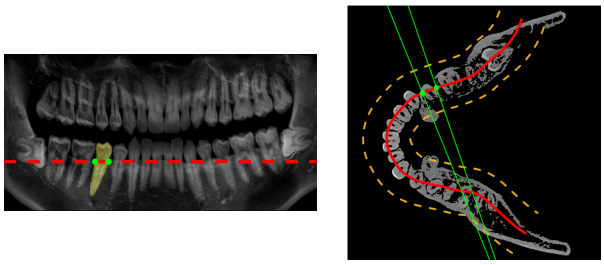


Fig. 5: Tooth segmentation in the panoramic image and back-projection for determining 3D ROIs.

C. Step 3: Determine 3D ROIs of individual teeth in the CBCT image

Now we are ready to determine 3D ROIs of individual teeth using Step 1 and Step 2. Each tooth's ROI in 3D CBCT data is determined by back-projection using the segmentation of tooth in the panoramic image and the reference curve, as shown in Fig.5.

D. Step 4: 3D tooth segmentation

We perform the 3D tooth segmentation in the individual 3D ROI obtained in the Step 3. The size of the rescaled ROI is $64 \times 64 \times 128$ which is much smaller than that of the entire 3D CBCT data. Taking advantage of the small size of the ROI, we can effectively learn the 3D tooth segmentation.

For 3D tooth segmentation in the small size of the ROI, we use M-Net [15] which has a multi-scale side-input and side-output to reinforce the network to obtain more robust and generalized result.

III. Result

This section demonstrates the feasibility of the proposed method using 3D dental data. We use dental dataset, provided by CBCT with circular trajectory (Dentri, HDXWILL, Korea). This dataset was used to train deep learning networks in Step 2 and Step 4. The size of the original CBCT image was $800 \times 800 \times 400$. The labeling of 2D segmentations in Step 2 and 3D segmentations in Step 4 was performed by authors and experts in the company HDXWILL. Training was implemented by Tensorflow on a CPU (Intel(R) Core(TM) i7-6850K, 3.60GHz) and a GPU (NVIDIA GTX-1080, 8GB) system.

Fig.6(a) shows segmentation results of tooth and jaws. Fig.6(b) shows the visualized segmentation and identification of teeth. The proposed method is found to be very accurate in segmentation and identification of individual teeth.

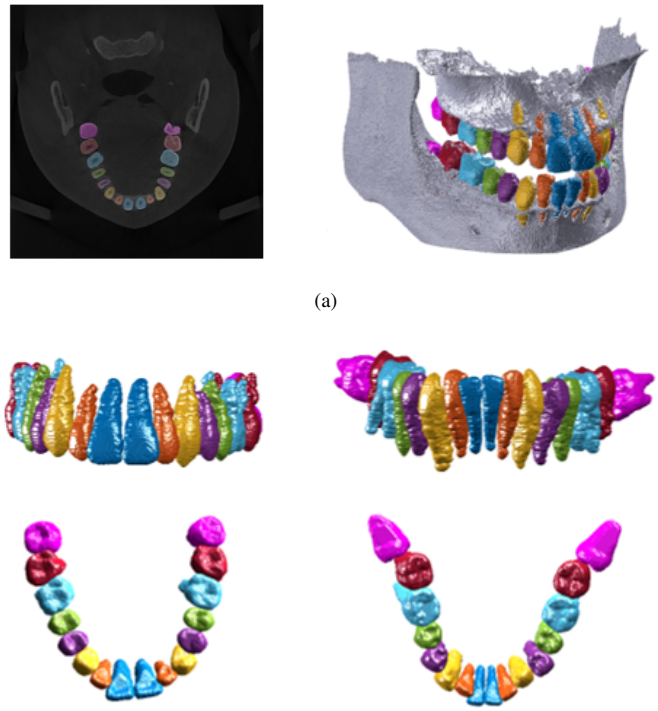


Fig. 6: Results of 3D tooth segmentation and identification.

IV. Conclusion

This paper proposes a fully 3D automated segmentation and identification method for individual teeth and jaws from CBCT images. Recently, there have been numerous researches on automated segmentation of teeth and jaws from 3D CT images to reduce the need of manual (cumbersome and time-consuming) intervention by dentists, but conventional automated methods have not been achieved at a level of accuracy that can be used immediately without dentist's interventions in a clinical setting.

Recently, deep learning methods using U-net have shown excellent performance in automatic 2D slice-by-slice segmentations, whereas accurate 3D segmentation with as limited training data is still difficult. The proposed hierarchical learning method is to deal with the difficulty of learning of high-dimensional data. For 3D segmentation of individual teeth, we initially generated a 2D panoramic image projected along the reference curve which represents a region-based shape feature of a dental arch. We use the deep learning methods for 2D tooth segmentation and identification in the 2D panoramic image. This plays an important role in determining 3D individual bounding boxes in 3D CBCT images. Then, we use another deep learning method to achieve accurate 3D tooth segmentation from the small bounding boxes.

The proposed method has scope for improvement. It would be necessary to integrate our multimodal hierarchical learning framework into a unified learning framework. To avoid the redundancies and reduce the computational cost, it is desirable to merge these multiple process models into a single framework.

Acknowledgements

This work was supported by the Samsung Science & Technology Foundation.

References

[1] S. Barone, A.Paoli, and A. V. Razionale, CT segmentation of dental shapes by anatomy-driven reformation imaging and B-spline modelling, *International journal for numerical methods in biomedical engineering*, 32(6), e02747, 2016.

[2] Y. Gan, Z. Xia, J. Xiong, G. Li, and Q. Zhao, Tooth and alveolar bone segmentation from dental computed tomography images. *IEEE journal of biomedical and health informatics*, 22(1):196–204, 2018.

[3] H. Gao and O. Chae, Individual tooth segmentation from ct images using level set method with shape and intensity prior. *Pattern Recognition*, 43(7):2406–2417, 2010.

[4] M. Hosntalab, R. A. Zoroofi, A. A. Tehrani-Fard, and G. Shirani. Segmentation of teeth in ct volumetric dataset by panoramic projection and variational level set. *International Journal of Computer Assisted Radiology and Surgery*, 3(3-4):257–265, 2008.

[5] D. X. Ji, S. H. Ong, and K. W. C. Foong. A level-set based approach for anterior teeth segmentation in cone beam computed tomography images. *Computers in biology and medicine*, 50:116–128, 2014.

[6] X. Zhou, Y. Gan, J. Xiong, D. Zhang, Q. Zhao, and Z. Xia, A method for tooth model reconstruction based on integration of multimodal images. *Journal of Healthcare Engineering*, 2018.

[7] H. Akhoondali, R. A. Zoroofi, and G. Shirani. Rapid automatic segmentation and visualization of teeth in ct-scan data. *Journal of Applied Sciences*, 9(11):2031–2044, 2009.

[8] H. T. Yau, T. J. Yang, and Y. C. Chen. Tooth model reconstruction based upon data fusion for orthodontic treatment simulation. *Computers in biology and medicine*, 48:8–16, 2014.

[9] Y. Gan, Z. Xia, J. Xiong, Q. Zhao, Y. Hu, and J. Zhang. Toward accurate tooth segmentation from computed tomography images using a hybrid level set model. *Medical physics*, 42(1):14–27, 2015.

[10] Y. Pei, X. Ai, H. Zha, T. Xu, and G. Ma. 3d exemplar-based random walks for tooth segmentation from cone-beam computed tomography images. *Medical physics*, 43(9):5040–5050, 2016.

[11] S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 4724-4732, 2016.

[12] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv*, preprint arXiv:1812.08008, 2018.

[13] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 7291-7299), 2017.

[14] O. Ronneberger, P. Fischer, and T.Brox. U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical image computing and computer-assisted intervention, Springer, Cham*, (pp. 234-241), 2015

[15] Mehta, Raghav, and Jayanthi Sivaswamy, M-net: A convolutional neural network for deep brain structure segmentation. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017).* IEEE, 2017.

[16] N. Otsu, A threshold selection method from gray-level histograms, *IEEE transactions on systems, man, and cybernetics*, 9(1), 62-66, 1979.

[17] R. Adams, and L.Bishof, Seeded region growing, *IEEE Transactions on pattern analysis and machine intelligence*, 16(6), 641-647, 1994.

[18] R. M. Haralick, S. R. Sternberg, and X. Zhuang, Image analysis using mathematical morphology. *IEEE transactions on pattern analysis and machine intelligence*, (4), 532-550, 1987.

[19] T. C. Lee, R. L. Kashyap and C. N. Chu. Building skeleton models via 3-D medial surface/axis thinning algorithms. *Computer Vision, Graphics, and Image Processing*, 56(6):462-478, 1994.

FDRN: Fast Deformable Image Registration Based on Unsupervised Convolutional Neural Networks

Kaicong Sun, Zhaowen Gong, and Sven Simon

***Abstract*—Deformable image registration is a fundamental and challenging task for MRI and CT image processing. Due to the large amount of computational demand for volumetric data, optimization-based methods usually face the difficulty on execution time in practice even using GPU (Graphics Processing Unit) acceleration. In this paper, we present a fast deformable registration network (FDRN) based on unsupervised learning. A lightweight autoencoder network structure has been proposed. By means of the spatial transformer layer, the moving image can be resampled according to the estimated displacement vector field (DVF) such that a similarity measure between the fixed and the resampled moving images is optimized without requiring the ground truth DVF. An auxiliary loss has been applied on the bottom encoder layer to constrain the local disturbance of the DVF and mitigate the gradient vanishing during the training process. To reduce the memory consumption, additive forwarding is adopted linking the associated encoder and decoder layers instead of concatenation. We carried out the comparison with the optimization-based and learning-based state-of-the-art methods. The experimental results demonstrate the effectiveness and the feasibility of the proposed FDRN.**

***Index Terms*—Deformable image registration, deep convolutional neural networks, unsupervised learning, multi-resolution.**

I. Introduction

In the past decades, deformable image registration has been intensively explored [1]. Unlike rigid and affine transformation, deformable registration establishes dense spatial correspondences with higher degrees of freedom and allows describing the transformation according to local anatomical structures. Typically, the expected transformation matrix T is encoded in the following loss function

$$Loss = D(F(\mathbf{x}), T \circ M(\mathbf{x})) + \lambda R(T) \tag{1}$$

where $\mathbf{x}$ indicates the 3D spatial coordinate with $\mathbf{x} \in \Omega$ and $\Omega \subset \mathbb{R}^3$. The notation D represents some dissimilarity criteria or distance between the fixed image F and the transformed moving image $T \circ M$. R serves as the regularization term which usually interprets the prior knowledge of the expected transformation. The weight λ scales the penalization of the regularization R. The optimization-based methods usually minimize the loss function given by Eq. (1) iteratively and suffer from the computational complexity of the volumetric data. On the contrary, learning-based methods perform purely forward propagation on the well-trained networks which alleviates the computational demand by offline training. However, the generation of reliable training datasets containing the ground-truth displacement vector field (DVF) for supervised learning is always a challenging and inevitable task. The performance of the networks is critically dependent on the fidelity of the training datasets. In this paper, we propose an unsupervised deep convolutional neural network (DCNN) containing following contributions:

- We present an unsupervised convolutional neural network for deformable registration based on an autoencoder architectured multi-resolution framework.
- An auxiliary loss has been devised to enhance the robustness of the transformation estimation and circumvent the gradient vanishing during the training process.
- The proposed FDRN provides the state-of-the-art performance of deformable registration for medical images within satisfactory inference time as shown in Fig. 1 which brings the FDRN towards practical registration applications.

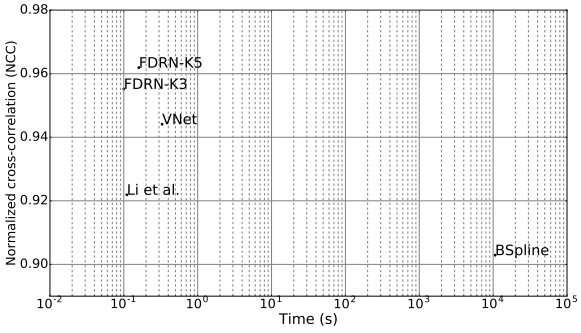


Fig. 1. Plot of performance of different methods in terms of the registration accuracy and the execution time.

II. Related Work

A. Medical Image Registration via Optimization

Deformable registration has been a considerable research field for decades [1]. Typically, the transformation is modeled in forms of DVF $d(\mathbf{x})$ [2] or velocity vector field $v(\mathbf{x}, t)$ [3]. The former one considers the transformation as the summation of the identity transform $\mathbf{x}$ with a smooth displacement vector field $d(\mathbf{x})$: $T(\mathbf{x}) = \mathbf{x} + d(\mathbf{x})$. Thirion treats the registration as a diffusion process and proposes Demons [2]. This assumption model does not guarantee an invertible bijective mapping. In other words, interchanging the two input images may not lead to an inverse transformation. In fact, some global registration

K. Sun, Z. Gong and S. Simon are with the Institute for Parallel and Distributed Systems, University of Stuttgart, Stuttgart, 70569, Germany e-mail: Kaicong.Sun@ipvs.uni-stuttgart.de.

Manuscript received xxx January, 2020; revised xxx, 2020.

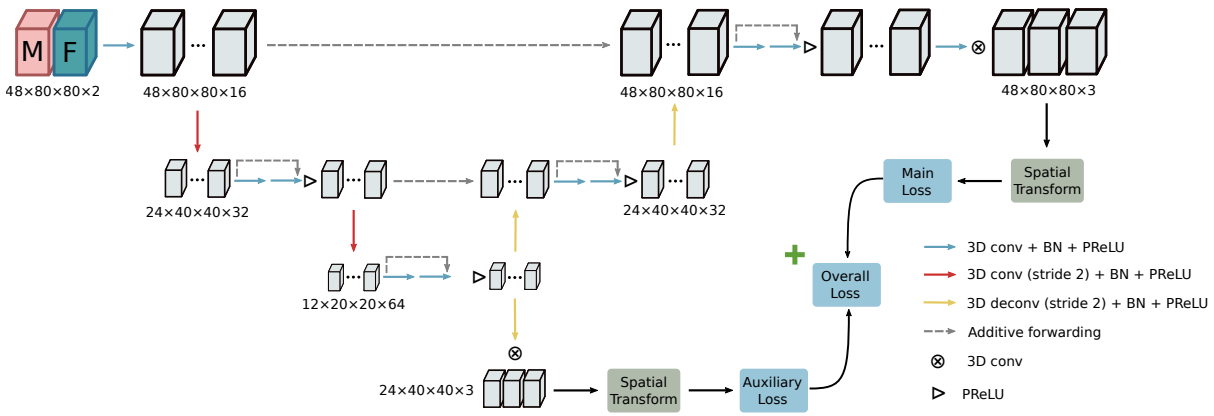


Fig. 2. Network architecture of the proposed FDRN. The dimension of the feature maps are illustrated at each stage.

like rigid or affine transformation is performed as preprocessing when tackling large deformation problems. The latter one addresses the deformable registration as a variational problem with constrains like diffeomorphism, topology preservation, symmetry and inverse consistency [4], [5]. In [4], the large deformations diffeomorphic metric mapping (LDDMM) is presented to estimate the optimal transformation in the space of smooth velocity vector field. Avants *et al.* [5] propose a symmetric diffeomorphic registration method steered by cross-correlation.

B. Medical Image Registration via Learning

In the last two decades, many studies for medical image registration have been carried out based on learning [6], [7]. Rohé *et al.* [8] apply a supervised DCNN model due to the outstanding performance on feature extraction. More recently, resorting to the spatial transformer network (STN) [9], some works [10], [11] propose to apply unsupervised CNN on deformable registration. Specially, Balakrishnan *et al.* present an autoencoder structured network VoxelMorph consisting concatenating forwarding. In [11], a multi-resolution scheme is applied on the fully convolutional networks (FCNs). Although auxiliary losses are adopted at multiple resolutions, they have not fully explored and utilized the localization information which is extremely critical for deformable image registration.

III. Method

The aim of FDRN is to perform a deformable registration and align the moving volume M with the fixed volume F. Specially, a displacement vector field $d(\mathbf{x})$ representing the spatial deformation in three dimensions (u, v, w) is generated as the output of FDRN. The transformation T is formulated as an addition of the identity transformation $\mathbf{x}$ with the corresponding displacement vector field $d(\mathbf{x})$ given by $T(\mathbf{x}) = \mathbf{x} + d(\mathbf{x})$. The network is trained such that the loss function $Loss_{total}$ formulated in Eq. (2) is minimized giving the fixed volume F and the transformed moving volume $T \circ M$.

The overall loss function of the proposed FDRN consists of two parts and is formulated as

$$Loss_{total} = \alpha_1 Loss_{main}(d_h) + \alpha_2 Loss_{aux}(d_l) \qquad (2)$$

where d_l and d_h denote the DVF in the low-resolution and high-resolution image grid, respectively. α_1 and α_2 act as the weights for balancing the main and auxiliary loss functions. As deformable image registration is inherently an ill-posed problem, regularization of the transformation $R(T)$ is inevitably required. In particular, the individual loss is expressed as

$$\begin{aligned} Loss_{main}(d_h) &= -NCC(F_h, T_h \circ M_h) + \lambda_1 TV(d_h), \\ Loss_{aux}(d_l) &= -NCC(F_l, T_l \circ M_l) + \lambda_2 TV(d_l), \end{aligned} \qquad (3)$$

where the normalized cross-correlation (NCC) is a similarity measure given by

$$NCC(F, T \circ M) = \frac{\sum_{\mathbf{x}}^{\Omega}(F(\mathbf{x}) - \bar{F})(T \circ M(\mathbf{x}) - \overline{T \circ M})}{N\sqrt{var(F)var(T \circ M)}}. \qquad (4)$$

The schematic illustration of the network structure of FDRN is demonstrated in Fig. 2. Two volumes F and M with dimension $48 \times 80 \times 80$ act as a two channel input of FDRN. There are three stages along the encoder path containing feature maps with different dimensions. At the first stage, a 3D convolutional layer followed by batch normalization (BN) and activation (PReLU) layer is adopted. BN and PReLU are used to facilitate the training effectiveness and involve the nonlinearity. Kernel size of 3 is used for all the convolutional layers. The size of the feature map is reduced to $1/8$ at the second stage by applying convolution with kernel size/stride of 2. Two other convolutions are performed in sequence at the same stage. An auxiliary loss is involved on lowest resolution image grid at the last stage to enhance the robustness of DVF and alleviate the gradient vanishing during the training. 3D deconvolutional layers are applied to assemble the low-resolution feature maps and restore the input dimension. Additive forwarding is adopted to incorporate the high-resolution localization information in the encoder path into the corresponding decoder one whilst without consuming additional memory consumption. Skip connection is employed within the stage to enable residual learning and enhance the training effectiveness.

The spatial transformer operation $T \circ M(\mathbf{x})$ marked by the gray block in Fig. 2 serves as a trilinear interpolation at the nonvoxel locations to construct a deformed moving volume based on the estimated DVF. Particularly, for each

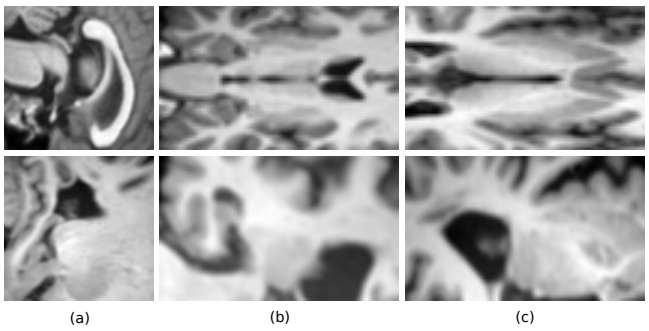


Fig. 3. Views of cropped regions in the training datasets ABIDE and ADNI. Upper row – ABIDE, bottom row – ADNI: (a) axial view (b) coronal view (c) sagittal view.

voxel located at $\mathbf{x}$, the gray value in the deformed volume can be interpolated and calculated by:

$$[T \circ M(\mathbf{x})]_{\mathbf{x}} = \sum_{\mathbf{x}_n \in \mathbf{N}(d(\mathbf{x}))} M(\mathbf{x}_n) \prod_{m \in \{u,v,w\}} (1 - |d_m(\mathbf{x}) - d_m(\mathbf{x}_n)|), \quad (5)$$

where u, v, w represent the three degrees of freedom of $d(\mathbf{x})$ at each voxel $\mathbf{x}$ and N denotes the voxel neighbors at the transformed location.

R_1 serves as the smoothness constrain on the transformation T. In this paper, we apply the 3D total variation (TV) on the displacement field $d(\mathbf{x})$ which is formulated as

$$R_1(d(\mathbf{x})) = \sum_{S_m} ||d(\mathbf{x}) - S_m d(\mathbf{x})||_1, \quad (6)$$

where S_m indicates the shifting operator along (u, v, w) directions with a window size of one. The subscript m denotes the moving direction and is expressed as $m = \{(u, v, w)|u, v, w \in \{0, 1\}\}$.

TABLE I
TRAINING PARAMETERS OF FDRN.

Patch size	Batch size	Optimizer	Learning rate	α_1	α_2	λ_1	λ_2
48×80×80	2	Adam	5×10^{-4}	0.3	0.7	10^{-5}	10^{-7}

IV. EXPERIMENTAL RESULTS

In the first experiment, the proposed FDRN is trained and evaluated on two public 3D MRI brain datasets: Autism Brain Imaging Data Exchange (ABIDE) [12] and Alzheimers Disease Neuroimaging Initiative (ADNI) [13]. It should be noted that FDRN is not limited to MR imaging but is also reliable and feasible for CT imaging. Specially, we selected 109 and 111 images from ABIDE I and ADNI I datasets, respectively. We split 109 images of ABIDE into 80, 16 and 13 for training, validation and testing. For ADNI, we chose 80 out of 111 for training, 16 for validation and 15 for testing. In oder to normalize the datasets, all of these images were firstly registered to the Montreal Neurological Institute (MNI) space using affine transformation by resorting to ICBM152 template [14]. We cropped a region of size 48×80×80 on both datasets containing left/right caudate in ABIDE and a part of cerebellum in ADNI as shown in Fig. 3. We implemented two FDRN variants denoted by FDRN-K3 and FDRN-K5 representing kernel size 3 and kernel size 5. The latter one has larger effective receptive field and more variables while in return consumes more inference time. The training parameters are detailed in Tab. I

We compared our proposed FDRN-K3 and FDRN-K5 with other learning-based state-of-the-art networks Li *et al.* [11], VNet [15] and the optimization-based B-spline transformation [16]. Specially, we implemented VNet and Li's model in Pytorch and trained them individually using the same datasets ABIDE and ADNI on Nvidia GeForce GTX 980M. B-spline registration was conducted on the CPU and we fine-tuned the registration parameters for B-spline transformation with a cross-correlation similarity measure in SimpleITK. The evaluation of different models on testing images is quantitatively depicted in Tab. II. The second column "Affine" shows the NCC of the patches before registration. It is shown that FDRN-K3 outperforms VNet and Li's model in terms of both NCC and inference time. FDRN-K5 provides better NCC than FDRN-K3 but performs slower. For qualitative evaluation, we conducted the registration of one fixed image with the rest 12 moving images in the testing dataset of ABIDE. The mean of the 12 registered images of different methods are demonstrated

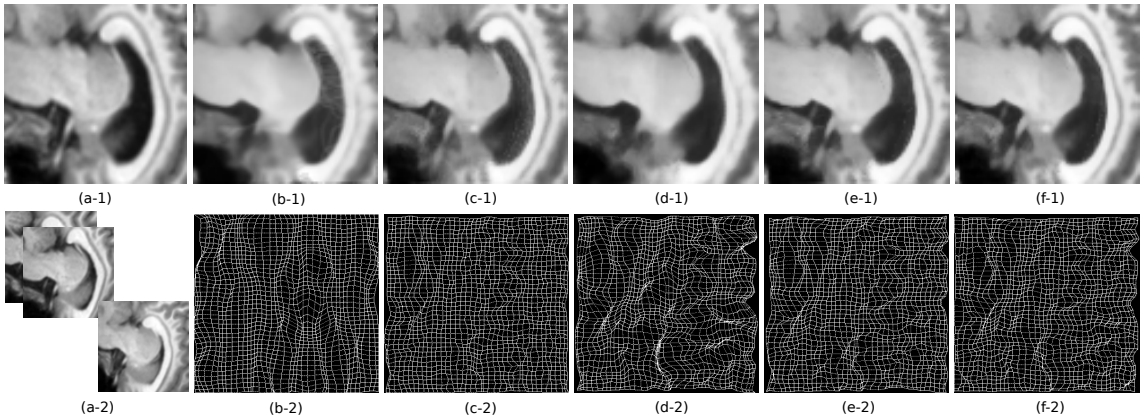


Fig. 4. Comparison of different methods on the mean of 12 registered testing images of ABIDE: (a-1) fixed image, (b-1) mean of B-spline, (c-1) mean of VNet, (d-1) mean of Li *et al.*, (e-1) mean of FDRN-K3, (f-1) mean of FDRN-K5, (a-2) 12 moving images, (b-2)∼(f-2) the corresponding deformation field.

TABLE II
REGISTRATION ACCURACY MEASURED BY NCC AND INFERENCE TIME OF DIFFERENT REGISTRATION METHODS.

Dataset	Affine[a]	B-spline[a]	VNet [15]	Li *et al.* [11]	FDRN-K3	FDRN-K5
ABIDE	0.5838±0.0696	0.8935±0.0237	0.9441±0.0125	0.9219±0.0184	0.9552±0.0118	0.9619±0.0100
ADNI	0.3997±0.1295	0.8256±0.0441	0.9205±0.0221	0.8868±0.0216	0.9376±0.0180	0.9392±0.0155
Inference time	–	7861.09s	0.33s	0.11s	0.10s	0.16s

[a] Experiments were carried out by the toolbox SimpleITK [16] and executed on Intel i7-6820HK CPU.

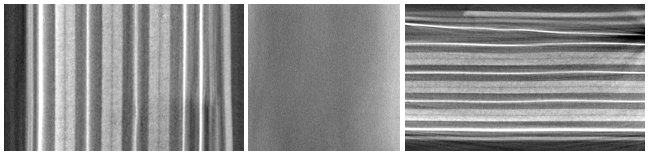


Fig. 5. Views of cropped custom CT volumes of pouch cells.

in Fig. 4. We can observe that our methods produce much better alignment and sharper images than the others. Furthermore, we evaluated the effect of different network components in FDRN, i.e., the additive forwarding, the skip connection and auxiliary loss. The performance of various configurations is exhibited in Tab. III. It is clearly shown that the auxiliary loss benefits the FDRN most in terms of the registration accuracy and then the additive forwarding.

TABLE III
ABLATION STUDY OF THE PROPOSED FDRN-K3.

Add. forward.	✗	✓	✗	✗	✗	✓	✓	✓
Skip connect.	✗	✗	✓	✗	✓	✗	✓	✓
Auxiliary loss	✗	✗	✗	✓	✓	✓	✗	✓
ABIDE	0.9299	0.9424	0.9353	0.9441	0.9510	0.9543	0.9481	0.9552
ADNI	0.8843	0.9112	0.8969	0.9139	0.9227	0.9358	0.9209	0.9376

In the second experiment, we trained the proposed model on custom CT datasets of pouch cells shown in Fig. 5. Specially, pouch cells were scanned by our Nikon HMX ST 225 CT scanner at 0% State of charge (SoC) and 100% SoC. The aim is to detect the deformation of pouch cells between 0% and 100% SoC. The CT volumes were cutted into 5392 pairs of patches with size of $64 \times 64 \times 64$ for training, 299 pairs for validation and 301 pairs for testing. We conducted a comparison with the state-of-the-art methods. The evaluation results depicted in Tab. IV demonstrate the outstanding performance of the proposed FDRN as expected and verify the feasibility and reliability of FDRN on CT data.

TABLE IV
EVALUATION ON CUSTOM CT DATASETS OF POUCH CELLS IN NCC AND INFERENCE TIME.

Dataset	Affine	B-spline	Li *et al.* [11]	FDRN-K3
Cells	0.8185±0.1330	0.9024±0.0522	0.9636±0.0403	0.9787±0.0287
Time	–	5873.10s	0.09s	0.08s

V. CONCLUSION

In this paper, we propose an unsupervised fast deformable registration network FDRN based on deep convolutional neural networks for MRI and CT images. A lightweight autoencoder structure in conjunction with an auxiliary loss and the spatial transformer layer endows the proposed FDRN with outstanding registration accuracy and inference time. A comparison of the proposed model with other state-of-the-art methods was conducted and the experimental evaluation demonstrates the feasibility and validity of our FDRN. Moreover, an analysis on the network structure exposes the significance and effectiveness of applying an auxiliary loss on the unsupervised FDRN.

REFERENCES

[1] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, 2013.
[2] J. Thirion, "Image matching as a diffusion process: an analogy with maxwells demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.
[3] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, pp. 95–113, 2007.
[4] M. F. Beg, M. I. Miller, A. TrouvÉ, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *Int. J. Comput. Vision*, vol. 61, no. 2, pp. 139–157, 2005.
[5] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, 2008.
[6] D. Rueckert *et al.*, "Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 1014–1025, 2003.
[7] B. Gutiérrez-Becker, D. Mateus, L. Peter, and N. Navab, "Guiding multimodal registration with learned optimization updates," *Medical image analysis*, vol. 41, pp. 2–17, 2017.
[8] M. M. Rohé *et al.*, "Svf-net: Learning deformable image registration using shape matching," in *MICCAI*, 2017, pp. 266–274.
[9] M. Jaderberg *et al.*, "Spatial transformer networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 2017–2025.
[10] G. Balakrishnan *et al.*, "An unsupervised learning model for deformable medical image registration," *CVPR*, pp. 9252–9260, 2018.
[11] H. Li and Y. Fan, "Non-rigid image registration using self-supervised fully convolutional networks without training data," *IEEE 15th International Symposium on Biomedical Imaging*, pp. 1–4, 2018.
[12] A. Martino *et al.*, "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," *Molecular psychiatry*, vol. 19, no. 6, pp. 659–667, 2014.
[13] S. G. Mueller, M. W.Weiner, L. J. Thal *et al.*, "Ways toward an early diagnosis in alzheimers disease: The alzheimers disease neuroimaging initiative (adni)," *Alzheimers & Dementia*, vol. 1, no. 1, pp. 55–66, 2005.
[14] V. Fonov *et al.*, "Unbiased average age-appropriate atlases for pediatric studies," *NeuroImage*, vol. 54, no. 1, pp. 313–327, 2011, [doi: 10.1016/j.neuroimage.2010.07.033].
[15] F. Milletari, N. Navab, and S. A. Ahmadi, "V-net: fully convolutional neural networks for volumetric medical image segmentation," in *IEEE 4th International Conference on 3D Vision (3DV)*, 2016, pp. 565–571.
[16] B. C. Lowekamp *et al.*, "The design of simpleitk," *Front. Neuroinform.*, vol. 7, pp. 1–14, 2013, [doi: 10.3389/fninf.2013.00045].

Low-dose CT image reconstruction with a deep learning prior

Hyoung Suk Park and Kiwan Jeon

***Abstract*—In low-dose computed tomography (LDCT), penalized weighted least squares (PWLS) approach that incorporates Poisson statics of X-ray photons can significantly reduce the excessive quantum noise. To improve the LDCT image quality, various prior information such as total variation or Markov random field has been imposed to the target image. However, they may be limited to reflect the characteristics of LDCT images, resulting in unexpected side effects (e.g. blurry images). In this paper, we proposed a deep PWLS method combined with a deep learning prior learned from standard dose CT (SDCT) images; the proposed model learns the noise reduction function that maps LDCT image to its corresponding SDCT images, and meanwhile estimating the prior distribution of SDCT images in the sense of the Pearson χ^2 divergence.**

***Index Terms*—Low-dose computed tomography, penalized weighted least squares, deep learning**

I. Introduction

X-ray computed tomography (X-ray CT) is a diagnostic imaging modality widely used in medical radiology. It produces cross-sectional images of the human body with a high resolution. However, there is an increasing demand for radiation dose reduction as concerns about the risk of radiation induced cancer are growing [6], [12]. In clinical practice, reducing X-ray source current is the most common way of reducing the X-ray dose [11]. The quality of low-dose CT (LDCT) is partly limited by excessive quantum noise, resulting in degradation of diagnostic performance.

Over the last few decades, numerous iterative reconstruction (IR) methods for noise reduction have been proposed. Some of them have incorporated a statistical method such as Maximum a Posteriori (MAP) approach for data fitting in sinogram space. With the assumption that X-ray photons reaching to X-ray detector follows the Poisson distribution, the MAP model can be reduced to penalized weighted least squares (PWLS) optimization, including a term relating to the prior distribution of target images [17], [14]. Due to the difficulty in calculating the prior distribution, alternative priors reflecting spatial correlation among neighboring pixels, such as total variation (TV) [5], [19], fractional-order TV [21], nonlocal TV [15], Markov random fields theory [17], [14], [18] or nonlocal means [4], [9], [22] have been imposed to the target image. Although the IR method improves the overall image quality, designing effective prior reflecting characteristics of the target image is still challenging.

In this paper, we propose a deep PWLS method combined with the deep learning prior. We estimate the prior distribution in the sense of Pearson χ^2 divergence using many SDCT samples. The proposed model can be converted to the least squares-GAN (ls-gan) [10] that adopts the least squares loss function for discriminator, including a PWLS objective. The proposed model is reformulated as the constrained optimization problem, and solved by using the alternating direction method of multipliers (ADMM) algorithm [3]. Numerical simulations are performed to explore the validity of the proposed method.

H.S. Park and K. Jeon were with the National Institute for Mathematical Sciences, Daejeon, 34047, Korea.

Manuscript received XXX; revised XXX

II. Method

Let $\boldsymbol{y} \in \mathbb{R}^M$ be a measured X-ray photon counts reaching to detector, and let $\boldsymbol{x} \in R^N$ be an unknown attenuation coefficient to be determined. CT system tries to recover $\boldsymbol{x}$ from measured $\boldsymbol{y}$. Define $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{M \times N}$ by forward operator whose element a_{ij} denotes the effective intersection length of projection line i with voxel j. Considering the monochromatic X-ray, the expected number of measured X-ray photons at detector j, denoted by $\tilde{y}_i$, can be simply modeled as

$$\tilde{y}_i = I^0 \exp\left(-\sum_{j=1}^{M} a_{ij} x_j\right), \quad 1 \le i \le M, \tag{1}$$

where I^0 is the number of X-ray photons generated at source.

We start from maximum a-posteriori (MAP) approach. Given $\boldsymbol{y}$, find $\hat{\boldsymbol{x}}$ that maximizes the conditional probability $p_{\boldsymbol{x}^*|\boldsymbol{y}}(\hat{\boldsymbol{x}}|\boldsymbol{y})$:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \log p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y}) = \arg\max_{\boldsymbol{x}} \log p_{\boldsymbol{x}}(\boldsymbol{x}) + \mathcal{L}(\boldsymbol{y}|\boldsymbol{x}), \tag{2}$$

where $\mathcal{L}(\boldsymbol{y}|\boldsymbol{x}) = \log p_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{x})$. Here, the distribution $p_{\boldsymbol{x}}(\boldsymbol{x})$ is generally unknown. We estimate $p_{\boldsymbol{x}}(\boldsymbol{x})$ using many noise-free images $\boldsymbol{x}$ sampled from the $p_{\boldsymbol{x}}(\boldsymbol{x})$.

Let $\boldsymbol{z}$ be a noisy image reconstructed from $\boldsymbol{y}$, and let $G : \boldsymbol{z} \to \boldsymbol{x}$ be a noise reduction function. Together with many samples $\{z_k\}_{k=1}$ and $\{x_k\}_{k=1}$, we find G maximizing the expectation of $\log p_{\boldsymbol{x}|G(\boldsymbol{z})}(\boldsymbol{x}|G(\boldsymbol{z}))$, which can be expressed as [16]:

$$G_{\text{MAP}} := \arg\min_{\boldsymbol{\theta}_g} E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[-\log(p_{\boldsymbol{x}}(G(\boldsymbol{z};\boldsymbol{\theta}_g))) - \mathcal{L}(\boldsymbol{y}|G(\boldsymbol{z}))\right] \tag{3}$$

$$= \arg\min_{\boldsymbol{\theta}_g} -E_{G(\boldsymbol{z})\sim p_{G(\boldsymbol{z})}}\left[\log(p_{\boldsymbol{x}}(G(\boldsymbol{z};\boldsymbol{\theta}_g)))\right] \tag{4}$$

$$- E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[\mathcal{L}(\boldsymbol{y}|G(\boldsymbol{z};\boldsymbol{\theta}_g))\right]$$

$$= \arg\min_{\boldsymbol{\theta}_g} H(p_{G(\boldsymbol{z})}, p_{\boldsymbol{x}}) - E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[\mathcal{L}(\boldsymbol{y}|G(\boldsymbol{z};\boldsymbol{\theta}_g))\right], \tag{5}$$

where $H(\boldsymbol{x}, \boldsymbol{y})$ is the cross-entropy of $\boldsymbol{x}$ and $\boldsymbol{y}$. However, due to the difficulty in handling $H(p_{G(\boldsymbol{z})}, p_{\boldsymbol{x}})$ only from the samples $\{G(\boldsymbol{z}_i)\}_{i=1}^K$, we instead estimate $p_{\boldsymbol{x}}$ by using the Pearson χ^2 divergence $\chi^2_{Pearson}(p_{G(\boldsymbol{z})} + p_{\boldsymbol{x}}||2p_{\boldsymbol{x}})$. Now, the problem (5) can be reduced to

$$G_* := \arg\min_{\boldsymbol{\theta}_g} \chi^2_{Pearson}(p_{G(\boldsymbol{z})} + p_{\boldsymbol{x}}||2p_{\boldsymbol{x}}) - E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[\mathcal{L}(\boldsymbol{y}|G(\boldsymbol{z};\boldsymbol{\theta}_g))\right]. \quad (6)$$

It can be rewritten with adversarial loss [7], which is given by

$$G_* = \arg\min_{\boldsymbol{\theta}_g} E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[D(G(\boldsymbol{z};\boldsymbol{\theta}_g))^2 - \mathcal{L}(\boldsymbol{y}|G(\boldsymbol{z};\boldsymbol{\theta}_g))\right], \quad (7)$$

where D is the optimum of the following problem.

$$D = \arg\min_{\boldsymbol{\theta}_d} E_{\boldsymbol{x}\sim p_{\boldsymbol{x}}}\left[(1 - D(\boldsymbol{x};\boldsymbol{\theta}_d))^2\right] + E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[(1 + D(G(\boldsymbol{z});\boldsymbol{\theta}_d))^2\right]. \quad (8)$$

The proof of (7) is almost same as that of the paper [10]. The only difference is with or without the term $E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[\mathcal{L}(\boldsymbol{y}|G(\boldsymbol{z};\boldsymbol{\theta}_g))\right]$. For more details, we refer to Appendix V.

We optimize the problem (7) and (8) in alternating steps. For fixed D, we optimize the problem (7). To decouple the denoising function G from the term $\mathcal{L}$, we convert the problem (7) to the following constraint format:

$$\arg\min_{\boldsymbol{x}} E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[D(\tilde{\boldsymbol{x}})^2 - \mathcal{L}(\boldsymbol{y}|\tilde{\boldsymbol{x}})\right] \quad \text{s.t. } \tilde{\boldsymbol{x}} = G(\boldsymbol{z};\boldsymbol{\theta}_g). \quad (9)$$

Second order Taylor's expansion of $\mathcal{L}(\boldsymbol{y}_i|\tilde{\boldsymbol{x}})$ provides that [17]

$$\mathcal{L}(\boldsymbol{y}|\tilde{\boldsymbol{x}}) = -\frac{1}{2}\|\mathbf{A}\tilde{\boldsymbol{x}} - \mathbf{p}\|^2_{\mathbf{W}}, \quad (10)$$

where $\mathbf{p} \in \mathbb{R}^M$ is the projection data whose element $p_i = -\log(y_i/I^0)$, $\mathbf{W}$ is a diagonal matrix whose elements $w_i = y_i$. Here, $\|\boldsymbol{x}\|^2_{\mathbf{W}}$ is the weighted norm defined as $\|\boldsymbol{x}\|^2_{\mathbf{W}} = \boldsymbol{x}^T\mathbf{W}\boldsymbol{x}$. The problem (9) with (10) can be solved iteratively using ADMM algorithm [13].

Now, for fixed G, we solve the problem (8). the update of $\boldsymbol{\theta}_d$ is performed by using the Adam optimization.

A. Networks architectures & training

For generator (or noise reduction function), we adopt the deep convolutional framelets [20]. It is a multi-scale convolutional network network (CNN) that is composed of encoder and decoder paths with skipped and concatenation connections between them. The encoder and decoder paths are respectively composed of three stages. Each stage consists of two repeated convolutions (conv) with a 3×3 window. Each convolution is followed by a batch normalization (bnorm) and a leaky rectified linear unit (LReLU). 2-D Haar wavelet decomposition (wave-dec) and recomposition (wave-rec) are applied for downsampling and upsampling of the features, respectively. High pass filters after wavelet decomposition skip directly to the decoder path at same stage, while low pass filters (LF) are concatenated with the features in the decoder path. At the end of decoder path, an 1×1 convolution layer is added to generate a output with single channel.

For discriminator we adopted the standard CNN that consists of three convolution layers with a 4×4 window and strides of two in each direction of the domain. Each layer is followed by a batch normalization and a leaky ReLU with a slope of 0.2. At the end of the architecture, a 1×1 convolution layer is added to generate 1-dimensional output data.

In our simulations, the generator/discriminator training was performed using the Adam solver [8] with a learning rate of 0.0002, and mini-batch size of 1. Training was implemented using Tensorflow [1] on a CPU (Intel(R) Core(TM) i9-7900X, 3.30GHz) and a GPU (NVIDIA, Titan Xp. 12GB) systems. In total, it took about 7 days to train the proposed network. The network weights were initialized following a Gaussian distribution with a mean of 0 and standard deviation of 0.01. No data normalization and augmentation were performed during training.

III. Results

We used datasets provided by "the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge". Clinical CT scans of ten patients were included, and acquired at commercial CT scanners (Somatom Definition AS+, or Somatom Definition Flash operated in single-source mode, Siemens Healthcare, Germany) with a reference tube voltage of 120 kVp and a reference tube current of 200 mAs. Quarter-dose CT (i.e., LDCT) scans were generated by adding Poisson quantum noise to clinical SDCT scans. All projection data (or sinograms) obtained from the helical CT scanners with curve-detector were rebinned to the fan beam projection data with flat-panel detector. Standard filtered-backprojection [2] with the Ram-Lak filter was applied to reconstruct the LDCT and SDCT images of size 512×512. Of the total 10 patients, we used 8 patients for training, and the remaining 2 patients for testing.

The proposed method was compared with the PWLS method with total variation regularization (PWLS-TV) [19]. The model of PWLS-TV is given by

$$\arg\min_{\mathbf{x}} \; TV(\boldsymbol{x}) + \alpha\|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\|^2_{\mathbf{W}}, \quad (11)$$

where $TV(\boldsymbol{x})$ is the total variation given by $TV(\boldsymbol{x}) = \|\nabla\boldsymbol{x}\|_1$, and the regularization parameters $\alpha = 50$, which was nearly optimal in our simulation setting. As in the proposed method, we solved it by using ADMM method.

Fig. 1 shows the results of the proposed (third column) and the PWLS-TV (second column) for the AAPM test datasets. The second, fourth, sixth rows show the zoomed regions-of-interests (ROIs) marked by the rectangular regions in the first, third, and fifth rows, respectively. As shown in Fig. 1, LDCT images suffer from the noise streaking artifacts due to low radiation dose. Both methods significantly reduce the noise streaking artifacts from the LDCT images (first column). However, PWLS-TV method produces quite a blurry image owing to the inherent nature of its TV, while the proposed method clearly recovers the morphological structures of tissues in the body.

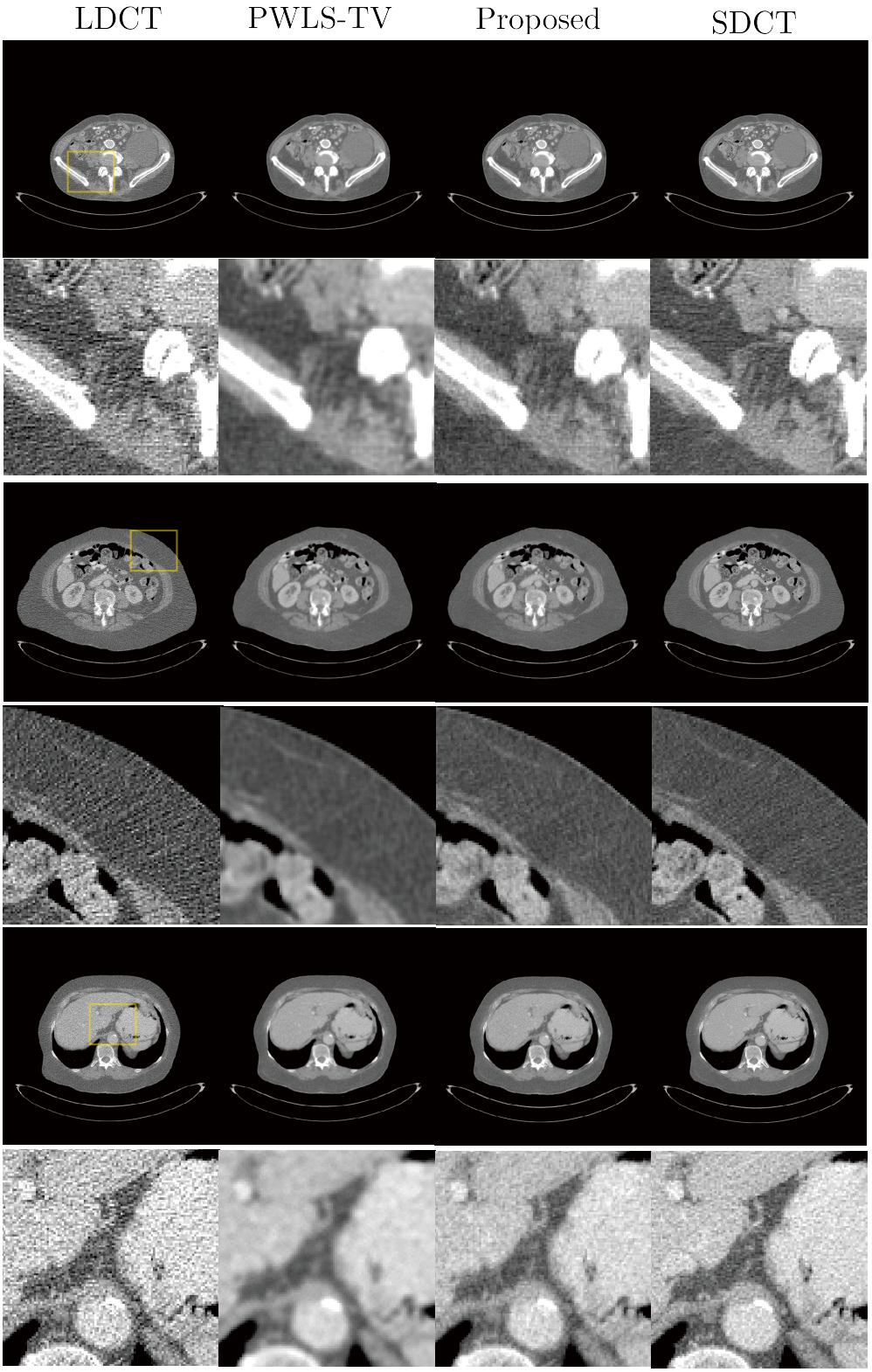


Fig. 1. Results of the proposed and other methods such as PWLS-TV, and IM-DL methods for the AAPM test datasets. Both methods significantly reduce the noise streaking artifacts in LDCT images. However, compared to the proposed method, PWLS-TV over-smoothed morphological structures of CT images. (C=0 HU/ W=800 HU for all CT images and C=0 HU/ W=500 HU for zoomed ROIs)

IV. Conclusion

In this paper, we propose a penalized weighted least squares method combined with a deep learning prior. We estimate the

prior distribution of target images in the sense of Pearson χ^2 divergence using many SDCT samples. The proposed model can be converted to the ls-gan with a PWLS objective, where the optimal generator is the denoising function that maps the LDCT images to its corresponding SDCT images. We reformulate the proposed model as constraint problem, and use ADMM method to decouple the denoising function and PWLS objective. By doing so, the standard backpropagation algorithm provided in Tensorflow can be used to learn the generator. While training is performed in an iterative way, the corrected image can be obtained directly from the trained generator. Numerical experiments show that the proposed method clearly reduces the quantum noise in LDCT images, while preserving morphological structures of the human body.

V. Appendix

In this Appendix, we prove the equation (7). Let $J_d(D)$ and $J_g(G)$ be functions given by

$$J_d(D) = E_{\boldsymbol{x}\sim p_{\boldsymbol{x}}}\left[1 - D(\boldsymbol{x})^2\right] + E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[(1 + D(G(\boldsymbol{z})))^2\right], \tag{12}$$

and

$$\begin{aligned} J_g(G) = {} & E_{\boldsymbol{x}\sim p_{\boldsymbol{x}}}\left[D(\boldsymbol{x})^2\right] + E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[D(G(\boldsymbol{z}))^2\right] \\ & - E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[\mathcal{L}(\boldsymbol{y}|G(\boldsymbol{z}))\right]. \end{aligned} \tag{13}$$

Given a fixed G, we first minimize $J_d(D)$. Writing both $G(\boldsymbol{z})$ and $\boldsymbol{x}$ to variable $\boldsymbol{\omega}$, optimal D_{opt} can be computed as

$$D_{opt}(\boldsymbol{\omega}) = \frac{p_{\boldsymbol{x}}(\boldsymbol{\omega}) - p_{G(\boldsymbol{z})}(\boldsymbol{\omega})}{p_{\boldsymbol{x}}(\boldsymbol{\omega}) + p_{G(\boldsymbol{z})}(\boldsymbol{\omega})}. \tag{14}$$

By plugging D_{opt} into $J_g(G)$, we have

$$\begin{aligned} J_g(G) = {} & E_{\boldsymbol{x}\sim p_{\boldsymbol{x}}}\left[D_{opt}(\boldsymbol{x})^2\right] + E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[D_{opt}(G(\boldsymbol{z}))^2\right] \\ & - E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[\mathcal{L}(\boldsymbol{y}|G(\boldsymbol{z}))\right] \qquad (15) \\ = {} & \int \frac{(p_{\boldsymbol{x}}(\boldsymbol{\omega}) - p_{G(\boldsymbol{z})}(\boldsymbol{\omega}))^2}{p_{\boldsymbol{x}}(\boldsymbol{\omega}) + p_{G(\boldsymbol{z})}(\boldsymbol{\omega})} d\boldsymbol{\omega} - E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[\mathcal{L}(\boldsymbol{y}|G(\boldsymbol{z}))\right] \qquad (16) \\ = {} & \chi^2_{Pearson}(p_{G(\boldsymbol{z})} + p_{\boldsymbol{x}} || 2p_{\boldsymbol{x}}) \\ & - E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[\mathcal{L}(\boldsymbol{y}|G(\boldsymbol{z}))\right], \qquad (17) \end{aligned}$$

where $\chi^2_{Pearson}(p,q) := \int \frac{(q(\boldsymbol{x})-p(\boldsymbol{x}))^2}{p(\boldsymbol{x})} d\boldsymbol{x}$. Since the first term in (13) is independent of G, minimizing $J_g(G)$ is equivalent to minimizing the equation (7), and this completes the proof.

Acknowledgment

This work was supported by the National Institute for Mathematical Sciences (NIMS) grant funded by the Korean government (No. NIMS-B20900000).

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, arXiv:160304467, 2016.

[2] R. N. Bracewell and A. C. Riddle. Inversion of fan-beam scans in radio astronomy. *The Astrophysical Journal*. 150. pp. 427, 1967.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn., 3, pp. 1–122, 2011.

[4] Y. Chen, D. Gao, C. Nie, L. Luo, W. Chen, X. Yin, and Y. Lin, Bayesian statistical reconstruction for low-dose x-ray computed tomography using an adaptive-weighting nonlocal prior, Computerized Medical Imaging and Graphics, 33, pp. 495–500, 2009.

[5] X. Dong, T. Niu, L. Zhu, Combined iterative reconstruction and image-domain decomposition for dual energy CT using total-variation regularization. Medical Physics, 41 pp. 051909, 2014.

[6] A.B. de González, and S. Darby, "Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries", The lancet, vol. 363, pp. 345–351, 2004.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozairy, A. Courville, Y. Bengio, Generative adversarial nets. In NIPS, 2014.

[8] D. P. Kingma, and J. Ba, Adam: A Method for Stochastic Optimization, arXiv:1412.6980, 2014.

[9] J. Ma, H. Zhang, Y. Gao, J. Huang, Z. Liang, Q. Feng, and W. Chen, Iterative image reconstruction for cerebral perfusion CT using a pre-contrast scan induced edge-preserving prior, Physics in Medicine and Biology, 57, pp. 7519–7542, 2012.

[10] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, Least Squares Generative Adversarial Networks, arXiv:1611.04076, 2017.

[11] L. F. McCaig, C.W. Burt, National Hospital Ambulatory Medical Care Survey: 2002 emergency department summary, Advance data, 340, pp. 1–34, 2004.

[12] D. L. Miglioretti, E. Johnson, A. Williams, R. T. Greenlee, S. Weinmann, L. I. Solberg, H. S. Feigelson, D. Roblin, M. J. Flynn, N. Vanneman, and R. Smith-Bindman, "The use of computed tomography in pediatrics and the associated radiation exposure and estimated cancer risk", JAMA Pediat., vol. 167, pp. 700–707, 2013.

[13] Y. Nesterov, A Method of Solving a Convex Programing Problem with Convergence Rate, Dokl. Akad. Nauk SSSR, 269, pp. 543–547, 1983.

[14] T. Li, X. Li, J. Wang, J. Wen, H. Lu, J. Hsieh, and Z. Liang, Nonlinear sinogram smoothing for low-dose x-ray CT, IEEE Transactions on Nuclear Science, 51, pp. 2505–2513, 2004.

[15] J. Liu, H. Ding, S. Molloi, X. Zhang, H. Gao, TICMR: Total Image Constrained Material Reconstruction via nonlocal total variation regularization for spectral CT. IEEE Transactions on Medical Imaging, 35, pp. 2578–2586, 2016.

[16] H.S. Park, J. Baek, S.K. You, J.K. Choi, and J.K. Seo, Unpaired image denoising using a generative adversarial network in X-ray CT, IEEE Access, 7, pp.110414-110425, 2019.

[17] K. Sauer, and C. A. Bouman, A local update strategy for iterative reconstruction from projections, IEEE Trans. Signal Processing, 41, pp. 534–548, 1993.

[18] J. Wang, T. Li, H. Lu, Z. Liang, Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. IEEE transactions on medical imaging, 25, pp. 1272–1283, 2006.

[19] Q. Xu, X. Mou,G. Wang, J. Sieren, E. A. Hoffman,and H. Yu, Statistical interior tomography, IEEE transactions on medical imaging, 30,, pp. 1116–1128, 2011.

[20] J. Ye, Y. Han, and E. Cha, "Deep convolutional framelets: a general deep learning framework for inverse problems", SIAM J. Imaging Sci., vol. 11, pp. 991–1048, 2018.

[21] Y. Zhang, Y. Wang, W. Zhang, F. Lin, Y. Pu, and J. Zhou, Statistical iterative reconstruction using adaptive fractional order regularization, Biomedical Optics Express, 7, pp. 1015–1029, 2016.

[22] H. Zhang, D. Zeng, J. Lin, H. Zhang, Z. Bian, J. Huang, Y. Gao, S. Zhang, H. Zhang, Q. Feng, Z. Liang, W. Chen, J. Ma, Iterative reconstruction for dual energy CT with an average image-induced nonlocal means regularization. Physics in Medicine and Biology, 62, pp. 5556–5574. 2017.

Advanced Scan Trajectories for Chest Tomosynthesis on Robotic X-Ray Systems

Chumin Zhao, Thomas Weber, Magdalena Herbst, Ludwig Ritschl, Sebastian Vogt, Steffen Kappler, Jeffrey H. Siewerdsen, and Wojciech Zbijewski

***Abstract*—Digital chest tomosynthesis (DCT) has the potential to improve visibility of lung structures compared to chest radiography. However, DCT exhibits loss of low frequency content, limited depth resolution and artifacts due to cardiac motion. The recent introduction of robotic x-ray systems may enable novel DCT scan trajectories with better sampling than the conventional unidirectional acquisitions. We use numerical x-ray system simulations and analytical imaging performance models to evaluate such novel orbits. In particular, we propose a cross-shaped orbit ("+"), which combines limited-angle rotations in the sagittal and axial planes, and a ring-shaped orbit ("O"), where the source and detector rotate in two coronal planes. The results demonstrate improved depth resolution, reduced artifacts, and more accurate lesion shape recovery using those advanced orbits, with best performance achieved by the ring trajectory. Lung nodule detectability computed from cascaded system analysis and numerical chest phantom simulations confirm adequate lesion visualization across a range of simulated exposures.**

***Index Terms*—Tomosynthesis, scan orbits, robotic x-ray systems, cascaded system analysis, detectability, lung nodule.**

I. Introduction

Early clinical studies of digital chest tomosynthesis (DCT) indicate that it might improve detection of small lung nodules compared to planar radiography [1], owing primarily to the ability to reduce out-of-plane anatomical clutter. Conventionally, DCT is performed with a stationary x-ray detector and an x-ray source that travels vertically, providing a set of posteroanterior (PA) projections over a limited range of view angles. Motivated by the recent introduction of robotic x-ray systems capable of fairly complex source-detector trajectories [2], we investigate alternative imaging orbits for DCT. Such advanced orbits might provide reduced image distortions and better depth resolution than the conventional linear scan, while maintaining the benefits of DCT in mitigating anatomical overlap in PA views of the lungs at relatively low radiation dose.

To facilitate the development of new DCT orbits, we use analytical imaging performance models based on cascaded system analysis [3] and an advanced numerical x-ray system simulator. We compare 3D resolution, noise, and detectability of pulmonary nodules for two unidirectional trajectories (a conventional vertical scan and a horizontal scan), for a cross-shaped combination of these two trajectories, and for a ring trajectory (Fig. 1), all with relatively low number of projection views (50-100). The simulation studies are parametrized based on the configuration of a twin-robotic x-ray system (Multitom Rax, Siemens Healthineers, Germany) [2], but the results are applicable to other x-ray devices with similar flexibility in the design of source-detector orbits, e.g. robotic C-arms. The results below focus primarily on the effects of scan geometry on DCT image quality; ongoing studies will investigate technical implementation of the most promising advanced DCT orbits identified here, including dose and scan time optimization. In particular, we will assess robustness to cardiac motion – often considered a challenge to clinical adoption of DCT - and benefits of algorithmic motion compensation [4].

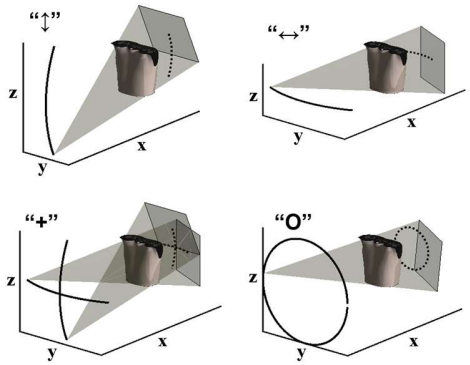


Fig. 1. Unidirectional and advanced multidirectional chest tomosynthesis trajectories for robotic x-ray systems: "↕" – vertical scan, "↔" – horizontal scan, "+" – cross trajectory (horizontal + vertical), and "O" – ring trajectory. Solid line indicates source motion, dashed line indicates the motion of the center of the detector. The coronal (PA) view commonly used in DCT corresponds to the *y*-*z* plane of the coordinate system.

II. Materials and Methods

A. Tomosynthesis Scan Trajectories

The studies assumed the configuration of a twin-robotic x-ray system (Multitom Rax) [2]. The x-ray source (120 kVp, 0.2 mm Cu filtration) and flat-panel detector (1420×1436 pixels, 0.3 mm pixel pitch) are mounted on individual robotized arms that provide independent translational and rotational motions.

We investigated four DCT orbits, all with source-axis distance (SAD) of 100 cm and axis-detector distance (ADD) of 40 cm. The trajectories (along with the coordinate system used throughout this work) are illustrated in Fig. 1: "↕" orbit – x-ray source and detector rotate in the *x*-*z* (sagittal) plane along a circular trajectory with a limited angular range of ±25° (50 views); "↔" similar to the "↕" orbit, but with rotation in the *x*-*y* (axial) plane; "+" (cross) orbit – a combination of "↕" and "↔" orbits (100 views); "O" (ring) orbit – a 360° rotation of the source and detector, each moving in a *y*-*z* (coronal) plane, with

C. Zhao, J. H. Siewerdsen, and W. Zbijewski are with the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21204, USA (e-mail: zhaocm@jhu.edu).

T. Weber, M. Herbst, L. Ritschl, S. Vogt and S. Kappler are with the Siemens Healthineers, Forchheim, Germany

the detector located diagonally from the source, i.e. with a 180° phase delay of its rotation, and with a fixed projection angle of 25° with respect to the *x*-axis (100 views). The detector performed an in-plane rotation throughout the "O" scan to maintain the detector rows (coordinate *u*) tangential to the path of the detector.

B. Numerical Simulation Studies

We used a high-fidelity numerical x-ray system simulator (previously validated for the Rax [5], [6]) to generate DCT data for the proposed orbits. Primary projections were computed using the Siddon method; x-ray scatter was estimated with a fast Monte Carlo (MC) [7]. Noise and noise correlations were added following a kernel-based approach [8]. To assess performance in a realistic setting, the simulations employed a digital chest model developed from a CT scan of the Lungman phantom (Kyoto Kagaku, Japan). Pulmonary nodules (~10 mm size, 100 HU) were digitally inserted in the central coronal plane of the chest (x=0); the lesions were spiculated [9] to provide both low- and high-frequency features (the latter will be affected by quantum noise) for comparison of the DCT orbits. The tube output was swept from 1 to 70 mAs per scan (nominal 50 mAs). The mAs/projection was therefore reduced by a factor of two for the "+" and "O" orbits (100 views) compared to the "↕" and "↔" scans (50 views) to ensure the same total mAs. Since this work focuses on the effects of DCT trajectory on 3D resolution and noise, we assumed an ideal scatter correction – however, the noise magnitude was based on the total signal (primary + scatter).

C. Tomosynthesis Reconstruction

Unfiltered backprojection (or shift-and-add), filtered backprojection (FBP) and iterative reconstruction methods have been used in tomosynthesis [10], [11]. Although unfiltered backprojection may be able to detect large spherical-shaped lung nodules, the fine details might be blurred due to limited depth resolution. We therefore adopted FBP reconstruction, implementing a Ramp filter with a Hann apodizer (cutoff at the Nyquist frequency) and a slice thickness filter (to remove noise aliasing) [12]. The Hann-apodized Ramp was applied in the direction tangential to the detector motion path, while the slice thickness filter was added in the depth direction (x). The reconstructed image voxels were 0.3 × 0.3 mm in the coronal plane (y-z), with a slice thickness (along x axis) of 1 mm.

D. Cascaded System Analysis of the Advanced DCT orbits

We developed a cascaded system analysis [3] model of the advanced DCT orbits to augment the numerical simulations with a rapid analytical estimation of 3D modulation transfer function (MTF) and noise power spectrum (NPS).

The input to the model includes detector MTF, $T_{2D}(f_u, f_v)$, and log-normalized projection NPS, $S_{2D}(f_u,f_v)$, obtained from the numerical x-ray system simulator.

The effect of the reconstruction process on the MTF and NPS were modeled as a series of transfer functions: $T_{\text{ramp}}(f_u, f_v) = |f_u|$ for the Ramp filter, $T_{\text{apd}}(f_u, f_v) = 0.5[1+\cos(2\pi f_u a_u)]$ for the Hann window, and $T_{\text{interp}}(f_u, f_v) = \text{sinc}^2(\pi f_u a_u)\cdot\text{sinc}^2(\pi f_v a_v)$ for interpolation prior to backprojection ($a_u = a_v = 0.3$ mm, and f_u is the direction tangential to the detector path) [3].

The 3D backprojection step introduced transfer function $T_{\text{bp}}(f_x, f_y, f_z) = M/f_r$, where M is the magnification and $f_r = M\cdot f_u$ [3]. This step involved a change of variables (Fourier slice theorem) from the projection frequencies (f_u, f_v, θ, where θ is the projection angle) to the object frequency domain (f_x, f_y, f_z).

For the "↕" orbit, the mapping from the projection to the object frequency space was:

$$\begin{cases} f_x = f_r \sin\theta = M \cdot f_u \sin\theta \\ f_y = M \cdot f_v \\ f_z = f_r \cos\theta = M \cdot f_u \cos\theta \end{cases}, \quad (1)$$

For the "↔" and "+" orbit, an analogous mapping was used, but with appropriate changes in the definition of the u axis – "vertical" for the "↕" orbit, "horizontal" for the "↔" orbit, and in the direction of the rotation θ (in the x-y plane for the "↕" orbit and in the x-z plane for "↔").

For the "O" orbit, the projection angle $\theta = 25°$ for all views. The object frequency spaced was sampled by rotating the detector in the y-z plane. We denote this rotation as φ to arrive at the following mapping:

$$\begin{cases} f_x = M \cdot f_v \sin\theta \\ f_y = M \cdot (-f_u \sin\varphi + f_v \cos\theta \cdot \cos\varphi). \\ f_z = M \cdot (f_u \cos\varphi + f_v \cos\theta \cdot \sin\varphi) \end{cases} \quad (2)$$

After 3D backprojection, the slice thickness filter was applied as a Hann window in the depth direction (x): $T_{\text{slice}}(f_x, f_y, f_z) = 0.5[1+\cos(2\pi f_x a_x)]$, where $a_x = 1$ mm is the slice thickness.

The final stages of the model included 3D sampling of the voxel matrix (III_{vox}) and post-reconstruction voxel binning (T_{bin} and III_{bin}) [3]. The resulting the 3D MTF and NPS were:

$$T_{3D}(f_x, f_y, f_z) = [T_{2D} T_{ramp} T_{apd} T_{interp}(f_u, f_v)][T_{bp} T_{slice} T_{bin}(f_x, f_y, f_z)], \quad (3)$$

$$S_{3D}(f_x, f_y, f_z) = \{S_{2D}[T_{ramp} T_{apd} T_{interp}(f_u, f_v)]^2 \frac{1}{M^2} \frac{\pi M^2}{m f_r} T_{slice}{}^2(f_x, f_y, f_z) *** III_{vox}\} T_{bin}{}^2(f_x, f_y, f_z) *** III_{bin}. \quad (4)$$

DCT is primarily viewed in coronal slice images, and thus the in-plane MTF (T_{ip}) is of particular interest. T_{ip} was obtained by integrating T_{3D} along the depth frequencies (f_x):

$$T_{ip}(f_y, f_z) = \int_{-f_{Nyq,x}}^{f_{Nyq,x}} T_{3D}(f_x, f_y, f_z) df_x. \quad (5)$$

E. Lung Nodule Detectability Index

The detectability (d') of a lung nodule was investigated assuming a non-prewhitening (NPW) observer [13]:

$$d'^2 = \frac{[\iint (\int T_{3D} \cdot W_{task} df_x)^2 df_y df_z]^2}{\iiint (T_{3D}^2 S_B + S_{3D}) df_x \cdot (\int T_{3D} \cdot W_{task} df_x)^2 df_y df_z}, \quad (6)$$

where S_{B} is the anatomical background noise obtained as 3D Fourier transform of nodule-free tissue clutter [14], and W_{task} is a task function given by Fourier transform of the lung nodule scaled by its contrast.

III. RESULTS

A. 3D MTF and NPS

Fig. 2 shows the 3D MTF $T_{3D}(f_x, f_y, f_z)$ (a-d, green) and NPS $S_{3D}(f_x, f_y, f_y)$ (e-h, blue) in the zero-frequency coronal, sagittal and axial planes for the "↕" (a, e), "↔" (b, f), "+" (c, g), and

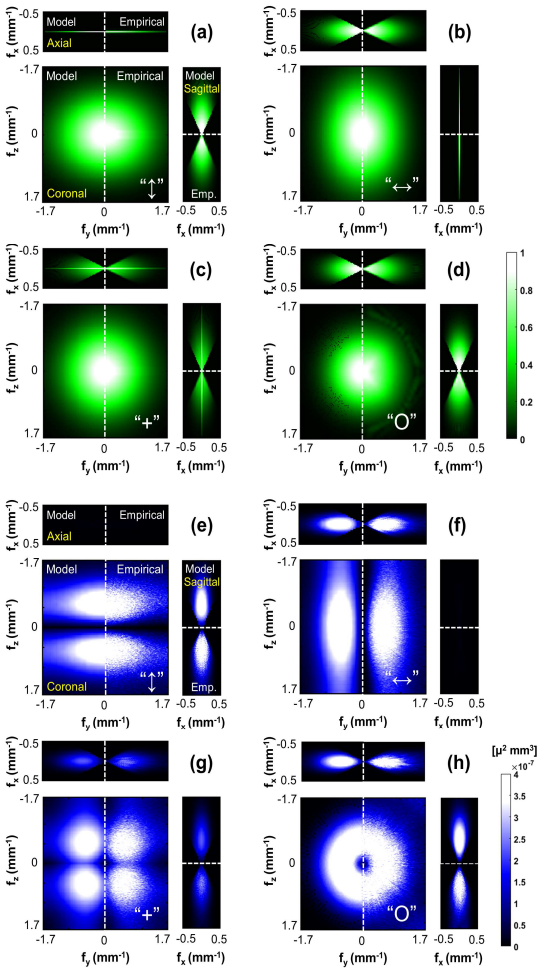


Fig. 2. Analytical models ("Model) and numerical simulations ("Empirical") of MTF (a – d) and NPS (e – f) in coronal ($f_x = 0$), sagittal ($f_y = 0$) and axial ($f_z = 0$) planes for the "↕" (a, e), "↔" (b, f), "+" (c, g) and "O" (d, h) orbits.

"O" (d, h) orbits. The 3D MTF and NPS obtained from the cascaded system model (left halves in the coronal and axial views, top half in sagittal view) are compared to the empirical MTF, measured from the numerical simulation of a 50 µm sphere impulse, and empirical NPS, measured from numerical simulation of air scans.

The "↕" and "↔" orbits (Fig. 2a, b) recover a wedge-shaped region of spatial frequencies, with wide frequency coverage in the coronal views ($f_x = 0$) and with severely limited MTF in the axial plane for the "↕" orbit and in the sagittal plane for the "↔" scan. The "+" orbit (Fig. 2c) combines the "↕" and "↔" frequency coverages, resulting in a degree of depth resolution in both sagittal and axial directions. The "O" orbit (Fig. 2d) achieves a similar frequency response as the "+" orbit, with somewhat improved MTF in the f_x and f_z directions, but slightly degraded coronal plane resolution.

The 3D NPS was computed at 50 mAs for all orbits (1 mAs/view for "↕" and "↔", and 0.5 mAs/view for "+" and "O"), resulting in similar noise magnitudes. However, the frequency distribution of noise differs between the trajectories. Most notable is the difference between the "+" and "O" scans, where the improved low-frequency axial and sagittal MTF of the "O" orbit results in a circularly symmetric coronal NPS.

B. In-plane and In-depth Resolution

The in-plane (coronal) MTF $T_{ip}(f_y, f_z)$ for each tomosynthesis orbit is shown in Fig. 3. Due to the limited angle acquisition and the filtering, the linear orbits exhibit a loss of low-frequency resolution in the f_z ("↕") or f_y ("↔") directions.

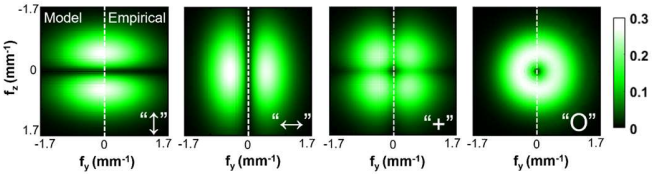


Fig. 3. In-plane coronal MTF for the "↕", "↔", "+" and "O" orbits (3D MTF integrated along f_x direction).

In Fig. 4(a), we show reconstructions of a numerical simulation of a 3 × 3 × 3 grid of spherical nodules (10 mm diameter, 40 mm distance, only the central nodule is displayed). The loss of low-frequency content in the "↕" and "↔" orbits results in strong edge enhancement in the scanning direction and distortions of the spherical shapes. We also observe the effects of limited depth resolution: shading artifacts in the coronal views (especially for "↕" and "+" orbits) that emanate from the neighboring nodules, as shown in Fig. 3(b) and (c). The "+" orbit yields reduced out-of-plane artifacts and improved shape recovery. The "O" orbit achieves the best depth-resolution among the investigated DCT trajectories, especially in f_x.

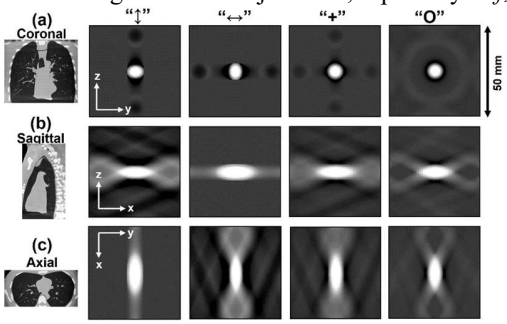


Fig. 4. In-plane (a – coronal view) and in-depth (b – sagittal view, and c – axial view) visualization of a 10 mm spherical insert placed in a 3 x 3 x 3 grid with 40 mm spacing.

C. Imaging Performance of the Chest Tomosynthesis Orbits

Fig. 5 shows reconstructed central slices of a numerical simulation of the chest phantom (50 mAs scans). The "↕" orbit resolves primarily the horizontal tissue structures, while the "↔" orbit detects predominately the vertical details. The "+" and "O" orbits show better removal of out-of-plane structures compared to the unidirectional scans (see yellow arrows). The "O" orbit provides slightly improved mitigation of out-of-plane artifacts (especially from the ribs) compared to the "+" scan with slightly reduced field-of-view at the periphery of the chest.

Fig. 6 shows a magnified view of a lung detail (yellow box in Fig. 5(a)), and a representative nodule selected from an area near the heart; Fig. 6 (a) uses the nominal dose (50 mAs), Fig. 6 (b) is at reduced exposure of 5 mAs/scan. A noise-free cone-beam CT (CBCT) reconstruction is provided as a reference. Detection of the nodule is challenged by the shading artifacts from the heart. This complex anatomy highlights the advantages of the "+" and "O" orbits in removal of out-of-plane structures and better preservation of the lesion shape compared

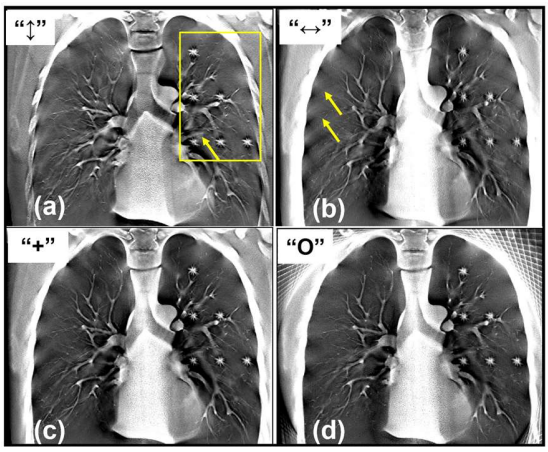


Fig. 5. Central coronal slice of DCT reconstructions of the chest phantom containing 10 mm spiculated lung nodules (highlighted in a) for the "↕" (a), "↔" (b), "+" (c), and "O" (d) orbits (50 mAs scans).

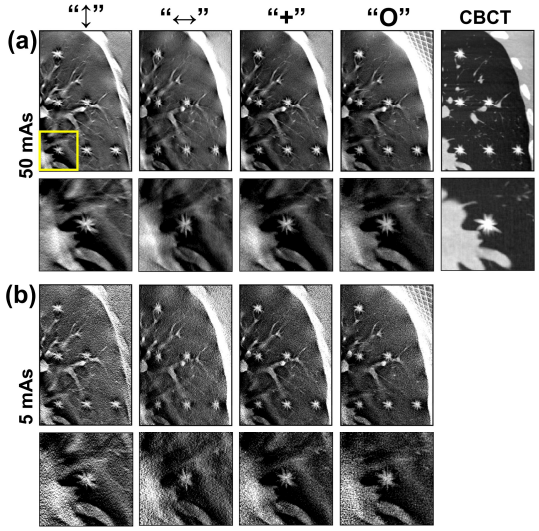


Fig. 6. Magnified views of a lung and of a selected pulmonary nodule for the different tomosynthesis orbits at the nominal (a, 50 mAs) and low (b, 5 mAs) dose. Noise-free CBCT images are shown for comparison.

to the unidirectional scans. The visualization of the lesions remains adequte after the 10-fold dose reduction.

D. Lung Nodule Detectability

Fig. 7 shows the detectability of the lung nodule visualized in Fig. 6 as a function of total scan exposure. The "O" and "+" trajectories achieve higher detectability mainly because of the reduced background noise (S_B) and better low frequency MTF. The theoretical *d'* estimates at the simulated 5 mAs exposure are only ~15% reduced compared to the 50 mAs scan, confirming the visual impression from Fig. 6(b).

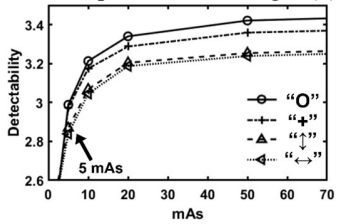


Fig. 7. Detectability index of the lung nodule of Fig. 6 at total scan exposures ranging from 1 to 70 mAs for the "↕", "↔", "+" and "O" orbits.

IV. Discussion and Conclusions

The recent introduction of robotic x-ray systems may enable development of advanced DCT orbits to improve lesion visibility while maintaining relatively low imaging dose. We are investigating the benefits and tradeoffs of such new scan trajectories using analytical and numerical x-ray system simulations. Our results quantified the gains in the "depth-direction" MTF and low-frequency MTF associated with the advanced multi-directional "+" and "O" orbits. As a result, the advanced trajectories achieved an appreciable reduction of out-of-plane artifacts and in-plane image distortions compared to conventional unidirectional scans. The detectability of lung nodules was thus improved. Those benefits of the advanced orbits were maintained across a broad range of scan exposures.

Ongoing work investigates the impact of cardiac motion in DCT with advanced orbits. Early results show reduced motion artifacts using projection-domain retrospective gating and reconstruction-domain cardiac phase registration [15].

This work is supported by research collaboration with Siemens Healthineers, XP Division. The presented method is not commercially available. Due to regulatory reasons, the Siemens Multitom RAX is not available in all countries, its future availability cannot be guaranteed.

References

[1] J. T. Dobbins *et al.*, "Digital tomosynthesis of the chest for lung nodule detection: Interim sensitivity results from an ongoing NIH-sponsored trial," *Med. Phys.*, vol. 35, no. 6, pp. 2554–2557, 2008.
[2] A. Fieselmann *et al.*, "Twin robotic x-ray system for 2D radiographic and 3D cone-beam CT imaging," in *Proc. SPIE 9783*, 2016, p. 97830G.
[3] D. J. Tward and J. H. Siewerdsen, "Cascaded systems analysis of the 3D noise transfer characteristics of flat-panel cone-beam CT," *Med. Phys.*, vol. 35, no. 12, p. 5510, 2008.
[4] A. Sisniega *et al.*, "Image-based deformable motion compensation for interventional cone-beam CT," in *Proc. SPIE 10948,* 2019, p. 109481O.
[5] C. Zhao *et al.*, "A robotic x-ray cone-beam CT system: trajectory optimization for 3D imaging of the weight-bearing spine," in *Proc. SPIE 10948,* 2019, p. 109481L.
[6] C. Zhao *et al.*, "Optimization of cone-beam CT scan orbits for cervical spine imaging," in *Proc. 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2019, p. 110720U.
[7] A. Sisniega *et al.*, "Monte Carlo study of the effects of system geometry and antiscatter grids on cone-beam CT scatter distributions," *Med. Phys.*, vol. 40, no. 5, p. 051915, 2013.
[8] A. S. Wang *et al.*, "Low-dose preview for patient-specific, task-specific technique selection in cone-beam CT," *Med. Phys.*, vol. 41, no. 7, p. 071915, 2014.
[9] J. Solomon and E. Samei, "A generic framework to simulate realistic lung, liver and renal pathologies in CT imaging," *Phys. Med. Biol.*, vol. 59, no. 21, pp. 6637–6657, 2014.
[10] J. Dobbins and H. McAdams, "Chest tomosynthesis: Technical principles and clinical update," *Eur. J. Radiol.*, vol. 72, no. 2, pp. 244–251, 2009.
[11] I. Sechopoulos, "A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications.," *Med. Phys.*, vol. 40, no. 1, p. 014302, 2013.
[12] B. Zhao and W. Zhao, "Three-dimensional linear system analysis for breast tomosynthesis.," *Med. Phys.*, vol. 35, no. 12, pp. 5219–5232, 2008.
[13] G. J. Gang *et al.*, "Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance.," *Med. Phys.*, vol. 38, no. 4, pp. 1754–1768, 2011.
[14] G. J. Gang, D. J. Tward, J. Lee, and J. H. Siewerdsen, "Anatomical background and generalized detectability in tomosynthesis and cone-beam CT," *Med. Phys.*, vol. 37, no. 5, pp. 1948–1965, 2010.
[15] C. Zhao *et al.*, "Digital chest tomosynthesis on robotic x-ray systems using retrospective cardiac gating and multidirectional scan trajectories," submitted to *RSNA* 2020.

Fluence Field Modulation in X-ray CT for ROI Imaging: Optimization of Primary Fluence and Comparison of Attenuator Scenarios

Sascha Manuel Huck[a,b], George S.K. Fung[c,d], Katia Parodi[b], Karl Stierstorfer[a]

***Abstract*—In our previous work we have presented a technical concept for fluence field modulation (FFM) in x-ray computed tomography (CT): the z-aligned sheet-based dynamic beam attenuator (z-sbDBA). We have built a prototype of the z-sbDBA and confirmed its capability by experiments on a clinical CT system installed with the prototype. [1]. Here, we first present a method to determine and optimize the primary fluence w.r.t. feasible z-sbDBA trajectories for a given scan scenario. The object is separated in two regions: a diagnostically relevant region-of-interest (ROI), where high image quality is desired, and the surrounding region, the organ-at-risk (OAR), which is supposed to be protected from radiation dose. Our trajectory optimization minimizes the product of i) the image variance inside the ROI and ii) the dose deposition in the OAR. Dose depositions are calculated from Monte-Carlo (MC) simulations. In a second step, we compare the z-sbDBA with other attenuators, including different shapes of bowtie filters combined with tube current modulation (TCM). At a prescribed level of image quality in the ROI, we compare the dose deposition caused in the OAR and the spatial distribution of the image variance across the attenuators. To summarize, the findings from our study show that 1) our method computes reasonable z-sbDBA trajectories as a prerequisite for ROI imaging. 2) The comparison with state-of-the-art bowtie filters demonstrates that the z-sbDBA is capable of reducing the OAR dose by about $20\,\%$ to $30\,\%$ at equal image quality. 3) The z-sbDBA is able to realize high image quality only where it is diagnostically important and effectively spares surrounding regions from radiation exposure.**

I. Introduction

A major challenge in computed tomography (CT) imaging is the weighting between sufficient image quality, necessary for gaining diagnostic information on the one hand, and the responsibility toward the patient to be protected from unnecessary radiation exposure on the other hand.

Different strategies for dose optimization in CT have been explored in literature [2], [3]. One of them is fluence field modulation (FFM) aiming to overcome the limitations of current technology – static bowtie filters combined with tube current modulation (TCM). In FFM, the primary fluence is dynamically modulated over the projections *and* over the fan beam angle. In the past, different hardware approaches have been proposed to facilitate FFM, including our sheet-based dynamic beam attenuator (sbDBA) [4]. Recently, we have presented an alternative concept, the z-aligned sbDBA (z-sbDBA) [1]. It constitutes a redesigned structuring and geometry of the x-ray attenuation sheets compared to the sbDBA. After verifying its feasibility for FFM, the question arose which primary fluence would be most suitable to cope with the above-mentioned weighting. Therefore, we developed a method to optimize the radiation exposure for a given scenario where a homogeneous phantom is divided into a region-of-interest (ROI) and an organ-at-risk (OAR). In addition, we compare the performances of different attenuator scenarios in terms of radiation dose and image quality.

[a]Siemens Healthcare GmbH, Forchheim, Germany.
[b]Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München, Garching, Germany.
[c]Siemens Medical Solutions USA, Inc., Malvern (PA), USA.
[d]Russell H. Morgan Department of Radiology and Radological Science, Johns Hopkins University, Baltimore (MD), USA
Contact: sascha.huck@siemens-healthineers.com

II. Methodology

A. The z-sbDBA concept

Our novel concept for fluence field modulation, the z-sbDBA, has been presented in our previous work [1]. It consists of an array of alternating (almost) x-ray opaque and x-ray transparent sheets, located in the beam path in proximity (about $100\,\mathrm{mm}$) to the x-ray source. The sheets are aligned parallel to the fan beam plane and focused toward the focal spot. To enable FFM over the fan beam angle β, the height of the sheets is minimal at the center of the fan beam and increases for larger fan beam angles, see fig. 1. The transmission profiles $t(\vartheta)$ at different attenuation angles ϑ of the z-sbDBA are known from previous experimental measurements. The data acquisition is reported in detail in Huck *et al.* [1]. The transmission profiles are the necessary input data for the trajectory optimization in section II-C.

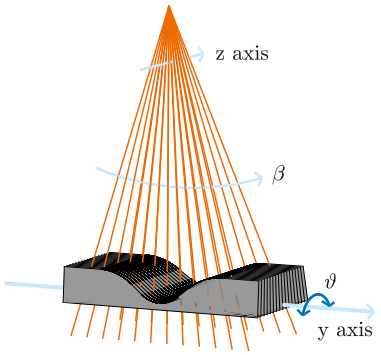


Fig. 1: A schematic drawing of the novel z-sbDBA concept.

The optimization of the trajectory and the attenuator comparison is divided into three major steps as described in sections II-B to II-D.

B. Phantom and simulation data base

In our method, experimentally measured data was combined with simulation data. So, the scanner geometry used for the measurements (Siemens SOMATOM Force, Siemens Healthcare GmbH, Forchheim, Germany) was mimicked in the simulations. In our sample scan scenario, we considered a centered, homogeneous water phantom (semi axes: $160\,\mathrm{mm} \times 120\,\mathrm{mm}$, length: $300\,\mathrm{mm}$). It was separated into the central ROI (semi axes: $80\,\mathrm{mm} \times 40\,\mathrm{mm}$) and the surrounding OAR, see fig. 2d.

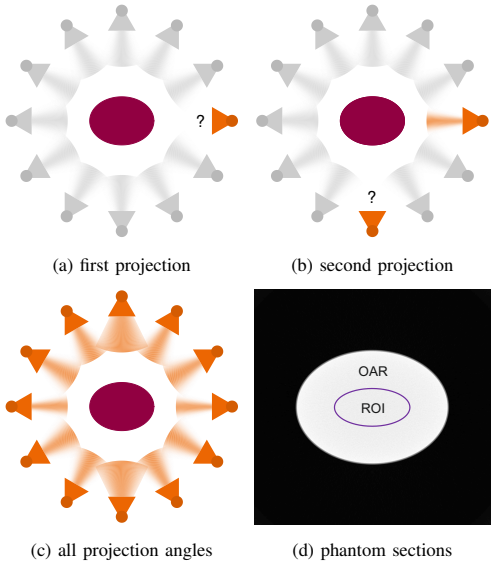
1) Phantom Dose Matrix: A dose matrix $M(\alpha, \beta, v)$ of the phantom, containing the dose contribution at each fan beam angle β and at each projection angle α to a specific dose voxel v of the phantom, was calculated. For this, the phantom was voxelized ($2\,\mathrm{mm} \times 2\,\mathrm{mm}$) in axial plane and – to also account for scatter radiation – the full phantom length in z-direction. In individual Monte-Carlo (MC) simulations, 10^6 photons (typical $120\,\mathrm{kV}$ CT spectrum) were propagated for each combination of fan beam angle ($\Delta\beta = 1°$, $\beta \in [-24°, 24°]$) and projection angle ($\Delta\alpha = 5°$, $\alpha \in [0°, 355°]$). The MC simulations were conducted with *MOCASSIM* (Siemens Healthcare GmbH, Forchheim, Germany) [5], a MC software package dedicated for diagnostic x-rays.

2) Phantom Attenuation Sinogram: The same setup was considered to calculate the attenuation sinogram $L(\alpha, \beta)$ of the phantom. It was simulated using *DRASIM* (Siemens Healthcare GmbH, Forchheim, Germany), a x-ray transmission software package. For a proper noise evaluation, Poisson noise was added according to the phantom attenuation and the propagated primary fluence Φ. Images were reconstructed with filtered backprojection.

Eventually, the dose distribution in the phantom and the image quality in the reconstructed image can be simulated for an arbitrary primary fluence $\Phi(\alpha, \beta)$ produced by an attenuator.

C. Optimization of the z-sbDBA trajectory

Our method optimizes the z-sbDBA trajectory $\vartheta(\alpha)$ such that the optimization criterion, the product of i) the resulting image variance inside the ROI and ii) the radiation dose deposition in the surrounding OAR, is minimized. Therefore, the primary fluence Φ was varied as described below. For each sample primary fluence, the relevant factors – the ROI variance and the OAR dose – and eventually the optimization criterion were calculated. TCM according to Gies, Kalender, Wolf, *et al.* [6] was incorporated. The method calculates the optimization criterion projection-wise to compile the full trajectory. Initially, a moderate attenuation angle $\vartheta = 2.4°$ was set for projection angles α not yet optimized. For the projection angle currently being optimized (orange), different transmission profiles $t(\vartheta)$ were evaluated regarding the optimization criterion while other projections (gray) remained unchanged, as shown in fig. 2a.



(a) first projection (b) second projection

(c) all projection angles (d) phantom sections

Fig. 2: (a)-(c) exemplifies the projection-wise determination of the best-matching attenuation angle ϑ per projection α for a given object (purple). Note the different fan beam profiles. (d) visualizes the phantom segmentation in a central ROI and a surrounding OAR region.

The possible range of attenuation angles at the current projection was constrained by the attenuation angles of the neighboring projections plus a maximum difference of $\Delta\vartheta = 0.5°$. Then, the attenuation angle ϑ providing the minimum result in terms of the optimization criterion was identified and kept for the evaluations of the remaining projections. Subsequently, the next projection angle – displaced by about $90°$ – was evaluated following the same scheme, see fig. 2b. After completing all projections, shown in fig. 2c, the full trajectory was smoothed. Both the smoothing and the range constraint aim to reduce a) strongly fluctuating trajectories, b) avoid unreasonable large angular steps between adjacent projections and c) promote a cyclic shape. The described scheme was repeated in multiple iterations until the trajectory remains unchanged compared to the previous iteration or a maximum number of iterations is reached.

D. Comparison of attenuator scenarios

In order to assess the z-sbDBA with respect to present technical solutions, the z-sbDBA was compared with a standard bowtie filter and a more narrow bowtie filter dedicated to cardiac applications – all combined with TCM [6]. Furthermore, two scenarios corresponding to a CT without attenuator, thus emitting homogenous fluence over the fan angle, were part of the comparison. Scenario 1 is similar to scenario 2 but with TCM activated. An overview of all considered scenarios is summarized in table I.

The primary fluence of the z-sbDBA was adopted from the results of the trajectory optimization described in sections II-C and III-A and the ones of the bowtie filters were derived from measurements [1].

For the comparison, we have simulated CT scans of the described phantom, fig. 2d, at equal image quality. Hence, the OAR dose, required to achieve the prescribed image quality,

was compared among the attenuator scenarios. Here, image quality was considered as the image variance inside the ROI of the difference image, resulting from two noise-independent reconstructions. In a separate image quality analysis of the two bowtie filters and the z-sbDBA, the variance at each pixel position was calculated based on 100 noise-independent difference images. The filter specific primary fluence was adopted from the first part of the comparison for the calculation of the variance maps.

TABLE I: Overview of the attenuator scenarios.

no.	name	attenuator	TCM	filter type
1	*noAtt_tcm0*	no	0 (off)	n/a
2	*noAtt_tcm1*	no	1 (on)	n/a
3	*bowtie1_tcm1*	standard bowtie	1 (on)	static
4	*bowtie2_tcm1*	narrow bowtie	1 (on)	static
5	*z-DBA_tcm1*	z-sbDBA	1 (on)	dynamic

III. Results and Discussion

A. Optimization of the z-sbDBA trajectory

The presented method for optimizing the z-sbDBA trajectory is applied to the described phantom. The resulting trajectory, optimizing the primary fluence, is shown in fig. 3. At projections α around $0°$, $180°$ or $355°$, the attenuation angle ϑ is larger compared to projections around $90°$ or $270°$. Accordingly, a relatively narrow transmission profile is propagated at projections facing the shorter side of the ROI and a broader profile when facing the wider side. The overall trajectory appears plausible for the considered setup and a cyclic, smooth course is realized.

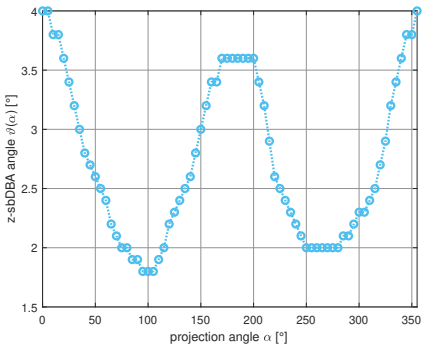


Fig. 3: The resulting z-sbDBA trajectory $\vartheta(\alpha)$ as a function of the projection angle α.

B. Comparison of attenuator scenarios

The dose deposition in the OAR caused to achieve a certain level of variance inside the ROI is compared among the scenarios, see fig. 4. For the given setup, TCM (*noAtt_tcm1*) reduces the OAR dose by $9\,\%$ w.r.t. the native, reference scenario (*noAtt_tcm0*). Using the standard bowtie (*bowtie1_tcm1*) in addition decreases the OAR dose by $36\,\%$. The narrow bowtie (*bowtie2_tcm1*) instead improves the dose deposition by $42\,\%$ compared to the reference scenario. The best result is achieved when using the z-sbDBA (*z-DBA_tcm1*) which is shown capable of improving the OAR dose exposure by $53\,\%$ over the reference scenario at same image quality. In

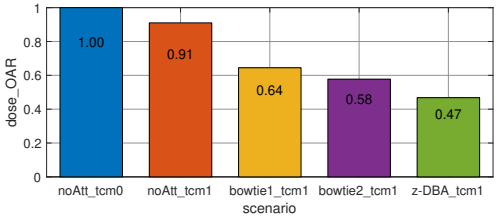


Fig. 4: The resulting dose deposition to the surrounding OAR for each attenuator scenario at equal image variance inside the ROI (normalized w.r.t. *noAtt_tcm0*).

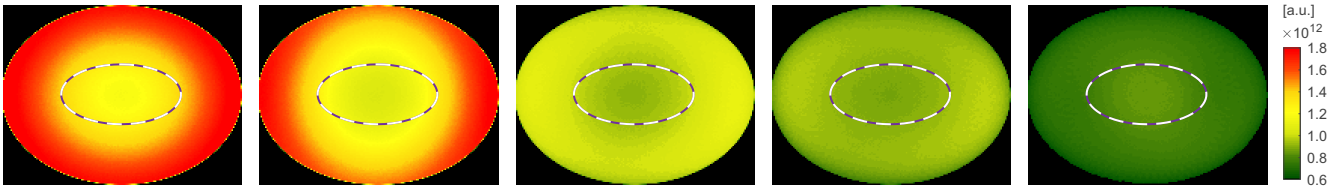
comparison to the standard or the narrow bowtie, it improves the OAR dose by $27\,\%$ or $19\,\%$, respectively. FFM not only over the projections but also over the fan beam angle – as facilitated by the z-sbDBA – can considerably lower patient radiation dose in ROI imaging.

1) Dose maps: On the basis of the primary fluence optimized to achieve equal ROI image variance, as explained in section III-B, radiation dose to the considered phantom is portrayed in dose maps, see fig. 5. Additionally, line plots along the short phantom axes are depicted in fig. 6. The overall phantom dose is the largest in scenarios without attenuator (*noAtt_tcm0*, *noAtt_tcm1*). Especially peripheral regions (e.g. skin) are exposed. Both the static bowtie filters (*bowtie1_tcm1*, *bowtie2_tcm1*) can noticeably reduce the radiation exposure. The z-sbDBA reveals the lowest phantom dose, though. Likewise, it is remarkable that only the z-sbDBA is able to reduce the radiation dose toward peripheral regions (OAR) compared to the diagnostically relevant ROI. The findings demonstrate that the z-sbDBA can shift radiation dose to regions where high image quality is required and reduce the radiation exposure of radio-sensitive regions.

2) Variance maps: Figures 7a to 7c visualize the image variance in the reconstructed image calculated from 100 noise-independent difference images. The variance along the long phantom axes is depicted in fig. 7d. The general shape is similar along the short semi axes. By construction, the variance is equal inside the ROI across the filter scenarios. In both bowtie scenarios, the variance continuously decreases toward the diagnostically irrelevant phantom edges. In contrast, the variance approximately doubles in the OAR compared to the ROI when using the z-sbDBA. This, in particular, emphasizes that high image quality is only realized in diagnostically relevant regions while radiation dose is reduced in the surrounding OAR.

IV. Conclusion

In order to operate a CT scanner equipped with a z-sbDBA we present a method to calculate the optimal primary fluence for a given scan scenario and determine the corresponding filter trajectory. The optimization method for FFM aims to reduce the image variance in a ROI but also the radiation dose delivered in an OAR. Our findings in this work demonstrate that 1) the optimization method calculates appropriate z-sbDBA trajectories for a given scan scenario. Furthermore, the comparison with other technologies emphasizes the beneficial properties of the z-sbDBA when realizing



(a) noAtt_tcm0 (b) noAtt_tcm1 (c) bowtie1_tcm1 (d) bowtie2_tcm1 (e) z-DBA_tcm1

Fig. 5: Resulting Dose maps for different attenuator scenarios visualizing the spatial dose distribution inside the phantom at equal image variance inside the central ROI (dashed ellipse).

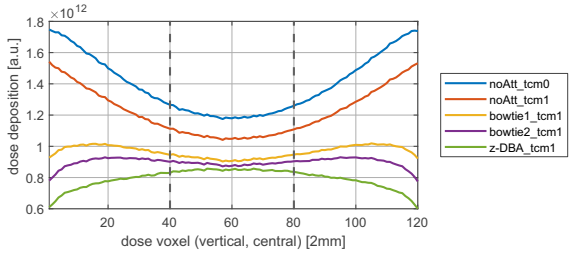
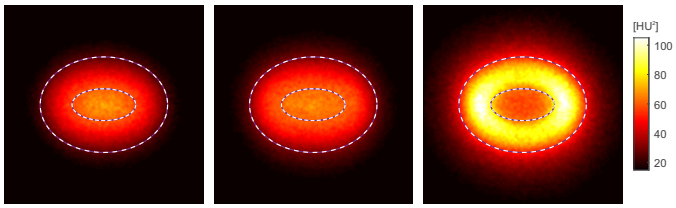
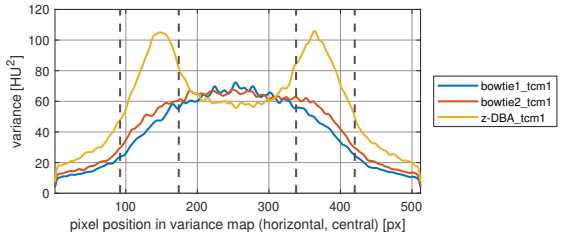


Fig. 6: The dose deposition along the short semi axis of the phantom. The ROI is located in between the dashed lines.



(a) bowtie1_tcm1 (b) bowtie2_tcm1 (c) z-DBA_tcm1



(d) line plots across the variance maps

Fig. 7: Figures 7a to 7c: Resulting variance maps visualizing the spatial image variance in the reconstructed image of the phantom (outer ellipse) and the central ROI (inner ellipse) in HU^2. Figure 7d: Variance along the long semi axes of the phantom. The ROI or the phantom edges are indicated by the inner or the outer dashed lines, respectively.

the optimized fluence: 2) Compared to state-of-the-art bowtie filteres, the z-sbDBA reduces the OAR dose by about 20 % to 30 % at equal image quality. 3) The z-sbDBA is able to deliver high image quality where it is really needed and preserve surrounding tissue from radiation dose. Building upon these findings, future work should include phantoms with anthropomorphic structures and investigations on off-centered ROIs or patients. Furthermore, a comparison with the sbDBA approach presented previously [4] might explore particular fields of applications and demonstrate individual strengths of the respective filters.

References

[1] S. M. Huck, G. S. Fung, K. Parodi, and K. Stierstorfer, "A new concept for fluence field modulation in x-ray ct: The z-sbdba," in *Medical Imaging 2020: Physics of Medical Imaging*, G.-H. Chen and H. Bosmans, Eds., International Society for Optics and Photonics, vol. 11312, SPIE, 2020, p. 1 131 202. DOI: 10.1117/12.2543185.

[2] M. K. Kalra, M. M. Maher, T. L. Toth, L. M. Hamberg, M. A. Blake, J.-A. Shepard, and S. Saini, "Strategies for ct radiation dose optimization," *Radiology*, vol. 230, no. 3, pp. 619–628, 2004.

[3] S. Bartolac, S. Graham, J. Siewerdsen, and D. Jaffray, "Fluence field optimization for noise and dose objectives in ct," *Medical physics*, vol. 38, no. 1, pp. 2–17, 2011.

[4] S. M. Huck, G. S. Fung, K. Parodi, and K. Stierstorfer, "Sheet-based dynamic beam attenuator–a novel concept for dynamic fluence field modulation in x-ray ct," *Medical physics*, vol. 46, no. 12, pp. 5528–5537, 2019. DOI: 10.1002/mp.13690.

[5] S. Kappler, D. Niederlöhner, S. Wirth, and K. Stierstorfer, "A full-system simulation chain for computed tomography scanners," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE*, IEEE, 2009, pp. 3433–3436.

[6] M. Gies, W. A. Kalender, H. Wolf, C. Suess, and M. T. Madsen, "Dose reduction in ct by anatomically adapted tube current modulation. i. simulation studies," *Medical physics*, vol. 26, no. 11, pp. 2235–2247, 1999.

Deterministic Boltzmann Transport Equation Solver for Patient-specific CT Dose Estimation: Validation for Realistic Scanner Configurations and Patient Models

S. Principi, A. Wang, A. Maslowski, T. Wareing, P. Jordan, T. G. Schmidt

***Abstract*—The purpose of this work was to validate and optimize a linear Boltzmann Transport Equation (LBTE) solver for CT dose estimation across a range of realistic scanner configurations and patient models. The studied software tool rapidly and deterministically solves the LBTE and is proposed as an alternative to Monte Carlo for patient-specific organ dose estimation. In this study, methods were developed for modeling CT scanner complexities within the LBTE framework, including bowtie filter, over range collimation, and tube current modulation. The LBTE solver requires discretization in the spatial, angular, and spectral dimensions, which may affect the accuracy of scanner modeling. To investigate these effects, this study evaluated the LBTE dose accuracy for different discretization parameters, scanner configurations and patient models (male, female, adults, pediatric). The results were validated against the Monte Carlo simulation code Geant4, which provided ground truth energy deposition maps. Results showed a very good agreement between the two methods, with root mean square error in organ dose below 3% for all the studied configurations. The LBTE accuracy was similar across the different protocols and phantoms. A relevant aspect was the short execution time. The LBTE simulations, implemented on a GeForce GTX 1080 graphic unit, required less than one minute for all cases. In conclusion, this study validated the LBTE solver for organ dose estimation considering realistic aspects of CT scanners and a range of phantom models. Future plans include the validation of the deterministic solver against physical measurements.**

***Index Terms*—Organ dose estimation, CT scanner modeling, Deterministic solver, Monte Carlo simulation.**

I. Introduction

THE risk related to the radiation exposure of CT scans is a matter of wide concern. The CTDI (CT dose index) is the mandated dose metric in CT dose reports. However, it is an inadequate metric of patient radiation dose, as it represents the average dose to a uniform plastic cylindrical phantom, which is not an appropriate model of the patient's anatomy. The alternative to provide accurate and patient-specific dose maps is Monte Carlo (MC) computational methods. MC represents the gold standard for radiation dose estimates and is widely used among the research scientific community. Whereas it was previously a computationally expensive tool, recent implementations of MC code on GPUs have drastically improved speeds. The alternative method for fast dose maps investigated in this work is Acuros CTD (Varian Medical Systems, Palo Alto, CA), a linear Boltzmann transport equation (LBTE) solver. Acuros CTD and MC both solve the Boltzmann transport equation, but they use different approaches. MC is a stochastic solver whose accuracy depends on the number of particles simulated. Acuros CTD is a deterministic solver. Thus, the problem is discretized in energy, space and angle, and the accuracy of the LBTE solutions depends on the level of discretization of these three parameters. The level of discretization may be limited by memory size. Acuros CTD was previously validated for simple CT scan configurations in Wang *et al.*'s work [1] using a reference dataset defined in the AAPM TG-195 report [2]. However, this previous study did not represent realistic scanner complexities that impact patient dose, such as over range collimation and tube current modulation (TCM). The modeling of these scanner complexities in Acuros CTD may be impacted by the level of discretization. To investigate this question, this work first developed methods to model scanner complexities within the Acuros CTD framework. Then, in order to verify the accuracy of the CT system modeling considering the required discretization, the energy deposition maps were validated against the gold standard MC method. Also, to broaden the validity of our study, different scanner parameters were varied, such as tube collimation, spectral distribution of the beam and TCM. It was also validated for a variety of anthropomorphic phantoms.

II. METHODS

The present study implements modern scanner aspects such

Submitted 19 January 2020.
This work was supported by the NIH Grant U01EB023822.
S. Principi and T. G. Schmidt are with the Biomedical Engineering Department, Marquette University and Medical College of Wisconsin, Milwaukee, WI, USA (e-mail: sprincipi@mcw.edu; tal.gilat-schmidt@marquette.edu).
A. Wang is with Stanford University, Stanford, CA, USA (e-mail: adamwang@stanford.edu)
A. Maslowski, T. Wareing, P. Jordan are with Varian, Palo Alto, CA, USA (e-mail: Alex.Maslowski@varian.com; Todd.Wareing@varian.com; Petr.Jordan@varian.com)

as the bowtie filter, the over range collimation and the TCM, both in Acuros CTD and in MC. The MC code Geant4 version 10.4.2 was used as the gold standard method and as our benchmark code for the validation of the LBTE results. Acuros CTD is based on the core algorithm of Acuros XB, which is a commercially available software by Varian Medical System for radiation therapy planning. In Acuros, each x-ray beam is defined by a discrete spectrum of 6 energy bins. In Geant4 the beam fluence varied continuously in energy and 10^7 photons were generated per each view in order to obtain results with statistical uncertainty below 1% for most organs. For the bowtie filter, a mathematical model of a medium aluminum bowtie was used. In Acuros, the spectral distribution of the beam varies discretely across the fan angle, whereas in MC the implementation is continuous. For the over range (OR) collimation, the beam varies across the cone angle. Because of the different discretization in space for the two codes (200 views per rotation in Geant4 and 18 in Acuros), 448 over range collimation adjustments were implemented in Geant4 whereas only 40 adjustments were implemented in Acuros. An existing MATLAB custom program was used to generate the longitudinal and angular tube current modulation profiles as a function of the projection angle. Again, given the different sampling of the views for Acuros and MC, only 18 current modulations per rotation were performed for Acuros versus 200 for MC. Furthermore, different parameters were varied to generate several configurations, as described in detail below. For all cases, a helical scan with a pitch of 0.9 was simulated including the bowtie filter, angular and longitudinal tube current modulation, and overrange collimation.

A. *Optimal voxel downsampling*

The voxelized anthropomorphic phantom employed in the AAPM Task Group 195 was used as a test case. The ground-truth Geant4 implementation used the original voxel sampling of 1x1x1 mm^3. Because Acuros is limited by the memory of the GPU, the phantom was down sampled to voxels of dimensions of: 10x8x1, 8x8x2, 5x5x5 and 4x5x4 mm^3. Relevant parameters for this study are listed in Table I.

TABLE I
STUDY A

2-cm collimation (120 kV)

	Acuros	MC
No.views per rotation	18	200
No. rotations	8	8
Scan length traveled per view (cm)	0.1	0.009
Phantom	AAPM (≠ voxel size)	AAPM (1x1x1mm^3 voxel)

4-cm collimation (120 kV)

	Acuros	MC
No.views per rotation	18	200
No. rotations	4	4
Scan length traveled per view (cm)	0.2	0.018
Phantom	AAPM (≠ voxel size)	AAPM (1x1x1mm^3 voxel)

B. *Optimal number of views per rotation*

Wang *et al*. work [1] suggested that 18 views per rotation are a good compromise between accuracy and low simulation time. However, in this study several relevant aspects of modern CT scanners are implemented. Both the TCM and overrange collimation models depends on the discretization in angle. Therefore, a higher number of projection angles might be required to achieve accuracy when considering these scanner effects. This study, described in Table II, investigated Acuros for various number of views per rotation: 4, 5, 8, 12, 18, 24, 30, 40 and 200.

TABLE II
STUDY B

4-cm collimation (120 kV)

	Acuros	MC
No.views per rotation	4, 5, 8, 12, 18, 24, 30, 40, 200	200
No. rotations	4	4
Scan length traveled per view (cm)	0.9, 0.72, 0.45, 0.3, 0.2, 0.15, 0.12, 0.09, 0.018	0.018
Phantom	AAPM (5x5x5 mm^3)	AAPM (1x1x1mm^3)

C. *Different source spectra*

Acuros defines the spectral distribution in six energy groups. This study investigated tube spectra with peak voltages of 80, 100 and 140 kV in addition to the previously modeled 120-kV beam. The configuration for this study is given in Table III.

TABLE III
STUDY C

4-cm collimation (**80 kV, 100 kV, 140 kV**)

	Acuros	MC
No.views per rotation	18	200
No. rotations	4	4
Scan length traveled per view (cm)	0.2	0.018
Phantom	AAPM (5x5x5 mm^3)	AAPM (1x1x1mm^3)

D. *Different anthropomorphic phantoms*

The purpose of this study was to validate Acuros across a range of patient model sizes. The tool XCAT2 from Duke University was used to generate voxelized anthropomorphic phantoms, one pediatric and two adults. The original phantoms have isotropic voxels of 1.5 mm size. In Geant4 this voxel size was maintained whereas in Acuros the voxels were down

sampled, as specified in Table IV, where height, weight and BMI percentiles for each patient model are also listed. The

TABLE IV
STUDY D

Phantom	Height (cm) Weight (kg)	BMI	kV
Female_Ped	119.5 cm 22.9 kg	16.04	80
Female_Adult	162 cm 66.4 kg	25.30	100
Male_Adult	176.5 cm 100 kg	32.10	120

down sampling for the new phantoms was 4.5x4.5x4.5 mm^3 for the pediatric and 6.0x4.5x4.5 mm^3 for the adults, due to memory constraint. The protocols from Table III were applied.

III. RESULTS

A. *Optimal voxel down sampling*

In Acuros four different voxel sizes were implemented, as specified in section II, with Geant4 providing the ground truth with voxels of 1x1x1 mm^3. The resulting dose maps were similar (Figure 1), although stochastic noise is evident in the Geant4 maps. The measure used to quantify the agreement between the two methods was the root-mean-squared organ dose error (RMSE), excluding the organs that absorb less than 0.1% of the total deposited energy. The RMSE between the Geant4 and Acuros implementation varied from 1.5% (for voxels dimension = 4x5x4 mm^3) to 1.9% (for voxels dimension = 10x8x1 mm^3) for the 2-cm collimation configuration, whereas RMSE varied from 1.8% (for voxels dimension = 4x5x4 mm^3) to 2.1% (for voxels dimension = 10x8x1 mm^3) for the 4-cm collimation. The organ with the highest error was the breast. For the 2-cm collimation, the breast error varied from 2.8 to 3.8%, whereas it ranged from 3.7% to 4.4% for the 4-cm collimation. Thinner voxels along the Z axis (8x8x2 and 10x8x1 mm^3) resulted in higher accuracy for organs with low exposure, such as the adrenals, as shown in Figure 2. The relative difference for adrenal is of 2.5% for the 10x8x1 mm^3 voxel and it increases up to 35% for the 5x5x5 and 4x5x4 mm^3 down sampling. However, for this organ the maximum absolute error was only 0.029 eV/photon (vs. total energy deposited of 7.8E3 and 7.2E3 eV/photon for 2-cm and 4-cm collimation respectively). Considering the most irradiated organs (organ deposited energy > 0.1% total deposited energy), the bone is the tissue most sensitive to the variation of the voxel down sampling. Isotropic voxels better model the bone tissue, with differences of 0.2% to 0.7% for voxels of 5x5x5 and 4x5x4 mm^3. The difference increases up to 2.5% for 10x8x1 mm^3 voxels. The improvement in bone accuracy with smaller voxels is likely due to the better phantom resolution along the axial plane (5x5 or 4x5 instead of 10x8 mm^2), given that the estimation of the dose at the interface between soft tissue and bone (high density and high Z_{eff}) can be critical. For the subsequent studies we chose isotropic voxels of 5x5x5 mm^3 to improve accuracy for the most exposed organs and to reduce runtime (simulation time = 22 s for 5x5x5 vs. 35 s for 4x5x4 mm^3). The spatial distribution of deposited energy is shown in Figure 1 for a window of [0 .0015] eV/photon/mm^3.

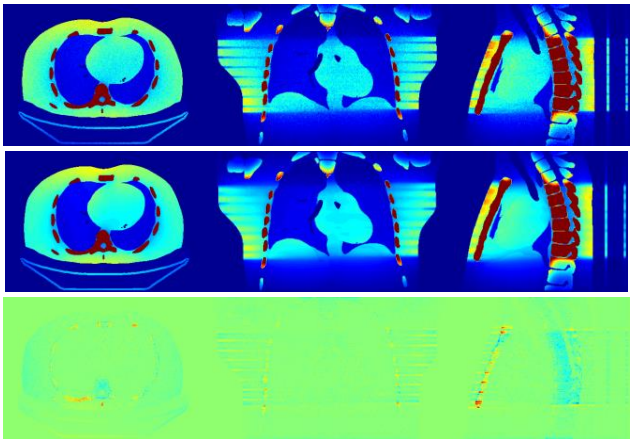


Fig. 1. Energy deposition maps in Geant4 (top) and Acuros (middle) for the case of 2-cm collimation and 120 kV, shown with window of [0 .0015] eV/photon/mm^3. Difference image Acuros-Geant4 (bottom), shown with window of [-0.0015 0.0015], for the axial, coronal and sagittal views, displayed from left to right.

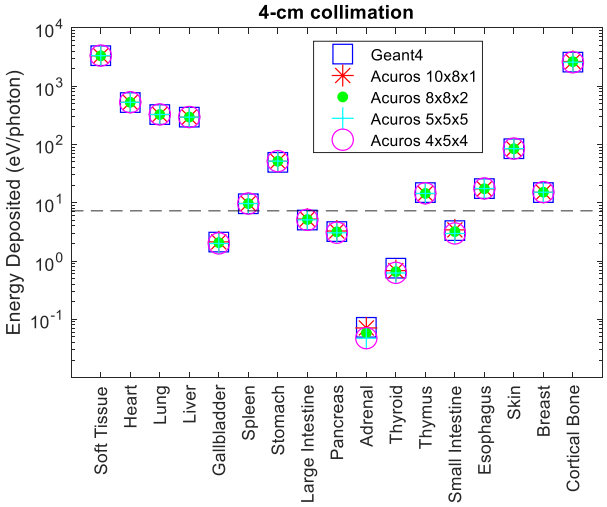


Fig. 2. Energy deposited in the 17 organs defined in the phantom for case modeling 4-cm collimation. The organs that absorb more than 0.1% of the total absorbed energy are above the dashed line. The ground-truth Geant4 organ doses are plotted along with Acuros organ doses resulting from different voxel down sampling schemes in Acuros.

B. *Optimal number of views per rotation*

Figure 3 plots the runtime and organ dose accuracy of Acuros for 4, 5, 8, 12, 18, 24, 30, 40 views per rotation. It shows that a good and conservative compromise between simulation time and accuracy, when accounting for scanner complexities, is reached when 18 views are modeled per rotation. Again, the breast is the organ with the highest error between Geant4 and Acuros. To investigate why the breast had the highest discrepancy among the most exposed organs, simulations were repeated in Acuros for 200 views per rotation, as for Geant4. The improvement of the overall agreement was modest: RMSE improved from 1.8% to 1.4% for 18 and 200 views per rotation respectively. More significant improvements were observed for

specific organs, with breast error decreasing from 4.1% to 2.9% and stomach error decreasing from 3.6 to 2.7%, for 18 and 200 views respectively.

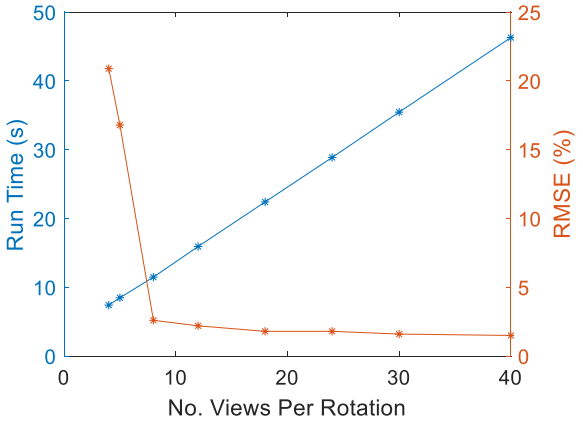


Fig. 3. Run time and RMSE as a function of the number of modeled views per rotation for the 4-cm collimation configuration.

C. *Different source spectra*

RMSE for organs that absorb more than 0.1% of the total deposited energy was approximately 2.5% for the studied spectra of 80 kV and 100 kV, while it increased to 4.2% for the 140-kV beam. Of organs receiving more than 0.1% of the total deposited energy, the breast demonstrated the highest deposited energy discrepancy between Acuros and Geant4. The maximum error was 4.7% for the breast for the 80-kV beam, 5.8% for the 100-kV beam, and 10.9% for the 140-kV beam. The error increase might be due to the upper limit of 125 kV for the highest energy group currently modeled in Acuros. The energy deposition maps for 80 kV are shown in Figure 4.

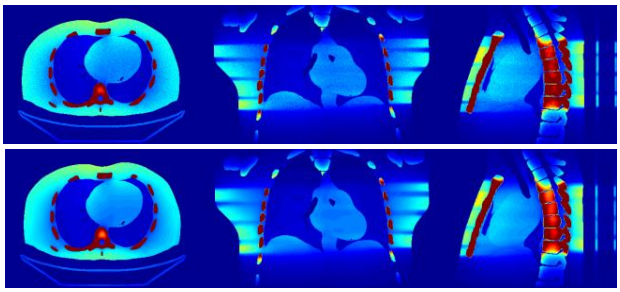


Fig. 4. Energy deposition maps in Geant4 (top) and Acuros (bottom) for the case of 4-cm collimation and 80 kV, shown with window of [0 .0015] eV/photon/mm^3.

D. *Different anthropomorphic phantoms*

Good agreement was achieved for the different patient models. The RMSE was below 1.7% for all three phantoms. The skin entails the highest errors among the most exposed organs, with errors of 4.1%, 2.9%, and 1.5% for the pediatric female, adult female, and adult male, respectively. The energy deposition maps for the male adult phantom are shown in Figure 5.

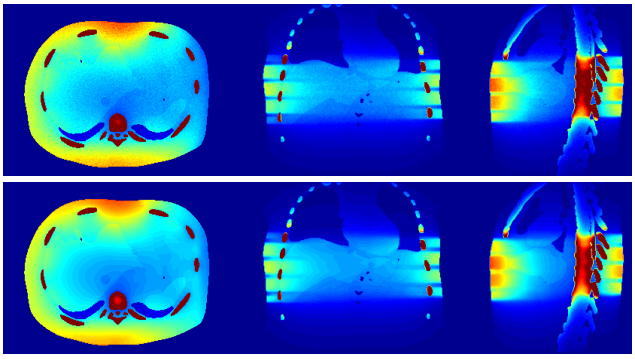


Fig. 5. Energy deposition maps in Geant4 (top) and Acuros (bottom) for the male adult phantom and 120 kV, shown with window of [0 .0015] eV/photon/mm^3.

IV. DISCUSSION AND CONCLUSIONS

The purpose of this work was to implement realistic scanner complexities in the Acuros framework, to obtain accurate organ doses for routine monitoring, and to validate Acuros against Monte Carlo for a range of configurations and patient models. In Acuros the source for each projection view can be specified individually. This aspect was exploited to model spatial and angular variation of the beam to implement bowtie filter, over range collimation and TCM. Acuros and Monte Carlo methods were compared for varying scanner parameters such as collimation, TCM, four x-ray spectra and four patient models of different size. Results show generally good agreement between Acuros and Geant4, with organ dose RMSE < 3% for all cases, except 140 kV (RMSE=4.2%). The results demonstrate that accurate organ doses can be obtained despite the discretization in space, angle, and energy performed in Acuros. The energy deposition maps demonstrate the effect of the bowtie filter that reduces the dose at the edges of the phantom, and of the over range collimation that limits the exposure at the beginning and end of the scan. Acuros organ dose accuracy depends on voxel size, especially for small organs that are not in the direct field of radiation but are only exposed to scatter. Isotropic voxels of 5x5x5 mm^3 appear to be a reasonable option for the TG195 phantom, whereas a down sampling of 4.5x4.5x4.5 mm^3 and 6.0x4.5x4.5 mm^3 was used for the patient models generated with XCAT2. In general, the selection of the voxel size parameter will depend on the desired run time and accuracy, subject to memory constraints of the GPU. The results demonstrate the intrinsic speed of the LBTE approach, with runtimes less than one minute for all cases running on a GeForce GTX 1080 GPU with 8 GB of memory, whereas Geant4 was run on a CPU.

REFERENCES

[1] A. Wang, A. Maslowski, T. Wareing, J. Star-Lack, T. Gilat-Schmidt, “A fast, linear Boltzmann transport equation solver for computed tomography dose calculation (Acuros CTD),” *Med Phys.*, vol. 46, no. 2, pp. 925-933, Dec. 2018.

[2] I. Sechopoulos, *et al*., “Monte Carlo reference data sets for imaging research:executive summary of the report of AAPM Research Committee Task Group 195,” *Med. Phys*, vol. 42, pp. 5679–5691, 2015.

Get two steps ahead with Dual Source CT

SOMATOM Force

Get two steps ahead in clinical excellence
At the top of our Dual Source CT portfolio, SOMATOM Force enables new levels of image quality, clinical outcomes, and ultimately precision medicine. Examine patients without beta-blockers, with no need for them to hold their breath, and with the lowest possible amount of contrast media. Make clearly quantified therapy evaluations with dose-neutral Dual Energy.

Get two steps ahead in workflow performance
Automated technologies support safe, standardized, and highly performant workflows – allowing for appropriate dose and reproducible precision from the smallest to the tallest patients.

Get two steps ahead in expert leadership
Thinking beyond today, you're connected to the future with an ever-growing expert community and exclusive access to our advanced research environment.

siemens-healthineers.com/somatom-force

SIEMENS Healthineers

The scientific overlay is not that of the individual pictured and is not from a device of Siemens Healthineers. It was modified for better visualization.

International version. Not for distribution or use in the U.S. · 7515 0519

Wednesday, August 5

Deep Learning

Motion Estimation and Compensation

Poster Session 2

Artifact and Noise Reduction

Generalizability test of a deep learning-based CT image denoising method

Rongping Zeng, Claire Yilin Lin, Qin Li, Jiang Lu, Jeffrey A Fessler and Kyle J Myers

Abstract— **Deep learning (DL) has been increasingly explored in low-dose CT image denoising. DL products have also been submitted to the FDA for premarket clearance. While having the potential to improve image quality over the filtered back projection method (FBP) and produce images quickly, generalizability of DL approaches is a major concern because the performance of a DL network can depend highly on the training data. In this work we take a residual encoder-decoder convolutional neural network (REDCNN)-based CT denoising method as an example. We investigate the effect of the scan parameters associated with the training data on the performance of this DL-based CT denoising method and identifies the scan parameters that may significantly impact its performance generalizability. This abstract particularly examines these three parameters: reconstruction kernel, dose level and slice thickness. Our preliminary results indicate that the DL network may not generalize well between FBP reconstruction kernels, but is insensitive to slice thickness for slice-wise denoising. The results also suggest that training with mixed dose levels improves denoising performance.**

Index Terms—**Deep learning, Generalizability test, Low-dose CT denoising**

I. Introduction

CT imaging is widely used in the clinic to assist the diagnosis of abnormalities and to monitor treatment response. It is essential to reduce the x-ray dose to a reasonable level for patient safety while maintaining the CT image quality for accurate decision making. Various approaches have been developed toward this goal through smarter hardware design, such as automatic exposure control, kV optimization and dynamic bowtie filter design, as well as through advanced image reconstruction/ denoising methods, such as statistical and model-based iterative reconstruction (IR) algorithms. Recently, deep learning (DL) methods are attracting high attention, thanks to the growth of big data and increased computation power. DL methods have the potential to improve image quality over FBP and produce image quality comparable to some IR methods (1).

Unlike traditional FBP and (IR) algorithms that were derived based on the imaging physics, DL methods rely on training data that usually contain noisy images and their corresponding high-quality (low noise or better resolution) target to optimize the DL network coefficients to estimate and remove the noise from a noisy input. This data-driven mechanism makes the DL performance less predictable when applying the DL network to process data that has properties that differ from the training data. In CT imaging, these properties include different anatomical regions and different acquisition settings. Considering the varieties of the CT data, generalizability is hence an important aspect in the performance evaluation of DL based denoising methods.

This work aims to investigate the generalizability of DL based CT denoising methods. We implemented an example DL denoising network(2). We categorized the training images in terms of the scan parameters and trained the network separately with each data category. We then evaluated the performances of the DL networks and compared how the performances may differ when tested on the same and different categories of testing data. This abstract reports our initial findings regarding the performance generalizability of DL-based CT denoising methods related to the CT acquisition parameters.

II. Methods

A. Low-dose CT denoising network

Let $\boldsymbol{x} \in R^{m\times n}$ denote a high-noise CT image; the DL-based denoising problem is to optimize the network $C(\boldsymbol{x})$: $R^{m\times n} \rightarrow R^{m\times n}$ that maps $\boldsymbol{x}$ to its corresponding low-noise image $\boldsymbol{y} \in R^{m\times n}$ by minimizing a loss function between $\boldsymbol{x}$ and $\boldsymbol{y}$ based on a giving set of training data. After the network is optimized, a noisy CT image can be passed through the network to produce an image that may have reduced noise.

Various network structures have been explored in the literature for low-dose CT denoising, including fully connected neural network (FCN), residual network (ResNet), UNet and others. In this work, we selected the residual encoder-decoder convolutional neural network (REDCNN) developed by Chen et al.(2) for performing a generalizability test. The loss function for training the denoising network was the mean squared error (MSE) between the DL output and the corresponding low-noise training target.

As illustrated in Fig. 1, REDCNN contains 10 layers of FCNs (5 convolutional and 5 deconvolutional layers) with a rectified linear units (ReLU) activation function following the FCN in each layer. Three shortcuts are added to connect the convolution layers and deconvolution layers to construct a

Corresponding author: Rongping Zeng, rongping.zeng@fda.hhs.gov

Affiliations: Zeng R, Li Q, Lu J and Myers KJ are from the Office of Science and Engineering Laboratories, within the Center for Devices and Radiological Health, US Food and Drug Administration at Silver Spring, Maryland. Lin CY is from the Department of Mathematics and Fessler JA is from the Department of Electrical Engineering and Computer Science in the University of Michigan at Ann Arbor, Michigan. Lin CY was supported in part by the FDA critical path funding and Fessler JA was supported in part by NSF grant IIS 1838179.

residual learning mechanism. More details about the network structure design can be found in the paper by Chen et al.(2).

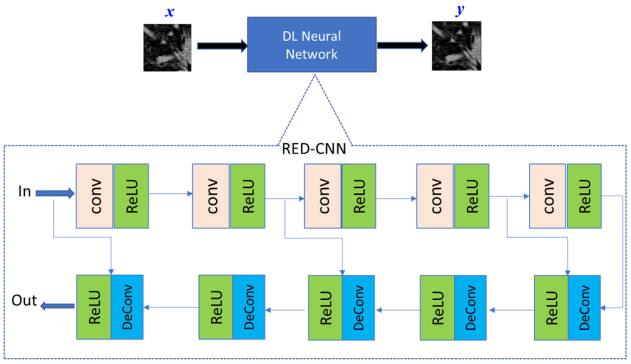


Fig. 1. Illustration of the REDCNN denoising network.

B. *Training data categorization*

We used the Low-dose Grand Challenge (LDGC) dataset(3) shared by the Mayo Clinic to train the DL denoising network. The LDGC dataset contained ten patient datasets of quarter and full dose scans acquired on a Siemens CT scanner reconstructed with two slice thicknesses (1 mm and 3 mm) and two reconstruction kernels (a sharp kernel D45 and a smooth kernel B30). The corresponding quarter- and full-dose image pairs were treated as training input and training target in the DL training process. Among the ten patient datasets, seven were used for training. A total of 350 slices of size 512x512 were randomly selected from the seven patients and each slice was divided into 55x55 patches excluding the outside of the body. This resulted in about 70,000 training patches.

Note that in the LDGC dataset there was only one reduced dose level (25%) available. To synthesize other dose levels, we extracted the noise by subtracting the quarter dose image from the full dose image and then blended a portion of the noise back into the full dose image. By varying the blending factor from 0.5 to 1.2, we created images corresponding to dose levels ranging about 17% to 80% of the full-dose scan.

The variety of reconstruction thickness, reconstruction kernel and augmented dose levels allowed us to examine the performance generalizability of the DL denoising network for the three parameters. Particularly we categorized the training images into the following six groups based on the combined acquisition parameter values:

- Kernel effect: **Sharp kernel** / 3 mm thickness / 25% dose level; **Smooth kernel** / 3 mm thickness / 25% dose level
- Thickness effect: Smooth kernel / **1 mm thickness** / 25% dose level; Smooth kernel / **3 mm thickness** / 25% dose level
- Dose level effect: Smooth kernel / 3 mm thickness / **25% dose level**; Smooth kernel / 3 mm thickness / **Mixed dose levels of 17%-80%**

With this categorization, we were able to obtain six trained DL networks. For convenience, we name networks according to the parameter setting of the training data as follows: *DLkernel-thickness-dose%.* For example, "DLsharp-3mm-mix%" represents the REDCNN trained with images of sharp kernel, 3mm thickness and mixed dose levels; "DLsmooth-1mm-25%" represents the REDCNN trained with images of smooth kernel, 1mm thickness and a single 25% dose level. The six networks will be evaluated to check how they may preserve their performance when denoising a different category of testing data relative to their performance on denoising the same category of data as used in training.

C. *Generalizability performance evaluation*

To evaluate the performance, we considered the following metrics: MSE, modulation transfer function (MTF), noise power spectrum (NPS), and low-contrast detectability (LCD). MSE reflects how well the network performs in terms of minimizing the loss function that the network is designed to do; MTF and NPS are standard performance metrics that characterize the image quality of linear imaging systems; LCD represents a task-based performance metric that augments MTF and NPS to challenge nonlinear smoothing algorithms such as IR and DL methods.

For the MSE measure, the slices from one of the three patients in the LDGC dataset that were not included in the training were used as the test set. The corresponding full-dose images were treated as the references for calculating the MSE. For the MTF, NPS and LCD measures, we simulated 2D phantom images of the CATPHAN600 contrast module (Fig. 2), a uniform water phantom and the CCT189 phantom (Fig. 2), respectively. The simulated fan-beam CT scanner will be described in the next paragraph. The contrast module in CATPHAN600 contains a few disks with varying HU values that allow the measurement of contrast-dependent MTF for non-linear image reconstruction(4). The NPS was estimated by taking the average of the modulus square of the Fourier transform of multiple repeats of the uniform phantom scans. The CCT189 phantom contains four low contrast disks with varying size/HU combinations (3mm/14HU, 5mm/7HU, 7mm/5HU, 10mm/3HU). The task of detecting the low-contrast disks challenges the image reconstruction/denoising methods that usually involve nonlinear smoothing. In this study, the detectability was evaluated using a Laguerre-Gauss channelized Hoteling model observer(5).

The simulated fan-beam CT scanner was set to have distances of 595 mm from the x-ray tube to the isocenter and 1085.6 mm to the detector, same as the Siemens CT scanner used to collect the LDGC dataset. Poisson noise was modeled at the detector and the air photon flux controlled the noise level in the reconstructed images. For FBP reconstruction, we applied two Hann filters of different cutoff frequencies (named Hann1 & Hann2) to be convolved with the sinogram to yield similar MTF50% and MTF10% as the D45 and B30 filters in LDGC (see Table 1). For convenience, we refer to Hann1 and D45 as sharp kernels, Hann2 and B30 as smooth kernels in this abstract. Since the simulated CT scanner was a 2D fan-beam scanner, slice thickness was not a modeled parameter; all the simulated data effectively corresponded to a very thin slice thickness.

Resolution (lp/cm)	D45 (Sharp)	Hann1 (sharp)	B30 (smooth)	Hann2 (smooth)
MTF50%	5.6	5.6	3.5	3.5
MTF10%	9.4	10.4	5.9	6.2

Table 1. The MTF50% and MTF10% values of the commercial (D45, B30) and simulated reconstruction kernels (Hann1 and Hann2).

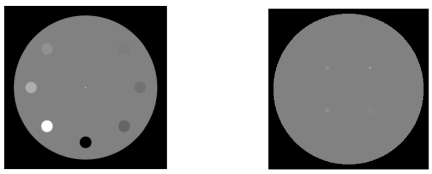


Fig. 2: Left: simulated contrast phantom that mimics the contrast layer in the CATPHAN600 for measuring the MTF. Right: simulated LCD phantom that mimics the CCT189 phantom containing four low-contrast disks for measuring the LCD performance.

III. Results and Discussions

A. MSE

Fig. 3 shows the box plots of the MSE reduction rate of the DL processed image relative to the original noisy image input. Comparison of the kernel effect in the plots clearly shows a significantly larger MSE reduction rate when the test data has the same reconstruction kernel as the training data, indicating that the kernel may be an important factor of the DL performance generalizability. Comparison of the thickness effect shows that the DL network trained with 3mm thickness has slightly better than or similar MSE reduction to the DL network trained with 1mm thickness data, indicating the slice thickness may not be critical. Comparison of the dose effect shows that the DL network trained with a single dose level does not preserve the MSE performance on denoising images at a different dose level but the DL network trained with mixed dose levels can, indicating that training with mixed dose levels is more robust.

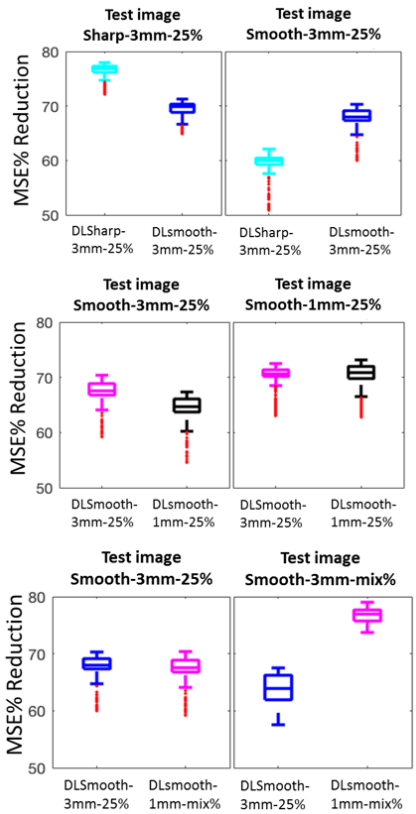


Fig. 3 Comparison of the MSE reduction of the DL networks.

B. MTF

We generated noiseless sinograms of the contrast phantom and reconstructed two images, by FBP of sharp and smooth kernels. We then applied the DL networks to process the two images and measured the MTF at the five contrasts: 990, 340, 200, 120 and 35 HU. The MTF50% value was estimated to represent the image resolution. Fig. 4 plots the MTF50% values vs the HU contrast for the trained DL networks for processing FBP sharp (Fig. 4a) and FBP smooth image (Fig. 4b).

In general, the MTF50% value decreases with the disk HU value, as shown in both plots in Fig. 4. The decreased image resolution with decreased HU is a common characteristic of non-linear image reconstruction or denoising methods. In Fig. 4a, it can be observed that the DL network trained with the sharp kernel data has better image resolution (higher MTF50% value) than the DL network trained smooth kernel, and the DL network trained with mixed dose data has better image resolution than the DL network trained with a single dose data for processing the FBP sharp image. It can also be seen that the DL network trained with 3mm thickness has better image resolution. Similar trends present in Fig 4b for processing the FBP smooth image, although with smaller magnitude of differences among the DL networks.

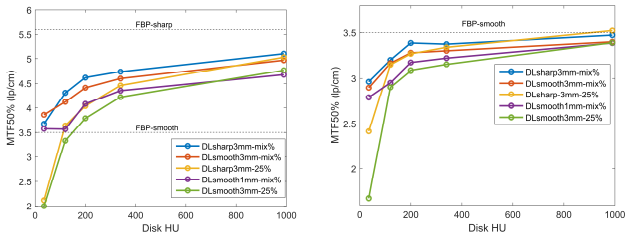


Fig. 4: Contrast-dependent MTF50% curves of the DL networks for processing the FBP sharp (a) and FBP smooth (b) images.

C. NPS

We generated 30 noisy scans of the cylindrical water phantom for the NPS estimation. Each noisy scan was reconstructed by FBP of both sharp and smooth kernels. Based on the MSE and MTF tests, we found that DL network trained with the thicker slice thickness and mixed dose had better performance than those trained with the thin thickness and single dose. Therefore, we decided to only compare the DLsharp-3mm-mixDose and DLsmooth-3mm-mixDose, referred as *DLsharp* and *DLsmooth* thereafter, in the NPS test, as well as in the LCD test later.

Fig.5 shows the 2D NPS images evaluated using the central ROIs of size 64x64. The NPS of the DLsharp appears to have stronger noise magnitude along the northwest diagonal direction, no matter for processing the FBP sharp or smooth images. The NPS of the DLsmooth has a normal round appearance when processing the FBP smooth images but contains much of high-frequency noise when processing the FBP sharp images. By examining the 1D NPS curves (radial binning of the 2D NPS image) shown in Fig. 6, it can be seen that the NPS curve of the DLsmooth has a bump (marked by the blue arrow in the Fig. 6) in the tail that almost matches the magnitude of the NPS tail of the FBP sharp images, indicating that the DLsmooth only suppressed the noise in the low and medium frequency bands but left the high-frequency noise largely untouched. We hypothesize that the smooth kernel data

did not contain high-frequency noise so the DLsmooth did not learn to suppress the high-frequency noise.

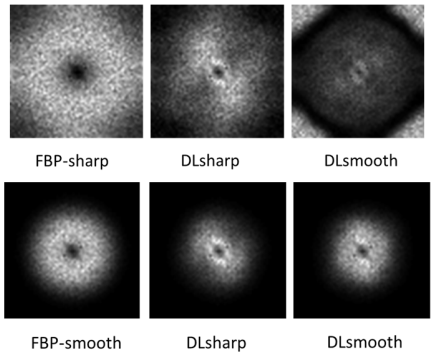


Fig. 5 Top: the NPS images of FBP sharp images and the corresponding DL processed images. Bottom: The NPS images of FBP smooth images and the corresponding DL processed images.

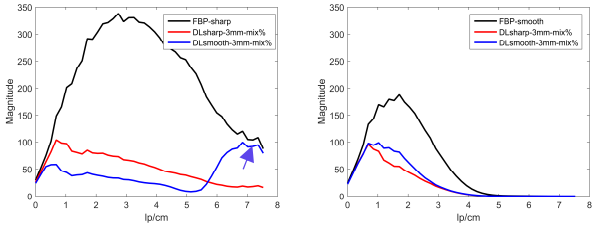


Fig. 6 The 1D NPS curve of the images in Fig. 5. Left: the FBP sharp and the corresponding DL processed images. Right: the FBP smooth and the corresponding DL processed images.

D. LCD

In the LCD test, we simulated 200 noisy scans of the LCD phantom and 100 noisy scans of a uniform phantom at each of the five dose levels (100% to 30%) we selected. The 100% dose level corresponded to an air photon count of 3×10^5. For each low-contrast insert, one signal-present (SP) ROI was extracted from each LCD phantom image and five signal-absent (SA) ROIs were extracted at the vicinity of the insert location from each of the uniform phantom image. As a result, a total of 200 SP and 500 SA ROIs for each insert were created for the evaluation of detectability.

Fig. 7 contains the detectability curves as a function of dose for two of the four inserts. For the two inserts of relatively small sizes (3mm/14HU, 5mm/7HU) the detectability values were very close among the FBP reconstructions and the DL denoising networks, with the AUC ranging from 0.8 to 0.95 in the simulated dose levels, so their detectability curves were not shown in this abstract. The detectability curves are more separated for the two inserts of slightly larger sizes. As clearly shown in the 10mm/3HU insert plot, the smooth kernel boosts the LCD performance in the original FBP reconstruction. The DLsmooth has better LCD performance than the DLsharp on processing both the FBP sharp or smooth images. However, the best LCD performances are achieved by the DLsmooth denoising the FBP sharp image, despite the high-frequency noise shown in the NPS test. We hypothesize that the LCD tasks focused more on the low and medium frequency band and the DLsmooth operated on FBP sharp images achieved an improved balance between image resolution and noise suppression within the signal's frequency band.

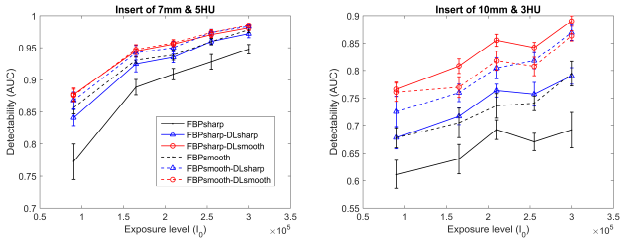


Fig. 7 Detectability curves for the two inserts (7mm&5HU, 10mm&3HU) in the original FBP sharp and FBP smooth images, and denoised FBP images with the DLsharp and DLsmooth.

IV. Conclusions and Future Work

This abstract presented our preliminary work in testing the performance (MSE, MTF, NPS and LCD) generalizability of a DL-based CT denoising method related to three CT acquisition parameters. Our results showed that the DL network was most sensitive to the reconstruction kernel. The DL network trained with thicker slice thickness data appeared to be slightly better than that trained with thinner slice thickness. The DL network trained with mixed dose levels was more robust than that trained with a single dose level. Future work is needed to explore the impact of the other acquisition parameters. Other tasks that can test the preservation of higher frequency information than the LCD task may be necessary for a more complete performance evaluation of the DL denoising networks.

References

1. MacDougall RD, Zhang Y, Callahan MJ, Perez-Rossello J, Breen MA, Johnston PR, Yu H. Improving Low-Dose Pediatric Abdominal CT by Using Convolutional Neural Networks. Radiology: Artificial Intelligence. 2019;1(6):e180087. doi: 10.1148/ryai.2019180087.
2. Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, Zhou J, Wang G. Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network. IEEE Transactions on Medical Imaging. 2017;36(12):2524-35. doi: 10.1109/TMI.2017.2715284.
3. McCollough CH, Bartley AC, Carter RE, Chen B, Drees TA, Edwards P, Holmes III DR, Huang AE, Khan F, Leng S, McMillan KL, Michalak GJ, Nunez KM, Yu L, Fletcher JG. Low-dose CT for the detection and classification of metastatic liver lesions: Results of the 2016 Low Dose CT Grand Challenge. Medical Physics. 2017;44(10):e339-e52. doi: 10.1002/mp.12345.
4. Richard S, Husarik DB, Yadava G, Murphy SN, Samei E. Towards task-based assessment of CT performance: System and object MTF across different reconstruction algorithms. Medical Physics. 2012;39(7Part1):4115-22. doi: 10.1118/1.4725171.
5. Vaishnav JY, Jung WC, Popescu LM, Zeng R, Myers KJ. Objective assessment of image quality and dose reduction in CT iterative reconstruction. Medical Physics. 2014;41(7):071904. doi: 10.1118/1.4881148.

Learned Multi-layer Residual Sparsifying Transform Model for Low-dose CT Reconstruction

Xikai Yang, Xuehang Zheng, Yong Long*, *Member, IEEE,* Saiprasad Ravishankar, *Member, IEEE*

***Abstract*—Signal models based on sparse representation have received considerable attention in recent years. Compared to synthesis dictionary learning, sparsifying transform learning involves highly efficient sparse coding and operator update steps. In this work, we propose a Multi-layer Residual Sparsifying Transform (MRST) learning model wherein the transform domain residuals are jointly sparsified over layers. In particular, the transforms for the deeper layers exploit the more intricate properties of the residual maps. We investigate the application of the learned MRST model for low-dose CT reconstruction using Penalized Weighted Least Squares (PWLS) optimization. Experimental results on Mayo Clinic data show that the MRST model outperforms conventional methods such as FBP and PWLS methods based on edge-preserving (EP) regularizer and single-layer transform (ST) model, especially for maintaining some subtle details.**

***Index Terms*—Low-dose CT, Statistical image reconstruction, Sparse representation, Transform learning, Unsupervised learning.**

I. Introduction

Signal models exploiting sparsity have been shown to be useful in a variety of applications such as compression, restoration, denoising, reconstruction, etc. Natural signals can be modeled as sparse in a synthesis dictionary (i.e., represented as linear combinations of few dictionary atoms or columns) or in a sparsifying transform domain.

Transforms such as wavelets [1] and the discrete cosine transform (DCT) are well-known to sparsify images. Methods such as synthesis dictionary learning [2] and analysis dictionary learning [3] allow adapting the model to the data. These dictionary learning problems are NP-hard in general and involve computationally expensive alternating-type algorithms that limit applicability to large-scale data. In contrast, the recently proposed sparsifying transform learning approaches [4] involve exact and highly efficient updates. In particular, the transform model approximates the signal as sparse in a transformed domain. Adopting a multi-layer sparsifying transform model enables sparsifying an input image successively over layers [5], creating a rich and more complete sparsity model, which forms the core of this work.

One of the most important application of such image models is for medical image reconstruction. In particular, an important problem in computed tomography (CT) is reducing the X-ray exposure to patients while maintaining good image reconstruction quality. A conventional method for CT reconstruction is the analytical filtered back-projection (FBP) [6]. However, image quality degrades severely for FBP when the radiation dose is reduced. In contrast, model-based image reconstruction (MBIR) achieves comparable or better image quality [7].

A typical MBIR method for low-dose CT (LDCT) is the penalized weighted least squares (PWLS) approach. The cost function for PWLS includes a weighted quadratic data-fidelity term and a penalty term or regularizer capturing prior information or model of the object [8, 9]. Recent works have shown promising LDCT reconstruction quality by incorporating data-driven models into the regularizer, where the models are learned from datasets of images or image patches. In particular, PWLS reconstruction with adaptive sparsifying transform-based regularization has shown promise for tomographic reconstruction [10–12]. Recent work has also shown that they may generalize better to new data than supervised deep learning schemes [13]. The adaptive transform-based image reconstruction algorithms can exploit a variety of image models [10, 11, 14] learned in an unsupervised manner, and involve efficient closed-form solutions for sparse coding.

In this work, we propose a new formulation and algorithm for learning a multi-layer transform model [5], where the transform domain residuals (the difference between transformed data and their sparse approximations) are successively sparsified over several layers. We refer to the model as a multi-layer residual sparsifying transform (MRST) model. The transforms are learned over several layers from images to jointly minimize the transform domain residuals across layers, while enforcing sparsity conditions in each layer. Importantly, the filters in deeper layers can help better exploit finer features (e.g., edges and correlations) in the residual maps. We investigate the performance of the (unsupervised) learned MRST model for LDCT reconstruction using PWLS. We propose efficient alternating minimization algorithms for both learning and reconstruction. Experimental results on Mayo Clinic data illustrate that the learned MRST model outperforms the conventional FBP as well as PWLS methods based on the non-adaptive edge-preserving (EP) regularizer and learned single-layer transform model, especially for maintaining some subtle details. In the following sections, we will first discuss the proposed model and formulation, followed by our algorithms and experimental results on Mayo Clinic data.

II. Problem Formulation

Here, we will introduce the proposed general multi-layer learning framework and the formulation for LDCT image reconstruction. The MRST learning cost and constraints are shown in Problem (P0), which is an extension of simple single-layer transform learning.

$$\min_{\{\mathbf{\Omega}_l, \mathbf{Z}_l\}} \sum_{l=1}^{L} \left\{ \|\mathbf{\Omega}_l \mathbf{R}_l - \mathbf{Z}_l\|_F^2 + \eta_l^2 \|\mathbf{Z}_l\|_0 \right\}$$
$$\text{s.t.} \quad \mathbf{R}_l = \mathbf{\Omega}_{l-1}\mathbf{R}_{l-1} - \mathbf{Z}_{l-1}, 2 \leq l \leq L, \mathbf{\Omega}_l^T \mathbf{\Omega}_l = \mathbf{I}, \forall l. \tag{P0}$$

Here, $\{\mathbf{\Omega}_l \in \mathbb{R}^{p \times p}\}$ and $\{\mathbf{Z}_l \in \mathbb{R}^{p \times N}\}$ denote the learned transforms and sparse coefficient maps for the $1 \leq l \leq L$ layers. Parameter η_l controls the maximum allowed sparsity level (computed using the ℓ_0 "norm" penalty) of $\mathbf{Z}_l$. The residual maps $\{\mathbf{R}_l \in \mathbb{R}^{p \times N}\}$ are defined in recursive form over layers, with $\mathbf{R}_1$ denoting the input training data. Here, we assume $\mathbf{R}_1$ to be a matrix, whose columns are (vectorized) patches drawn from

X. Yang, X. Zheng and Y. Long are with the University of Michigan–Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: yangxk@sjtu.edu.cn; zhxhang@sjtu.edu.cn; yong.long@sjtu.edu.cn)

S. Ravishankar is with the Department of Computational Mathematics, Science and Engineering and Department of Biomedical Engineering, Michigan State University, East Lansing, MI 48824, USA (e-mail: ravisha3@msu.edu)

*Yong Long is the corresponding author.

image data sets. The unitary constraints for $\{\mathbf{\Omega}_l\}$ enable closed-form solutions for the sparse coefficient and transform update steps in our algorithms. The learned MRST model can then be used to construct a data-driven regularizer in PWLS as shown in Problem (P1).

$$\min_{\mathbf{x}\geq\mathbf{0}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2+\beta \mathsf{S}(\mathbf{x}), \tag{P1}$$

$$\mathsf{S}(\mathbf{x}) \triangleq \min_{\{\mathbf{Z}_l\}} \sum_{l=1}^{L} \left\{ \|\mathbf{\Omega}_l\mathbf{R}_l-\mathbf{Z}_l\|_F^2+\gamma_l^2\|\mathbf{Z}_l\|_0 \right\}$$

$$\text{s.t. } \mathbf{R}_l=\mathbf{\Omega}_{l-1}\mathbf{R}_{l-1}-\mathbf{Z}_{l-1}, 2\leq l\leq L, \mathbf{R}_1^j=\mathbf{P}^j\mathbf{x}, \forall\, j.$$

In particular, we reconstruct the image $\mathbf{x}\in\mathbb{R}^{N_p}$ from noisy sinogram data $\mathbf{y}\in\mathbb{R}^{N_d}$ by solving (P1), where $\mathbf{A}\in\mathbb{R}^{N_d\times N_p}$ is the system matrix of the CT scan and $\mathbf{W}=\mathsf{diag}\{w_i\}\in\mathbb{R}^{N_d\times N_d}$ is the diagonal weighting matrix with elements being the estimated inverse variance of y_i. Operator $\mathbf{P}^j\in\mathbb{R}^{p\times N_p}$ extracts the jth patch of $\mathbf{x}$ as $\mathbf{P}^j\mathbf{x}$. The jth colums of $\mathbf{R}_l$ and $\mathbf{Z}_l$ are denoted $\mathbf{Z}_l^j$ and $\mathbf{R}_l^j$. The non-negative parameters $\{\gamma_l\}$ control the sparsity of the coefficient maps in different layers, and $\beta>0$ captures the relative trade-off between the data-fidelity term and regularizer.

III. Algorithms for Learning and Reconstruction

A. Transform Learning Algorithm

We propose an exact block coordinate descent (BCD) algorithm for the nonconvex Problem (P0) that cycles over updating each $\mathbf{Z}_l$ for $1\leq l\leq L$ (sparse coding step) and each $\mathbf{\Omega}_l$ for $1\leq l\leq L$ (transform update step). In each step, the remainder of the variables (that are not optimized) are kept fixed. The exact BCD algorithm ensures that the objective in (P0) is monotone decreasing over iterations and that it converges. In particular, under the unitarity condition on the transforms, every subproblem in the block coordinate descent minimization approach can be solved exactly. We initialize the algorithm with 2D DCT for $\mathbf{\Omega}_1$ and the identity matrix for $\{\mathbf{\Omega}_l\}(l\geq 2)$ respectively. The initial $\{\mathbf{Z}_l\}$ are all-zero matrices. Since the residuals are defined recursively in (P0), to simplify the algorithmic description, we first define matrices $\mathbf{B}_p^q(p<q)$, which can be regarded as backpropagation matrices from the qth to pth layers.

$$\mathbf{B}_p^q=\mathbf{\Omega}_{p+1}^T\mathbf{Z}_{p+1}+\mathbf{\Omega}_{p+1}^T\mathbf{\Omega}_{p+2}^T\mathbf{Z}_{p+2}+\ldots+\mathbf{\Omega}_{p+1}^T\mathbf{\Omega}_{p+2}^T\ldots\mathbf{\Omega}_q^T\mathbf{Z}_q$$
$$=\sum_{k=p+1}^{q}\left(\prod_{s=p+1}^{k}\mathbf{\Omega}_s^T\right)\mathbf{Z}_k \tag{1}$$

1) Sparse Coding Step for $\mathbf{Z}_l$: Here, we solve (P0) for $\mathbf{Z}_l$ with all other variables fixed. The corresponding subproblem is as in (2).

$$\min_{\mathbf{Z}_l}\sum_{i=l}^{L}\left\{\|\mathbf{\Omega}_i\mathbf{R}_i-\mathbf{Z}_i\|_F^2\right\}+\eta_l^2\|\mathbf{Z}_l\|_0 \tag{2}$$

Although the cost function is nonconvex and the residual maps also depend on $\mathbf{Z}_l$ (as in (P0)), we can exploit the unitarity of the transforms to simplify and rewrite Problem (2) as $\min_{\mathbf{Z}_l}(L-l+1)\times\|\mathbf{Z}_l-(\mathbf{\Omega}_l\mathbf{R}_l-\frac{1}{L-l+1}\sum_{i=l+1}^{L}\mathbf{B}_l^i)\|_F^2+\eta_l^2\|\mathbf{Z}_l\|_0$. This problem has a similar form as the single-transform sparse coding problem [4], and the optimal solution $\hat{\mathbf{Z}}_l$ is obtained as in (3), where $H_\eta(\cdot)$ denotes the *hard-thresholding* operator that sets elements with magnitude less than the threshold η to zero.

$$\hat{\mathbf{Z}}_l=\begin{cases}H_{\eta_l/\sqrt{L-l+1}}\left(\mathbf{\Omega}_l\mathbf{R}_l-\frac{1}{L-l+1}\sum_{i=l+1}^{L}\mathbf{B}_l^i\right), & 1\leq l\leq L-1,\\ H_{\eta_L}(\mathbf{\Omega}_L\mathbf{R}_L), & l=L.\end{cases} \tag{3}$$

2) Transform Update Step for $\mathbf{\Omega}_l$: Here, we fix $\{\mathbf{Z}_l\}$ and all $\mathbf{\Omega}_l$ (except the target $\mathbf{\Omega}_l$ in (P0)) and solve the following subproblem:

$$\min_{\mathbf{\Omega}_l}\sum_{i=l}^{L}\left\{\|\mathbf{\Omega}_i\mathbf{R}_i-\mathbf{Z}_i\|_F^2\right\}\quad\text{s.t.}\quad\mathbf{\Omega}_l^T\mathbf{\Omega}_l=\mathbf{I}. \tag{4}$$

Incorporating the recursive dependence of the residual maps on the transforms (as in (P0)) along with the unitarity of the transforms simplifies the solution to (4). First, denoting the full singular value decomposition (SVD) of the matrix $\mathbf{G}_l$ below by $\mathbf{U}_l\mathbf{\Sigma}_l\mathbf{V}_l^T$, the optimal solution to (4) is $\hat{\mathbf{\Omega}}_l=\mathbf{V}_l\mathbf{U}_l^T$.

$$\mathbf{G}_l=\begin{cases}\mathbf{R}_l\left(\mathbf{Z}_l+\frac{1}{L-l+1}\sum_{i=l+1}^{L}\mathbf{B}_l^i\right)^T, & 1\leq l\leq L-1,\\ \mathbf{R}_L\mathbf{Z}_L^T, & l=L.\end{cases} \tag{5}$$

B. Image Reconstruction Algorithm

The proposed PWLS-MRST algorithm for low-dose CT image reconstruction exploits the learned MRST model. We reconstruct the image by solving the PWLS problem (P1). We propose a block coordinate descent (BCD) algorithm for (P1) that cycles over updating the image $\mathbf{x}$ and each of the sparse coefficient maps $\mathbf{Z}_l$ for $1\leq l\leq L$.

1) Image Update Step for $\mathbf{x}$: First, with the sparse coefficient maps $\{\mathbf{Z}_l\}$ fixed, we optimize for $\mathbf{x}$ in (P1) by optimizing the following subproblem:

$$\min_{\mathbf{x}\geq\mathbf{0}}\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2+\beta\mathsf{S}_2(\mathbf{x}), \tag{6}$$

where $\mathsf{S}_2(\mathbf{x})\triangleq\sum_{l=1}^{L}\left\{\|\mathbf{\Omega}_l\mathbf{R}_l-\mathbf{Z}_l\|_F^2\right\}$, with $\mathbf{R}_l=\mathbf{\Omega}_{l-1}\mathbf{R}_{l-1}-\mathbf{Z}_{l-1}$, $2\leq l\leq L$, and $\mathbf{R}_1^j=\mathbf{P}^j\mathbf{x}$. We use the *efficient* relaxed linearized augmented Lagrangian method with ordered-subsets (relaxed OS-LALM) algorithm [15] to obtain the solution to (6). The algorithmic details are shown in **Algorithm 1**. In each iteration of the OS-LALM method, we update the image M times corresponding to M ordered subsets. The matrices $\mathbf{A}_m$ and $\mathbf{W}_m$, and the vector $\mathbf{y}_m$ are sub-matrices of $\mathbf{A}$ and $\mathbf{W}$, and sub-vector of $\mathbf{y}$ for the mth subset respectively. Matrix $\mathbf{D}_\mathbf{A}$ represents a diagonal majorizing matrix of $\mathbf{A}^T\mathbf{W}\mathbf{A}$. We precompute the Hessian matrix of $\mathsf{S}_2(\mathbf{x})$ as $\mathbf{D}_{\mathsf{S}_2}$ in (8) to accelerate the algorithm, and the gradient of $\mathsf{S}_2(\mathbf{x})$ is as in (7).

$$\nabla\mathsf{S}_2(\mathbf{x})=2\beta\sum_{j=1}^{N}(\mathbf{P}^T)^j\left\{L\mathbf{P}^j\mathbf{x}-\sum_{k=1}^{L}(\mathbf{B}_0^k)^j\right\} \tag{7}$$

$$\mathbf{D}_{\mathsf{S}_2}\triangleq\nabla^2\mathsf{S}_2(\mathbf{x})=2L\beta\sum_{j=1}^{N}(\mathbf{P}^T)^j\mathbf{P}^j \tag{8}$$

2) Sparse Coding Step for $\mathbf{Z}_l$: Similar to the sparse coding step during transform learning, the solution of (P1) with respect to each sparse coefficient map $\mathbf{Z}_l$ is shown in (10), and is the solution of (9).

$$\min_{\mathbf{Z}_l}\sum_{i=l}^{L}\left\{\|\mathbf{\Omega}_i\mathbf{R}_i-\mathbf{Z}_i\|_F^2\right\}+\gamma_l^2\|\mathbf{Z}_l\|_0$$
$$\text{s.t.}\quad\mathbf{R}_i=\mathbf{\Omega}_{i-1}\mathbf{R}_{i-1}-\mathbf{Z}_{i-1}, l\leq i\leq L \tag{9}$$

$$\hat{\mathbf{Z}}_l=H_{\eta_l/\sqrt{L-l+1}}\left\{\mathbf{\Omega}_l\mathbf{R}_l-\frac{1}{L-l+1}\sum_{i=l+1}^{L}\mathbf{B}_l^i\right\} \tag{10}$$

IV. Experimental Results

In this section, we evaluate the image reconstruction quality for the proposed PWLS-MRST algorithm and compare it with conventional **FBP** and the **PWLS-EP** (Edge Preserving) method [16].

Algorithm 1 Image Reconstruction Algorithm

Input: initial image $\tilde{\mathbf{x}}^{(0)}$, all-zero initial $\{\tilde{\mathbf{Z}}_l^{(0)}\}$, pre-learned $\{\mathbf{\Omega}_l\}$, thresholds $\{\gamma_l\}$, $\alpha = 1.999$, $\mathbf{D_A}$, $\mathbf{D_{S_2}}$, number of outer iterations T_O, number of inner iterations N, and number of subsets M.

Output: reconstructed image $\tilde{\mathbf{x}}^{(T)}$.

for $t = 0, 1, 2, \cdots, T_O - 1$ **do**

1) Image Update: $\{\tilde{\mathbf{Z}}_l^{(t)}\}$ fixed,

Initialization: $\rho = 1$, $\mathbf{x}^{(0)} = \tilde{\mathbf{x}}^{(t)}$, $\mathbf{g}^{(0)} = \boldsymbol{\zeta}^{(0)} = M\mathbf{A}_M^T\mathbf{W}_M(\mathbf{A}_M\mathbf{x}^{(0)} - \mathbf{y}_M)$ and $\mathbf{h}^{(0)} = \mathbf{D_A}\mathbf{x}^{(0)} - \boldsymbol{\zeta}^{(0)}$.

for $n = 0, 1, 2, \cdots, N - 1$ **do**

for $m = 0, 1, 2, \cdots, M - 1$ **do** $r = nM + m$

$$\begin{cases} \mathbf{s}^{(r+1)} = \rho(\mathbf{D_A}\mathbf{x}^{(r)} - \mathbf{h}^{(r)}) + (1-\rho)\mathbf{g}^{(r)} \\ \mathbf{x}^{(r+1)} = [\mathbf{x}^{(r)} - (\rho\mathbf{D_A} + \mathbf{D_{S_2}})^{-1}(\mathbf{s}^{(r+1)} + \nabla \mathsf{S}_2(\mathbf{x}^{(r)}))]_+ \\ \boldsymbol{\zeta}^{(r+1)} \triangleq M\mathbf{A}_m^T\mathbf{W}_m(\mathbf{A}_m\mathbf{x}^{(r+1)} - \mathbf{y}_m) \\ \mathbf{g}^{(r+1)} = \dfrac{\rho}{\rho+1}(\alpha\boldsymbol{\zeta}^{(r+1)} + (1-\alpha)\mathbf{g}^{(r)}) + \dfrac{1}{\rho+1}\mathbf{g}^{(r)} \\ \mathbf{h}^{(r+1)} = \alpha(\mathbf{D_A}\mathbf{x}^{(r+1)} - \boldsymbol{\zeta}^{(r+1)}) + (1-\alpha)\mathbf{h}^{(r)} \end{cases}$$

decreasing ρ.

end for

end for

$\tilde{\mathbf{x}}^{(t+1)} = \mathbf{x}^{(NM)}$.

2) Sparse Coding: with $\tilde{\mathbf{x}}^{(t+1)}$ fixed, for each $1 \le l \le L$, update $\tilde{\mathbf{Z}}_l^{(t+1)}$ by (10).

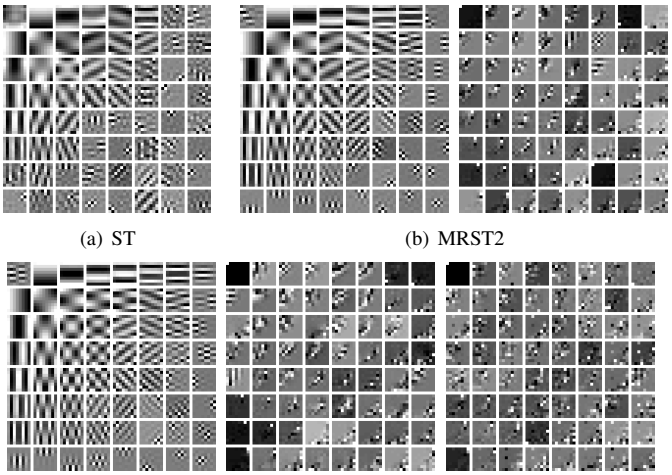
end for

A. Transform Learning

To better understand the potential of the MRST model, we vary the number of layers and pre-learn transforms for ST, MRST2, MRST3, MRST5, and MRST7, which possess 1, 2, 3, 5, and 7 layers, respectively. We used 7 slices of the Mayo Clinic data to train the models. For each model, we run 1000 to 2000 iterations of the learning algorithm to ensure convergence. Fig. 1 shows some of the learned transforms, with each transform matrix row displayed as a square patch for simplicity. The single layer transform displays edge-like and directional structures that sparsify the image. However, with more layers, finer level features are learned to sparsify transform domain residuals in deeper layers. Nonetheless, transforms in deep layers could be more easily contaminated with noise in the training data, since the main image features are successively filtered out over layers.

B. Image Reconstruction

Here, we simulated low-dose CT measurements from full-dose CT images in Mayo Clinic dataset with GE 2D LightSpeed fan-beam geometry corresponding to a monoenergetic source with 10^4 incident photons per ray and no scatter. We ran 100 iterations of the PWLS-EP algorithm with FBP reconstructions as initializations. We used the relaxed OS-LALM algorithm with 4 ordered subsets and regularization parameter $\beta = 2^{15.5}$. For the MRST model, we used the OS-LALM algorithm for the image update step with 2 inner iterations and 4 subsets. We used $T_O = 1000$ and $T_O = 1500$ outer iterations for (ST, MRST2) and (MRST3, MRST5, MRST7), respectively. We set the regularization parameters (after tuning over ranges of values) as $(\beta, \gamma) = (7\times10^4, 20)$ for ST, $(\beta, \gamma_1, \gamma_2) = (3\times10^4, 30, 12)$ for MRST2, $(\beta, \gamma_1, \gamma_2, \gamma_3) = (3\times10^4, 30, 12, 10)$ for MRST3, $(\beta, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (4\times10^4, 30, 20, 10, 10, 10)$ for MRST5,



(a) ST (b) MRST2 (c) MRST3

Fig. 1: Pre-learned transforms for ST, MRST2, and MRST3 (each transform matrix row is shown as a square patch).

and $(\beta, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7) = (7\times10^4, 30, 12, 10, 10, 10, 10, 10)$ for MRST7, respectively.

To compare the performance between various models quantitatively, we compute the root mean square error (RMSE), peak signal to noise ratio (PSNR), structural similarity index measurement (SSIM) of the reconstructions in a region of interest (ROI). The RMSE in Hounsfield units (HU) computed between the ground truth image and reconstruction image is defined as $\text{RMSE} = \sqrt{\Sigma_{i\in\text{ROI}}(\hat{x}_i - x_i^*)^2/N_{\text{ROI}}}$, where $\hat{x}_i$ and x_i^* denote the pixel intensities of the reconstructed and ground truth images, respectively, and N_{ROI} is the number of pixels in the ROI. The ROI here was a circular (around center) region containing all the phantom tissues. Since the full-dose Mayo Clinic CT images still contain some considerable noise, we cannot solely rely on RMSE and PSNR as metrics of image quality, as they would include the noise. SSIM helps better evaluate the preservation of structural details in the reconstructed image.

We conduct experiments on three slices (L067-slice100, L192-slice100, L506-slice100) of the Mayo Clinic data. Fig. 2 shows the reconstruction of L067-slice100 using FBP, PWLS-EP, PWLS-ST, PWLS-MRST2, PWLS-MRST3, PWLS-MRST5, and PWLS-MRST7, respectively at incident photon intensity $I_0 = 10^4$. TABLE I lists the RMSE, PSNR, and SSIM values of reconstructions of the three test slices, with the best values bolded. The two-layer model (MRST2) provides the best RMSE and PSNR values among the methods. However, when we consider the SSIM criterion, MRST5 and MRST7 outperform ST and MRST2. So which MRST model is better? By observing the reconstructed images, we see that although MRST2 and ST have lower RMSE and higher PSNR values than MRST5 and MRST7, they sacrifice some sharpness of the central region and suffer from loss of details. The deeper models have a more positive effect in maintaining subtle features, which is clearly more essential to clinic medical diagnosis. Furthermore, after considerable parameter tuning, we have observed that the deeper models offer more stable image quality as β is varied, i.e., they are more robust to oversmoothing.

V. Conclusion

This paper proposes a general framework for Multi-layer Residual Sparsifying Transform (MRST) learning wherein the transform domain residual maps over several layers are jointly sparsified. Our work then applies MRST learning to low-dose CT (LDCT) image reconstruction by using a PWLS approach with

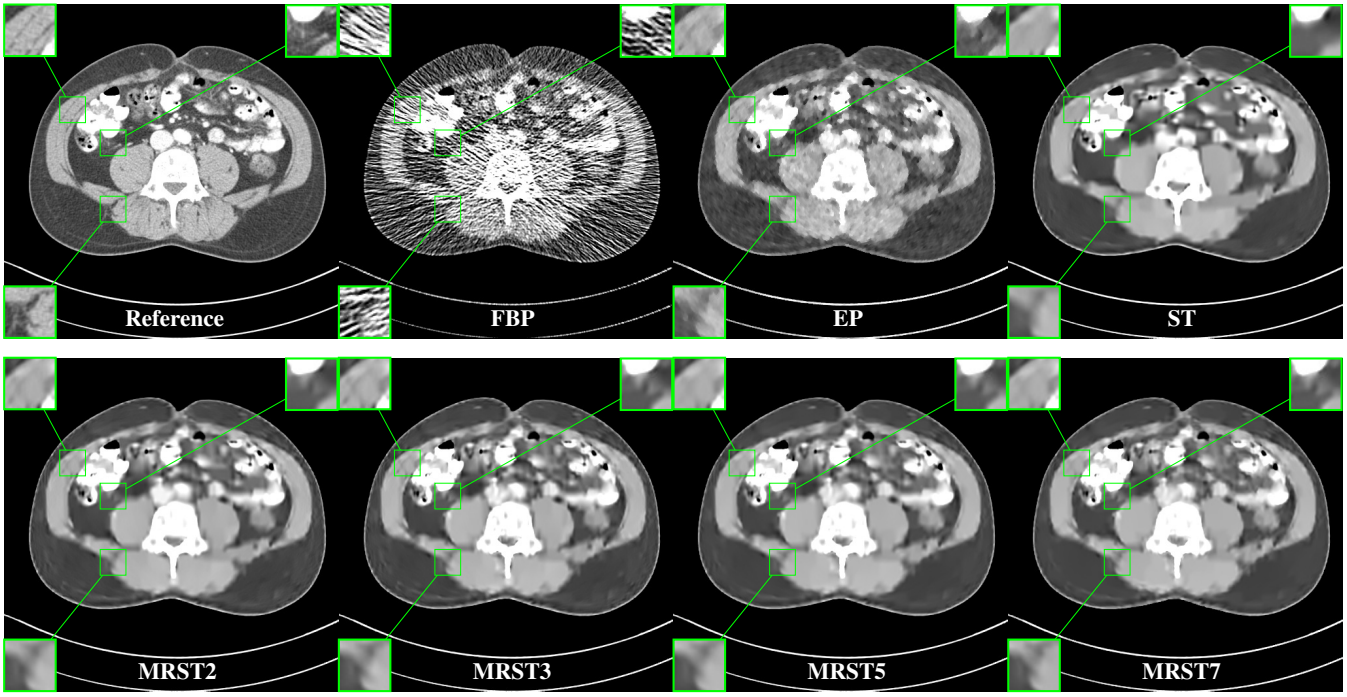


Fig. 2: Comparison of reconstructions of L067-slice100 with FBP, PWLS-EP, PWLS-ST, PWLS-MRST2, PWLS-MRST3, PWLS-MRST5, and PWLS-MRST7, respectively at incident photon intensities $I_0 = 10^4$. The display window is [800, 1200] HU.

TABLE I: RMSE in HU (first row), PSNR in dB (second row) and SSIM (third row) of reconstructions with FBP, PWLS-EP, PWLS-ST, PWLS-MRST2, PWLS-MRST3, PWLS-MRST5, and PWLS-MRST7, for three slices of Mayo Clinic data at incident photon intensity $I_0 = 1 \times 10^4$.

	FBP	EP	ST	MRST2	MRST3	MRST5	MRST7
L067 slice100	101.1	34.2	30.3	**29.1**	29.8	30.5	31.4
	27.4	36.8	37.8	**38.2**	38.0	37.8	37.5
	0.359	0.728	0.724	0.731	0.731	**0.732**	0.731
L192 slice100	60.0	26.6	23.5	**22.2**	22.3	23.0	23.8
	32.4	38.9	39.9	**40.4**	**40.4**	40.1	39.8
	0.440	0.786	0.790	0.798	0.800	0.807	**0.810**
L506 slice100	56.8	32.6	27.1	**26.0**	26.2	27.5	28.7
	32.8	37.6	39.2	**39.6**	39.5	39.1	38.7
	0.489	0.780	0.790	0.791	0.793	0.797	**0.802**

a learned MRST-based regularizer. Experimental results illustrate the promising performance of the multi-layer scheme over single-layer learned sparsifying transforms. Learned MRST models also offer significant improvement over typical nonadaptive methods. In the future, we will investigate additional strategies such as pooling operations to reduce noise over layers.

VI. Acknowledgment

The authors thank Dr. Cynthia McCollough, the Mayo Clinic, the American Association of Physicists in Medicine, and the National Institute of Biomedical Imaging and Bioengineering for providing the Mayo Clinic data.

References

[1] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Asilomar Conf. on Signals, Systems and Computers*, 1993, pp. 40–44 vol.1.

[2] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Sig. Proc.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[3] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Sig. Proc.*, vol. 61, no. 3, pp. 661–677, Feb. 2013.

[4] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Sig. Proc.*, vol. 61, no. 5, pp. 1072–1086, Mar. 2013.

[5] S. Ravishankar and B. Wohlberg, "Learning multi-layer transform models," in *Allerton Conf. on Comm., Control, and Computing*, 2018, pp. 160–165.

[6] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, June 1984.

[7] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 89–99, Feb. 2002.

[8] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Sig. Proc.*, vol. 41, no. 2, pp. 534–548, Feb. 1993.

[9] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, Nov. 2007.

[10] L. Pfister and Y. Bresler, "Model-based iterative tomographic reconstruction with adaptive sparsifying transforms," in *Proc. SPIE*, 2014, vol. 9020, pp. 90200H–1–90200H–11.

[11] X. Zheng, S. Ravishankar, Y. Long, and J. A. Fessler, "PWLS-ULTRA: An efficient clustering and learning-based approach for low-dose 3D CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1498–1510, June 2018.

[12] I. Y. Chun, X. Zheng, Y. Long, and J. A. Fessler, "Efficient sparse-view X-ray CT reconstruction using ℓ_1 regularization with learned sparsifying transform," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, June 2017, pp. 115–119.

[13] S Ye, S. Ravishankar, Y. Long, and J. A. Fessler, "SPULTRA: Low-Dose CT image reconstruction with joint statistical and learned image models," *IEEE Trans. Med. Imag., (Early Access)*, Aug. 2019.

[14] W. Zhou, J-F. Cai, and H. Gao, "Adaptive tight frame based medical image reconstruction: a proof-of-concept study for computed tomography," *Inverse Prob.*, vol. 29, no. 12, pp. 125006, Dec. 2013.

[15] H. Nien and J. A. Fessler, "Relaxed linearized algorithms for faster X-ray CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1090–1098, Apr. 2016.

[16] J. H. Cho and J. A. Fessler, "Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3D X-ray CT," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 678–689, Feb. 2015.

Unsupervised domain adaptation for practical metal artefact reduction in X-ray CT

Muge Du, Hewei Gao, Kaichao Liang, Yinong Liu, Yuxiang Xing*

***Abstract*—Metal artefacts in CT can severely damage the reconstructed images. Current deep learning-based metal artefact reduction (MAR) methods have better performance than traditional methods, but most of them are supervised training and use simulation to generate inputs and labels. The difference between simulated artefacts and practical artefacts results in the degraded performance on real images. To solve this problem, we propose an unsupervised domain adaptation method for MAR to transfer the learned MAR capability from labeled simulated data to unlabeled practical data. The proposed UDAMAR method makes the low-level features extracted for MAR domain-invariant through adversarial training, therefore the performance of MAR is similar on both simulated data and practical data. By simultaneous MAR learning and unsupervised domain adaptation, the proposed method enhances the current supervised MAR method and significantly improves the reduction of practical metal artefacts. The additional cost of training and data collection is small. Experimental studies on dental CT images demonstrate our method is superior to supervised MAR method. Moreover, compared with other deep-learning based MAR methods targeted at practical metal artefacts, our method is more robust and effective on reducing severe practical artefacts in multiple-metal situation.**

***Index Terms*—CT, metal artefact reduction, deep learning, unsupervised domain adaptation, GAN**

I. Introduction

X-RAY computed tomography (CT) has been used in medical diagnostics for decades. CT images containing metals can be affected by metal artefacts, leading to reduced image quality and declined credibility. Traditional metal artefact reduction (MAR) methods model physical effects in a CT scan and correct artefacts accordingly. The main limitation of traditional MAR methods is inaccurate modeling of physical effects. Compared with traditional MAR methods, recently developed MAR methods based on deep learning replace the manual crafted modeling of physical processes with data-driven learning of metal artefact features and achieved better performance. Current deep learning-based MAR methods are mostly supervised learning which requires labels built on artefact-free CT images[1][2][3][4][5][6]. However, it is generally not possible to obtain labels for practical artefact-affected CT images, therefore simulated artefact-affected images and corresponding artefact-free images are used as inputs and labels. A common problem with these MAR methods is the difference between simulation and real artefacts, which causes the method to degrade on practical images. Such problem can be attributed to the domain gap between simulated and practical artefacts.

There are multiple works of supervised and unsupervised MAR trying to reduce the practical metal artefacts. A representative supervised MAR method is I-DL-MAR[7] that starts with interpolation-based MAR (IMAR), followed by a CNN to improve the result in image domain. I-DL-MAR can be directly transferred to practical data, but may introduce new artefacts or wrong structure in the situation of multiple metals or inaccurate segmentation. Unsupervised MAR methods using GAN framework included CycleGAN-MAR[8] and AND[9] are also proposed to tackle the domain gap problem. The results are promising but the uncertainty and inefficiency of unsupervised learning is still a drawback.

To address the domain gap problem and effectively reduce practical metal artefacts, we proposed a new MAR method, called UDAMAR, which can improve the performance of current deep-learning based MAR methods on practical images. Our contributions are as follows:

- We introduce unsupervised domain adaptation for the MAR problem, which aims to transfer the learned MAR capability from simulated data to unlabeled practical data in a novel way at the same time of MAR learning. The key of the method is to use UDA to make the low-level features domain-invariant, while maintaining the ability of MAR.
- The practical images contain realistic artefacts, while simulation helps achieve accurate supervised learning. The proposed method can utilize both labeled simulated images and unlabeled practical images, which has advantages over previous methods.
- We evaluate our method on both simulated and practical dental CT images. Compared with supervised MAR method, our method significantly improves the reduction of practical artefacts. Compared with two state-of-the-art MAR methods for practical artefact reduction[7][8], our method can achieve better results in the presence of multiple metals and severe artefacts.
- To achieve the desired improvement, only a small amount of additional cost is required. The cost of training is an additional domain classifier network, and the minimal cost of

The submitted date is 2020/1/19. Muge Du, Hewei Gao, Kaichao Liang, Yinong Liu, Yuxiang Xing are with the Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education and the department of Engineering Physics in Tsinghua University, Beijing, China. Yuxiang Xing is the corresponding author (E-mail: xingyx@mail.tsinghua.edu.cn). This work was supported by Grants from National key research and development program (No. 2016YFF0101304) and the National Natural Science Foundation of China (Nos. 61771279, 61527807).

data collection is about 200 unlabeled practical CT images, which is acceptable compared to the original dataset.

II. Methods

The domain gap between simulated artefacts and practical artefacts limits the performance of supervised MAR on practical images. Unsupervised domain adaptation (UDA) is a class of transfer learning, which attempts to learn domain invariant representations from source domain and target domain so that the adapted network achieves comparable performance on both domains[10]. It can solve the domain gap issue when target domain has no labels.

In this work, we propose a method called UDAMAR, which learns MAR using simulated images, and simultaneously transfers the MAR ability to unlabeled practical images through unsupervised domain adaptation. We define CT images with simulated artefacts as belonging to source domain $\mathbb{S}$, while CT images with practical artefacts as belonging to target domain $\mathbb{T}$. In Fig. 1, we illustrate our method to combine MAR learning on source domain $\mathbb{S}$ and UDA between source domain $\mathbb{S}$ and target domain $\mathbb{T}$. The complete learning objective can be written as:

$$\Phi(\boldsymbol{\theta}_R) = \sum_{n=1}^{N_\mathrm{S}} \left\| R\left(\boldsymbol{\mu}_n^{\mathrm{a,S}}; \boldsymbol{\theta}_R\right) - \boldsymbol{\mu}_n^{*} \right\|_1 + \lambda \varphi^{\mathrm{DD}}(\mathbb{S}, \mathbb{T}; \boldsymbol{\theta}_R)$$

$$\widehat{\boldsymbol{\theta}}_R = \underset{\boldsymbol{\theta}_R}{\operatorname{argmin}} \Phi(\boldsymbol{\theta}_R) \qquad (1)$$

Here, the superscript " a " means "with artefacts", the superscripts "S" or "T" mean in source or target domain, and λ is a hyper-parameter to adjust the weights of the two terms. The first term in (1) is a supervised learning to train an MAR network R parameterized by $\boldsymbol{\theta}_R$ to effectively reduce metal artefacts on source domain $\mathbb{S}$. Artefact-free images $\boldsymbol{\mu}^*$ are served as labels and the L1 norm of the difference between network output and corresponding labels is used as the loss function.

Obviously, the learned feature extraction of MAR in source domain $\mathbb{S}$ only might be not suited for images in target domain $\mathbb{T}$, thus the capability of MAR is not effective in domain $\mathbb{T}$. The $\varphi^{\mathrm{DD}}()$ in (1) is a function of domain distance computed from features extracted by $\boldsymbol{\theta}_R$. It is to be minimized to get domain invariant features so that R can be valid in the target domain. We refer this part as domain adaption.

A. Domain adaptation with adversarial-training

For domain distance minimization to realize domain adaption, we propose an adversarial loss between domain $\mathbb{S}$ and $\mathbb{T}$ working on the partial features extracted in the network R, in particular, the low-level part of R, so that the extracted low-level features on two domains can be aligned to each other. The reason for adapting low-level features is because we think the difference of CT metal artefacts between simulation and reality lies mostly on small local structures. High-level features are more of semantic information which shall be same for source domain and target domain images. Hence, no adaption is needed for high-level features. We denote the set of features from network R as $\mathbf{f}$, and $\mathbf{f}_{\mathrm{sub}}$ is a subset of $\mathbf{f}$ corresponding to low-level features. We rewrite $\varphi^{\mathrm{DD}}()$ as an adversarial loss:

$$\varphi^{\mathrm{DD}}(\boldsymbol{\theta}_R, \boldsymbol{\theta}_C) = N_\mathrm{T} \sum_{n=1}^{N_\mathrm{S}} \left[C\left(\mathcal{T}\left(\mathbf{f}_{\mathrm{sub}}(\boldsymbol{\mu}_n^{\mathrm{a,S}}; \boldsymbol{\theta}_R)\right); \boldsymbol{\theta}_C\right) - c(\boldsymbol{\mu}_n^{\mathrm{a,T}}) \right]^2$$

$$+ N_\mathrm{S} \sum_{n=1}^{N_\mathrm{T}} \left[C\left(\mathcal{T}\left(\mathbf{f}_{\mathrm{sub}}(\boldsymbol{\mu}_n^{\mathrm{a,T}}; \boldsymbol{\theta}_R)\right); \boldsymbol{\theta}_C\right) - c(\boldsymbol{\mu}_n^{\mathrm{a,S}}) \right]^2, \forall \boldsymbol{\theta}_C \qquad (2)$$

Here, similar to GAN, we introduce a domain classification network C parameterized by $\boldsymbol{\theta}_C$ for the metric of domain distance. The class label is defined by $c(\boldsymbol{\mu}) = \begin{cases} 1 & \text{if } \boldsymbol{\mu} \in \mathbb{S} \\ 0 & \text{if } \boldsymbol{\mu} \in \mathbb{T} \end{cases}$. It is worth noting that the inputs and labels in (2) are reversely paired because domain distance is measured by domain confusion. The multiplication of N_T and N_S in each term is to balance the effect of different number of samples. In order to combine the subset features $\mathbf{f}_{\mathrm{sub}}$ to uniform size, we introduce a feature transformation $\mathcal{T}$ by that each intermediate feature is up-sampled into the same size and concatenated. The combined feature $\mathcal{T}\left(\mathbf{f}_{\mathrm{sub}}(\boldsymbol{\mu}^{\mathrm{a}}; \theta_R)\right)$ is then input to C.

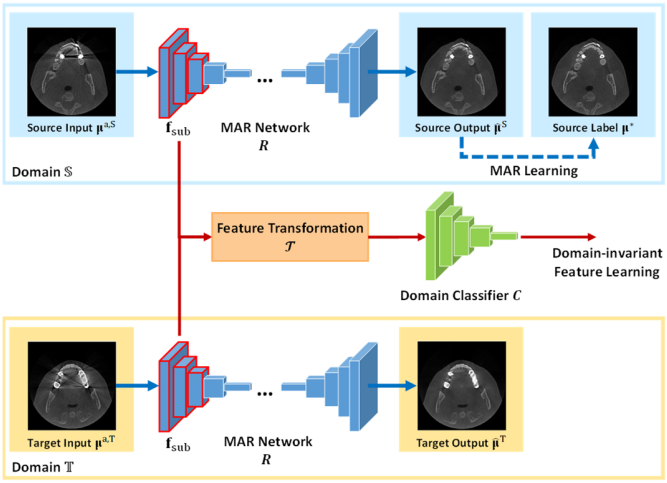


Fig. 1 Illustration of UDAMAR method

Combining (1) and (2), we get our final optimization problem:

$$\widehat{\boldsymbol{\theta}}_R, \widehat{\boldsymbol{\theta}}_C = \arg \min_{\boldsymbol{\theta}_R} \max_{\boldsymbol{\theta}_C} \textstyle\sum_{n=1}^{N_\mathrm{S}} \left\| R\left(\boldsymbol{\mu}_n^{\mathrm{a,S}}; \boldsymbol{\theta}_R\right) - \boldsymbol{\mu}_n^{*} \right\|_1 +$$

$$\lambda N_\mathrm{T} \textstyle\sum_{n=1}^{N_\mathrm{S}} \left[C\left(\mathcal{T}\left(\mathbf{f}_{\mathrm{sub}}(\boldsymbol{\mu}_n^{\mathrm{a,S}}; \boldsymbol{\theta}_R)\right); \boldsymbol{\theta}_C\right) - c(\boldsymbol{\mu}_n^{\mathrm{a,T}}) \right]^2 +$$

$$\lambda N_\mathrm{S} \textstyle\sum_{n=1}^{N_\mathrm{T}} \left[C\left(\mathcal{T}\left(\mathbf{f}_{\mathrm{sub}}(\boldsymbol{\mu}_n^{\mathrm{a,T}}; \boldsymbol{\theta}_R)\right); \boldsymbol{\theta}_C\right) - c(\boldsymbol{\mu}_n^{\mathrm{a,S}}) \right]^2 \qquad (3)$$

Here, the least-square form is borrowed from least squares GAN[11] that helps to stable the training. This optimization can be done by adversarial training between C and R. The operator $\underset{\boldsymbol{\theta}_C}{\operatorname{argmax}}$ in (3) is to discriminate whether the original input image $\boldsymbol{\mu}^{\mathrm{a}}$ is from domain $\mathbb{S}$ or $\mathbb{T}$. C is optimized to find domain-specific features, while R is optimized to confuse C through generating domain-invariant features.

B. Network structure

The MAR network R adopts an encoder-decoder structure. The intermediate outputs of first k layers in R's encoding path are used as $\mathbf{f}_{\mathrm{sub}}$. The domain classifier C adopts a patch-GAN discriminator[12] architecture.

III. Experimental study

A. Experiment Configurations

Data preparation For source domain dataset, we first acquire CT images without metal components from a commercial cone-beam dental CT. Simulated metals are added randomly to these images to form digital phantoms. A 120kVp spectrum is simulated for getting projections of the X-ray path passing metal regions. The original data together simulated metal-path data are reconstructed to form images with artefacts. More detail of the simulation can be found in [7]. In total, 2704 metal-free CT images are used to generate simulated artefact-affected images. The image pairs are used as the input and label in source domain dataset.

For target domain dataset, 1601 practical images with metal implants are collected. Data was acquired from the same type commercial cone-beam dental CT, and we only use the mid-plane for required 2D images.

The images in both domains are 640×640, with pixel size 0.25mm. Both datasets are split into training set, test set and validation set by 7:2:1.

Implementations Both network R and C are trained using Adam optimizer with learning rate of 0.0002. The $\mathbf{f}_{\text{sub}}$ is selected to be the outputs of first 3 layers of R. Both UDA-MAR and supervised MAR are trained for 150 epochs. In order to prevent the fluctuation of early-stage adversarial loss from hindering the supervised training of R, we use following strategies: We firstly train R for optimizing only with $\lambda = 0$ for 70 epochs, then train C and R with linearly increasing λ in the 71st-150th epochs. The maximum λ is 0.01.

B. Experiment 1: practical metal artefacts reduction

We first performed simulation study to test the learned MAR capability of UDAMAR and supervised MAR on simulated images. The results are evaluated quantitatively by relative-root-mean-square-error (RRMSE) for the nonmetal regions:

$$\text{RRMSE}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}^*) = \frac{\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_2}{\|\boldsymbol{\mu}^*\|_2} \quad (4)$$

The qualification results of reducing simulated metal artefacts is showed in Fig. 2. Table I indicates that the adversarial training in UDAMAR would cause a small increase in RRMSE. From Fig. 2, we can see that supervised MAR and UDAMAR can effectively remove the simulated metal artefacts to a similar level.

TABLE I
MEAN RRMSE FOR SIMULATED ARTEFACTS REDUCTION

Method	RRMSE
Supervised MAR	0.0614
UDAMAR	0.0698

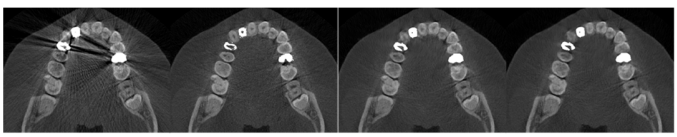


Fig. 2 Example results of simulation study in experiment 1

We also conducted a practical study comparing the capabilities of the proposed UDAMAR with supervised MAR and two methods, I-DL-MAR and CycleGAN-MAR, in terms of reducing real metal artefacts. I-DL-MAR is trained on paired input and labels generated from source domain dataset for 150 epochs, and CycleGAN-MAR is trained on unpaired images in two domains for 300 epochs, where one domain is the artefact-affected images from our target domain $\mathbb{T}$ and the other domain is the artefact-free images from our source domain $\mathbb{S}$. In the experiment, the MAR network structure of I-DL-MAR is set as the same encoder-decoder as R in UDAMAR. However, for CycleGAN-MAR, we found that the training on full images is easy to fall into "mode collapse" when using the same encoder-decoder network, so we use a U-net structure same as in [8], and adopt patch-based learning with patch size of 320×320.

The summary of datasets used for above methods is given in Table II. The qualification results are showed in Fig. 3. Three cases with artefacts of different severity are displayed.

TABLE II
DATASETS FOR DIFFERENT METHODS

Method	Dataset used in training
Supervised MAR	Paired images with and without artefacts from simulation
UDAMAR	Paired images with and without artefacts from simulation; practical images with artefacts
I-DL-MAR	Images without artefacts and simulated metal traces
CycleGAN-MAR	Unpaired practical images with and without artefacts

The first case represents a common situation in reality, when there are only a few metal implants with mild artefacts. Supervised MAR can merely reduce part of the artefacts, while UDAMAR successfully transfers the MAR capability to target domain and reduces artefacts. I-DL-MAR is slightly better than UDAMAR, showing its powerfuness in handling common cases with mild practical metal artefacts. CycleGAN-MAR's performance is between supervised MAR and UDA-MAR.

In the second case with more severe artefacts, the results are similar to the first case, except that supervised MAR performs even worse. UDAMAR can still reduce most of the artefacts effectively. I-DL-MAR also performs nicely except that there is some steaking artefacts due to under-segmentation. The performance of CycleGAN-MAR is similar to supervised MAR.

The third case is with very severe artefacts. In this case, supervised MAR and CycleGAN-MAR almost fail. Since interpolation step discards too much information when multiple metals exist, I-DL-MAR leads to severe new artefacts and incorrect structures. In the contrast, UDAMAR did well in reducing most of the artefacts and recover most structures.

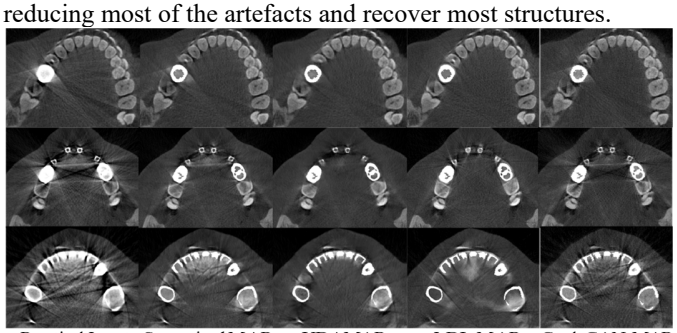


Fig. 3 Example results of practical study in experiment 1.

C. *Experiment 2: Ablation study on few-shot unsupervised domain adaptation*

To further examine the performance of UDAMAR, we conducted a study on few-shot unsupervised domain adaptation, to test the minimal number of samples in target domain that can effectively improve the performance.

In the study, the number of target domain images is varied from 50 to 1339, to test the quality of MAR results on target domain when limited target domain samples are used for unsupervised domain adaptation. To ensure the results are not biased by the sample distribution, we divide the entire target domain dataset into three groups, according to the number of metal implants in the image, and make the sampling propotion the same for each group. In addition, we ensure that the smaller dataset is included by the larger dataset, i.e., the study is in the sense of dataset growing.

The results are shown in Fig. 4. The performance of UDAMAR is similar to supervised MAR when using 50 target domain images. UDAMAR shows obvious improvement compared to supervised MAR when increasing target domain samples to 100, which demonstrating the effect of overcoming domain gap. UDAMAR using 200 target samples can ensure a rather good performance in MAR in this experiment. As a dataset of 200 unlabeled samples is fairly accessible in reality, our method is very practicable.

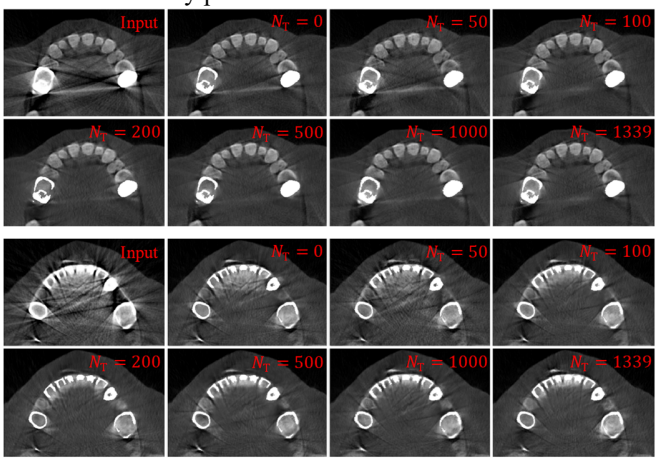


Fig. 4 Example results for practical MAR with different N_T, i.e. size of target domain dataset. N_T=0 stands for supervised MAR. Two cases are shown here.

IV. Conclusion

Most MAR methods using deep-learning are based on supervised learning, which requires simulation of artefacts for training. Due to the domain gap between simulated artefacts and practical artefacts, the performance degrades on practical images. To effectively reduce practical artefacts, we propose an unsupervised domain adaptation method for MAR, called UDAMAR. The method combines MAR learning on simulated data, and adversarial-based UDA between simulated and practical data with artefacts. Even though practical data has no labels, the proposed method effectively transfers the learned MAR ability to practical data. Experimental study proves the proposed method can remove practical metal artefacts far more effectively than supervised methods trained on simulated data, with a small additional cost of training and data collection. Compared with state-of-the-art supervised method for reducing practical metal artefacts, the proposed method has better performance in the existence of multiple metal implants and severe artefacts. Compared with other unsupervised MAR methods, the proposed method can learn MAR more effectively with simpler network architecture and also is easy to be trained.

The proposed method is a novel way to utilize unlabeled medical images in pixel-level optimization tasks, therefore has potential value for tasks that require transfer learning but lack labeled data, such as transferring deep learning methods across CT datasets with different scan parameters. In the future, we will explore our method on more potential applications.

References

[1] L. Gjesteby *et al.,* "Deep Learning Methods for CT Image-Domain Metal Artifact Reduction," *Developments in X-Ray Tomography Xi*, vol. 10391, Proceedings of SPIE, 2017.

[2] X. Huang, J. Wang, F. Tang, T. Zhong, and Y. Zhang, "Metal artifact reduction on cervical CT images by deep residual learning," *Biomed Eng Online,* vol. 17, no. 1, p. 175, Nov 27 2018, DOI: 10.1186/s12938-018-0609-y.

[3] H. S. Park, S. M. Lee, H. P. Kim, J. K. Seo, and Y. E. Chung, "CT sinogram-consistency learning for metal-induced beam hardening correction," *Med Phys,* vol. 45, no. 12, pp. 5376-5384, Dec 2018, DOI: 10.1002/mp.13199.

[4] Y. Zhang and H. Yu, "Convolutional Neural Network Based Metal Artifact Reduction in X-Ray Computed Tomography," *IEEE Trans Med Imaging,* vol. 37, no. 6, pp. 1370-1381, Jun 2018, DOI: 10.1109/TMI.2018.2823083.

[5] W.-A. a. L. Lin, Haofu and Peng, Cheng and Sun, Xiaohang and Zhang, Jingdan and Luo, Jiebo and Chellappa, Rama and Zhou, Shaohua Kevin, "DuDoNet: Dual Domain Network for CT Metal Artifact Reduction," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2019, pp. 10512-10521.

[6] C. Zhang and Y. Xing, CT artifact reduction via U-net CNN (SPIE Medical Imaging). SPIE, 2018.

[7] K. Liang, L. Zhang, H. Yang, Y. Yang, Z. Chen, and Y. Xing, "Metal artifact reduction for practical dental computed tomography by improving interpolation-based reconstruction with deep learning," *Med Phys,* vol. 46, no. 12, pp. e823-e834, 2019 2019, DOI: 10.1002/mp.13644.

[8] M. Du, K. Liang, and Y. Xing, "Reduction of metal artefacts in CT with Cycle-GAN," in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC),* 10-17 Nov. 2018 2018, pp. 1-3, DOI: 10.1109/NSSMIC.2018.8824544.

[9] H. Liao, W. Lin, S. K. Zhou, and J. Luo, "ADN: Artifact Disentanglement Network for Unsupervised Metal Artifact Reduction," *IEEE Trans Med Imaging,* pp. 1-1, 2019, DOI: 10.1109/TMI.2019.2933425.

[10] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial Discriminative Domain Adaptation," in *The IEEE International Conference on Computer Vision (ICCV),* 2017.

[11] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least Squares Generative Adversarial Networks," arXiv e-prints. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv161104076M

[12] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV),* 22-29 Oct. 2017 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.

Local Motion Estimation for Improved Cone-Beam CT Deformable Motion Compensation

A. Sisniega, S. Capostagno, W. Zbijewski, J.W. Stayman, C. R. Weiss, T. Ehtiati, J. H. Siewerdsen

***Abstract*—Cone-beam CT (CBCT) is commonplace for 3D guidance in interventional radiology (IR), but it is affected by degradation from complex, non-periodic, soft-tissue deformable motion associated with long image acquisition time. Multi-region autofocus methods were shown to provide deformable motion compensation through a combination of locally rigid motion trajectories. However, convergence of the optimization is challenged by non-convexity and high-dimensionality of the problem. Preconditioning with estimates of local motion characteristics can help to achieve convergence, but they require a reliable way to automatically estimate local features of motion. This work presents an automatic framework to obtain local estimates of motion amplitude and direction of the deformable motion field.**

The principal direction and amplitude of local motion were obtained with a multi-branch deep convolutional neural network (CNN) acting on three orthogonal planes of a volumetric ROI of the motion-corrupted volume, yielding estimations of the three components (antero-posterior, lateral, and superior) of the local motion vector. Local motion estimates were used to drive a preconditioning strategy to favor more likely solutions through spatially varying penalty designs acting on the motion trajectory and spatially varying tuning of the search range for the optimization algorithm. Training and validation data were generated with realistic simulations obtained from a publicly available database of abdominal CT datasets. Simple motion fields were induced in the CT volumes to obtain synthetic motion-corrupted CBCT data.

Predictions of individual components of local motion and total motion amplitude showed good agreement with the true values, with errors < 20 mm for the range of motion amplitudes explored. Volumetric distributions of the individual motion components in representative cases showed RMSE < 2.1 mm for the full volume. Spatially varying regularization and search range driven by motion characteristics yielded improved motion compensation after 1000 iterations in an example case featuring complex deformable motion.

Spatially varying preconditioning through estimation of local motion properties showed improved performance of deformable motion estimation. The proposed approach shows the feasibility for automatic local motion estimation with learning-based methods.

***Index Terms*—Motion compensation, cone-beam CT, CBCT, intraoperative imaging, soft-tissue imaging.**

I. Introduction

Cone-beam CT (CBCT) is now commonplace for 3D guidance and quality assurance in surgical and interventional radiology (IR) applications, including spinal [1], cardiac [2], and vascular procedures [3]. Intraprocedural 3D guidance provided by CBCT offers major benefit to minimally invasive IR procedures in the abdomen: for example, in trans-arterial chemoembolization (TACE), the visibility of small feeder vessels, not achievable with conventional 2D imaging, is crucial for selective embolization to spare as much healthy tissue as possible. However, the relatively long acquisition time of CBCT (4 s - 30 s) often results in image quality degradation from involuntary patient motion, with up to 10% [4] of CBCT datasets in TACE resulting in non-diagnostic quality images, and 25% of cases [5] showing some degree of image quality degradation. Even under breath-hold conditions, various sources of involuntary motion (e.g. cardiac, or peristalsis) challenge the delineation of small feeder vessels.

The muti-source nature of abdominal motion, which results in a complex deformable motion field, combined with the stringent workflow requirements in IR interventions, confounds conventional motion compensation methods, which often invoke assumptions of periodicity or use external trackers.

Recent work has demonstrated the feasibility of deformable motion compensation in soft-tissue abdominal CBCT by posing motion as an "autofocus" optimization problem [6]. The autofocus concept was previously employed for rigid motion compensation in high-contrast CBCT imaging by our [7] and other groups [8]. Deformable motion was estimated as a combination of locally rigid motion trajectories through minimization of a cost-function including an autofocus term acting on a set of regions of interest (ROIs) distributed across the CBCT field of view. Spatio-temporal regularization was added to penalize unrealistic abrupt motion trajectories and abrupt spatial transitions in the motion field. The resulting method made no assumptions on motion periodicity and required solely the acquired CBCT data, with no need of external trackers or physiological signals. The multi-ROI joint autofocus cost-function poses a non-convex problem with multiple local minima that challenge the optimization, slowing convergence or even resulting in non-convergence for very complex motion fields.

Characterization of motion characteristics (e.g. amplitude, approximate direction, and principal frequency) provides improved initialization of parameters and reduces the complexity of the optimization problem. In addition, characterization of local (i.e. per ROI) motion characteristics yields further preconditioning through spatially varying regularization techniques and narrowing of the local solution search space. While evaluation of motion characteristics through visual inspection is not feasible, ongoing work [9] has shown the potential of deep learning architectures to estimate the amplitude of local motion and integration of this knowledge into the optimization problem through spatially varying temporal regularization and search range of the optimizer.

In this work we propose a new method for characterization of local motion properties to achieve local estimation of the principal direction of motion and its amplitude through an automatic, multi branch, decision network acting on orthogonal planes (axial sagittal, and coronal) of local volumes of interest in the motion corrupted volume.

II. Methods

Fig. 1 summarizes the complete architecture for motion compensation, including the proposed multi-branch decision network and the motion compensation method, with illustration of

This work was supported by industry-academy collaboration with Siemens Healthineers.
A. Sisniega, S. Capostagno, W. Zbijewski, J. W. Stayman, and J. H. Siewerdsen are with the Dept. of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA.
C. R. Weiss is with the Russell H. Morgan Department of Radiology, Johns Hopkins University, Baltimore, MD, USA.
T. Ehtiati is with Siemens Healthineers, Forchheim, Germany.

the potential for motion compensation with proper initialization (Fig. 1B).

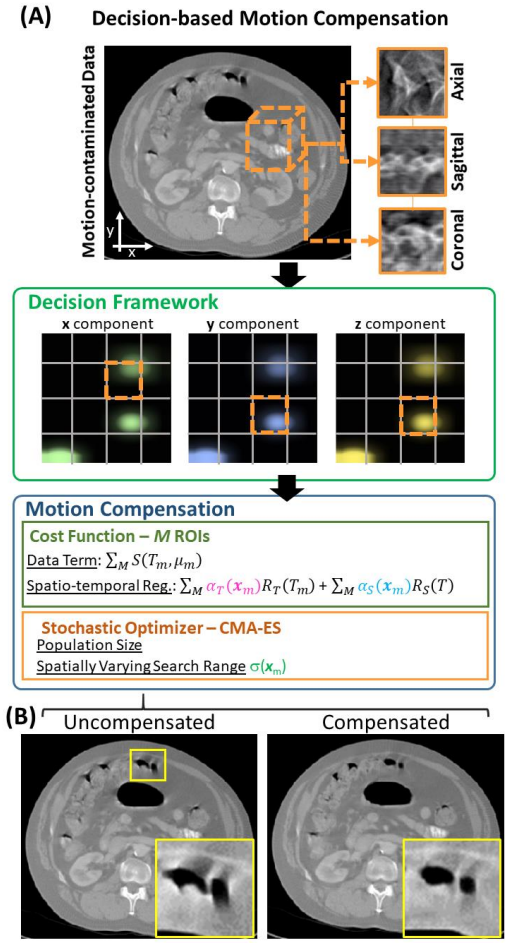


Figure 1. (A) Schematic illustration of the motion compensation solution, including the automatic decision framework and the autofocus motion compensation. The decision framework provides local estimation of the motion amplitude and direction that is used for selection of local spatially varying regularization in the autofocus cost-function.

A. Autofocus-based deformable motion compensation

The deformable motion field is assumed smooth enough to be considered rigid within a local neighborhood. The complete motion field can then be modelled as a set of local motion trajectories associated with a grid of M ROIs, each following a 6 degrees of freedom (DoF) trajectory T_m (m = 1, ..., M). Each DoF in the local motion trajectorie (T_m) is represented with a cubic b-spline with N knots (< N_θ, number of angular projections) to reduce the dimensionality of the optimization problem. A multi-region joint autofocus problem is posed with the following cost function:

$$\hat{T} = \arg\min_{T} \sum_M S(T_m, \mu_m) + \sum_M \alpha_T(\boldsymbol{x}_m) \cdot R_T(T_m) + \alpha_S(x_m) \cdot R_S(T) \quad (1)$$

where $S(T_m, \mu_m)$ is an autofocus metric computed in the reconstructed volume, μ_m (for ROI m) obtained using motion trajectory T_m. The regularization term $R_T(T_m)$ computes the L2 norm of the first order difference trajectory in time to penalize abrupt changes in time of the local motion, while the penalty term $R_S(T)$ encourages similarity between motion trajectories for ROIs close in space. The two penalty terms are weighted by the spatially varying scalar maps $\alpha_T(\boldsymbol{x}_m)$ and $\alpha_S(\boldsymbol{x}_m)$, dependent on local motion amplitude and principal direction, respectively, where $\boldsymbol{x}_m$ denotes the position of ROI m in the volume.

A stochastic optimization method (CMA-ES [10]) was used to minimize (1). CMA-ES obtains a solution through random search of the motion space following a multi-dimensional normal distribution with search space defined by its standard deviation (σ). The standard deviation was individually set per ROI and DoF, creating a spatial distribution $\sigma(\boldsymbol{x}_m)$, according to the estimated motion amplitude and direction.

B. Characterization of motion properties

Following our previous work for automatic estimation of local motion amplitude [9], and previous work for estimation of motion direction in cardiac CT [11], we propose a novel automatic decision framework to provide coarse estimates of local motion amplitude and principal direction for preconditioning of the autofocus problem defined in Eq. (1).

The proposed framework is built using a custom multi-input, multi-branch convolutional neural network (CNN) architecture. Local amplitude of motion in three orthogonal directions (aligned with the antero-posterior, lateral, and superior planes of the volume) was estimated in arbitrary volumetric ROIs placed anywhere in the volume. The network (Fig. 2) consists of three equivalent branches acting on the three central orthogonal slices (axial, sagittal, and coronal) of the motion-contaminated ROI, respectively. Each branch features an initial convolutional layer followed by five residual blocks (denoted ResBlock, similar to ResNet architectures [12]). Each ResBlock performs three convolution stages, each followed by a batch normalization layer, and a final rectified linear unit (ReLU), acting on the addition of the main path and the residual. Average pooling layers were added after each ResBlock, with a final global average pooling at the end of each branch. The outputs of the individual branches were concatenated and input to a final dense layer with linear activation that generated the final three motion components, one for each of the three spatial directions. The network was applied to a grid of ROIs to generate the spatial distribution of the principal direction of motion for the motion-corrupted volume.

Training data (Fig. 3) were generated from 75 real abdominal MDCT volumes from the "CT Lymph Nodes" collection of The Cancer Imaging Archive (TCIA) database. Each training instance was generated by applying a realistic deformable motion field to a 160 mm long subvolume centered at a random location between the lower abdomen and diaphragm. Deformable motion was induced at a random location ($\boldsymbol{x}_m$) in the volume, with maximum amplitude ranging from 0 to 60 mm, and fading in space according to a three-dimensional Gaussian function with full-width at half-maximum ranging from 20 mm to 80 mm, set independently for the three directions. The direction of motion was randomly selected, favoring physiologically realistic motion fields by allocating at least 40 % of the motion in the antero-posterior direction (denoted AP) and 30 % of the motion in the inferior-superior direction (denoted SUP), leaving lower amplitude in the less likely lateral direction (denoted LAT). The motion trajectory followed a cosine function in time, with frequency randomly set to a range between one and two periods in the simulation, with random phase. The data were further augmented by including a random in-plane rotation (± 10 deg).

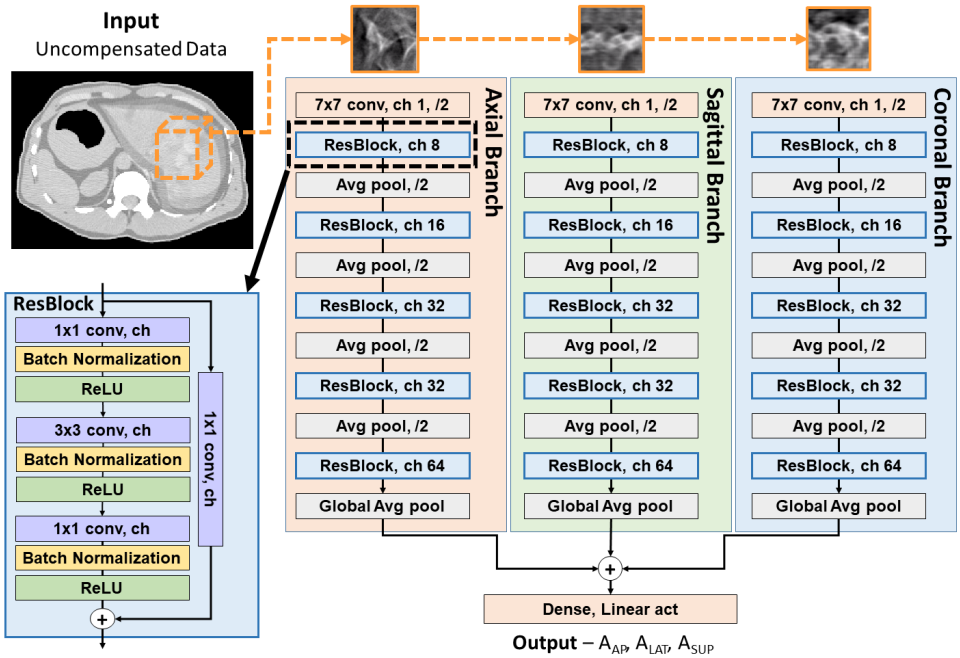


Figure 2. Topology of the multi-branch CNN-based decision framework proposed to predict the local amplitude and principal direction of motion. The network receives as input the central axial, sagittal, and coronal slices of a motion-corrupted volumetric ROI and outputs a 3-component vector with the estimated amplitude of motion for the AP, LAT, and SUP directions, assuming the motion inside the ROI is locally rigid.

CBCT motion-corrupted synthetic data were generated by forward projection of the deformed volume according to the motion field and temporal trajectory. The simulation used a system geometry with source-to-detector distance (SDD) of 1200 mm, and source-to-axis distance (SAD) of 900 mm, pertinent to interventional C-arm systems. Motion-contaminated volumes were obtained with the FDK algorithm on a 320 x 320 x 160 voxel grid (1 x 1 x 1 mm voxel size). For each motion-corrupted volume, 10 individual training instances were generated by selecting 10 ROIs (64 x 64 x 64 pixels), randomly placed in the volume.

Training was performed using the Adam optimizer with learning rate 10^{-3} and L2 regression loss. The network was trained on ~$6x10^4$ images with a batch size of 64 for 150 epochs.

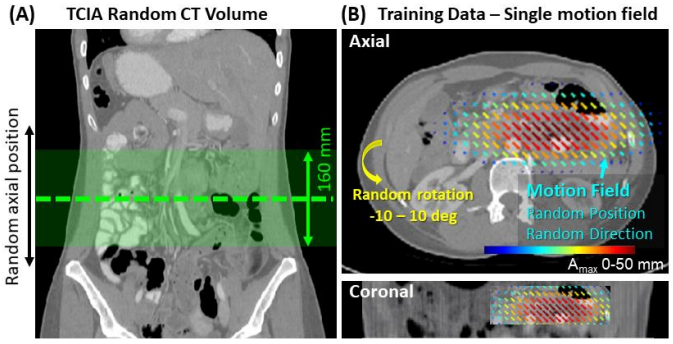


Figure 3. (A) Example abdominal MDCT volume from the TCIA database used for training and performance evaluation. (B) Training data were obtained from a randomly selected subvolume featuring a deformable motion field with random amplitude and motion direction, placed at a random position in the volume.

The performance of the decision framework was assessed on a dataset of 500 ROIs obtained from 20 volumes generated analogously to the training instances. A total of 25 ROIs was generated per volume, arranged following a regular grid in the central axial slice. Each ROI was independently input to the network and the estimated motion properties were posteriorly rearranged into a grid for visualization of the local motion properties into the context of the motion corrupted volume.

The benefit of spatially varying regularization and preconditioning was evaluated in a separate instance of motion-contaminated data. In the current study, only the total motion amplitude was incorporated in the autofocus motion compensation through spatial variation of the temporal penalty strength ($\alpha_T(\boldsymbol{x}_m)$). Ongoing work aims to incorporate motion direction in the spatial penalty strength ($\alpha_S(\boldsymbol{x}_m)$). The spatial map $\alpha_T(\boldsymbol{x}_m)$ was derived following an inverse linear relationship with the local predicted motion amplitude, acting on a nominal scalar α_T, empirically selected according to previous results on autofocus deformable motion compensation [6]. The local search range for CMA-ES, $\sigma(\boldsymbol{x}_m)$ was set to 5% of the estimated local motion amplitude. Motion compensation was performed using 9 ROIs and 1000 CMA-ES iterations.

III. Results

Fig. 4 shows the agreement between local motion direction and amplitude illustrated as spatial maps of true and estimated motion components for 25 instances in the evaluation dataset, for the AP (Fig. 4B), LAT (Fig. 4C), SUP (Fig. 4D), and total motion amplitude (Fig. 4E). Motion components were individually estimated for each ROI in Fig. 4A and rearranged into spatial maps. Reasonable agreement is observed for individual components, with RMSE = 1.49 mm, 0.48 mm, and 1.10 mm for the AP, LAT, and SUP directions, respectively. The spatial map of total motion amplitude showed RMSE = 2.1 mm.

Aggregated performance metrics for the complete test dataset showed errors below 20 mm over the entire motion range. However, the limited size of the training dataset resulted in non-optimal training results, with only moderate generalization to the validation dataset. Training with a larger dataset, featuring a wider variety of motion patterns is underway and anticipated to yield corresponding improvements in pre-conditioning.

The effect of preconditioning and spatially varying regularization in autofocus deformable motion compensation is illustrated in Fig. 5. The volume featured regions with severe motion degradation while some other areas remained almost steady. This heterogeneity of the

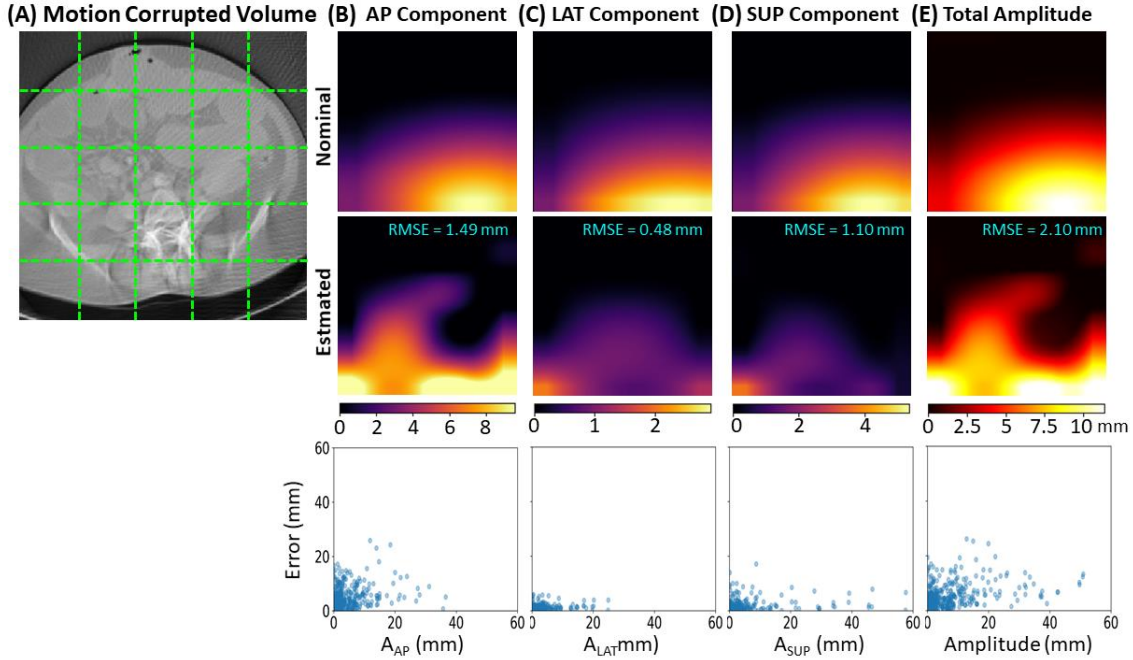


Figure 4. Local motion direction and amplitude estimation for a case with a single motion field and 25 ROIs arranged in a regular grid (A). Motion estimations showed good agreement between the predicted and nominal local motion components, both in spatial distribution and magnitude with RMSE = 1.49 mm, 0.48 mm, and 1.10 mm for the AP (B), LAT (C), and SUP (D) directions, respectively. The total motion amplitude, derived from the three components (D), showed RMSE = 2.10 mm. Aggregated results from all cases in the test dataset errors below 20 mm for the whole range of motion amplitude explored.

motion field challenged the convergence of the autofocus optimization, which was not able to achieve motion compensation within 1000 iterations using the conventional (un-conditioned) approach. The conventional method gave only partial motion compensation in motion-contaminated regions and – interestingly – distorted other areas that were only mildly affected by motion. Inclusion of spatially varying regularization and search space maps generated from the estimated motion amplitude yielded superior motion compensation in regions affected by motion, preserved regions with little or no motion, and provided faster convergence in areas undergoing large motion deformation, yielding satisfactory compensation at 1000 iterations.

VI. Conclusion

Preconditioning of autofocus deformable motion compensation was achieved with spatially varying regularization and search space limitation driven by local estimates of motion characteristics (motion amplitude and principal direction). The preconditioned optimization achieved faster convergence thanks to a simplified search space that facilitates avoidance of local minima in the non-convex optimization problem posed by multi-region autofocus.

The learning-based motion characterization network provided fairly accurate estimation of local motion amplitude and direction, using motion corrupted volumetric ROIs as inputs. Preliminary studies with limited size training datasets show the potential of the proposed approach, with ongoing work targeting more extensive training, validation, and integration of the estimated motion properties on the spatial regularization term of the autofocus cost-function.

References

[1] M.J. Wallace, et al., J. Vasc. Interv. Radiol. 19(6), 799–813 (2008).
[2] M. Akhtar et al., Catheter. Cardiovasc. Interv. 66(1), 43–49 (2005).
[3] C.B. Lightfoot et al., J. Vasc. Interv. Radiol. 24(7), 981–988 (2013).
[4] A. Tognolini, et al., J. Vasc. Interv. Radiol. 21(3), 339–347 (2010).
[5] I.J. Lee et al., J. Vasc. Interv. Radiol. 25(9), 1369–1379 (2014).
[6] A. Sisniega, et al., Proc. SPIE 10948, Medical Imaging 2019, 109481O (2019).
[7] A. Sisniega, et al, Phys. Med. Biol. 62 3712, (2017).
[8] J. Hahn, et al. Proc. SPIE 9783, Medical Imaging 2016, 97831A (2016).
[9] A. Sisniega, et al., Proc. SPIE, Medical Imaging 2020, Accepted (2020).
[10] N. Hansen and S. Kern, Proc. 8th Int. Conf. Parallel Probl. Solving, 282–291 (2004).
[11] T. Lossau, et al. Med. Image Anal., 52, 68-79, (2019).
[12] K. He, et al., Arxiv.Org 7(3), 171–180 (2015).

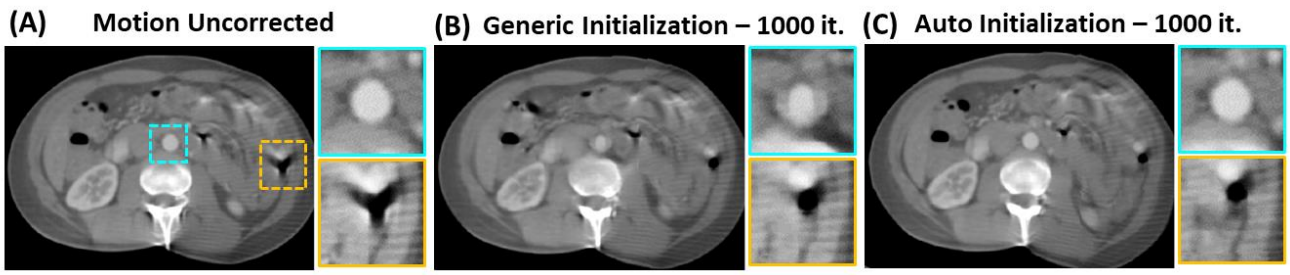


Figure 5. The benefit of preconditioning on autofocus deformable motion compensation with spatially varying temporal regularization and CMA-ES search range derived from estimated local motion amplitude. (A) The motion-corrupted volume showed highly non-homogeneous motion degradation with regions highly distorted by motion (orange ROI) and areas with minimal motion-induced degradation (blue ROI). (B) The conventional approach (involving motion-invariant regularization) did not achieve satisfactory motion compensation after 1000 iterations, showing only partial compensation of motion artifacts (orange) and – worse – distortion of regions not affected by motion (blue), which is a consequence of intermediate (non-converged) degenerate solutions. (D) Application of spatially varying regularization and restriction of the CMA-ES search space prevented such intermediate degeneracies and yielded faster convergence in areas with significant motion-induced image quality degradation.

Motion Resilient Iterative 3D CBCT Reconstruction Based on Gradient Weighting

Igor Peterlik, Adam Strzelecki, Mathias Lehmann, Philippe Messmer, Peter Munro, Pascal Paysan, Mathieu Plamondon and Dieter Seghers

***Abstract*—In cone-beam computed tomography (CBCT), the iterative reconstruction (IR) techniques provide an advantage over the analytical approach based on the filtered back-projection: The underlying iterative optimization typically employs a regularizer capable of better suppression of noise as well as the cone-beam artifacts and in the case of statistical IR, another type of artifacts due to photon starvation is reduced. However, due to longer acquisition times (up to dozens of seconds in some protocols), the acquired data often suffer from motion-related inconsistencies, which in turn lead to significant streak artifacts.**

Based on an analysis of a large number of reconstructions of 3D CBCT images acquired during radiation therapy, we propose a modified 3D CBCT IR method: The first-pass estimation of the 3D volume computed via algebraic reconstruction is refined with a statistical reconstruction method with modified gradients where signals transmitted exclusively through the air are down-weighted in the projection space.

Evaluation on four clinical scans acquired with Varian treatment devices shows that the proposed method significantly reduces motion-related artifacts in iteratively reconstructed CBCT images, while it maintains the benefits of IR. Moreover, as proposed modification of gradients requires a negligible additional computation cost, the resulting reconstruction method remains fast enough to be employed clinically.

***Index Terms*—3D CBCT, iterative image reconstruction, algebraic reconstruction technique, maximum log-likelihood, motion compensation.**

I. Introduction

The techniques based on the analytical approach of filtered back-projection (FBP) were employed in computed tomography for decades [**?**]. FBP-based algorithms provide adequate image quality in a robust and computational fast manner. However, due to shortcomings like high image noise level, artifacts and poor low-contrast, another important class of algorithms, the iterative reconstruction (IR) methods, have gained popularity. While improving image quality, the iterative methods typically require computations which have become compatible with the daily clinical practice only thanks to recent progress in hardware, namely GPU acceleration. Although the first iterative methods had been proposed already in late 1960s and early 1970s [**?**], [**?**], they have been introduced to diagnostic machines only less than 15 years ago [**?**], [**?**], [**?**], [**?**]. In standard CBCT applications such image-guided radiation therapy (IGRT), the analytical *Feldcamp-Davis-Kress* (FDK) algorithm [**?**] is still employed in many treatment protocols, however, iterative techniques including the statistical image reconstruction (SIR) methods are nowadays an indisputable part of clinical routine [**?**], [**?**].

Unfortunately, the iterative methods bring their specific shortcomings: While the SIR is highly efficient in improving the quality in the case of photon starvation, it is very sensitive to inconsistencies that stem from motion occurring during the acquisition. Therefore, despite their advantage, SIR is typically employed in protocols where sufficient immobilization can be ensured such as head-and-neck and pelvis scans. Abdominal anatomies, highly prone to voluntary or involuntary motion, are consequently reconstructed with analytical or non-statistical methods.

In this paper, we propose a modified iterative reconstruction method where the algebraic reconstruction technique (ART) is followed by a statistical reconstruction based on penalized maximum likelihood (PL). Unlike the conventional PL, our method employs modified gradients which are down-weighted in the areas capturing the signal that is transmitted exclusively through the air. We demonstrate that the proposed method is capable of reducing significantly the motion-related artifacts in abdominal images while maintaining the benefits of the statistical approach, and thus improves the image quality when compared to both FDK and the original iterative method.

I. Peterlik, A. Strzelecki, M. Lehmann, P. Messmer, P. Munro, P. Paysan, M. Plamondon and D. Seghers are with Varian Medical Systems Imaging Laboratory, Taefernstrasse 7, 5405 Daettwil, Switzerland (e-mail: Igor.Peterlik@varian.com, Adam.Strzelecki@varian.com)

II. Method

A. Iterative CBCT reconstruction as optimization problem

In the iterative 3D CBCT reconstruction, the image volume $\mathbf{f}^*$ is computed through an optimization of cost function typically composed of terms corresponding to data fidelity and regularization. The data fidelity penalizes the discrepancy between the forward projections $\mathbf{Af}$ of the actual estimation of $\mathbf{f}$, while the regularization introduces an additional prior knowledge about the reconstructed image. In the case of ART regularized with well-known total variation (TV) regularizer [**?**], the cost function is given as

$$\mathbf{f}^*_{\text{ART}} = \arg\min_{\mathbf{f}} \sum_{\alpha} \|\mathbf{p}_\alpha - [\mathbf{Af}]_\alpha\|_2^2 + \lambda_{\text{ART}} \|\mathbf{f}\|_{TV} \qquad (1)$$

where $\mathbf{p}_\alpha$ is the log-normalized pre-processed projection under angle α in the attenuation space (e.g., with scatter and beam-hardening corrections applied), $[\mathbf{Af}]_\alpha$ represents the forward projection of the actual volume $\mathbf{f}$ under the projection angle α, and λ is the scalar weight of the TV regularization term.

Unlike ART, PL performs the optimization in the transmission space, nonetheless, the regularizer is again employed to

improve the convexity of the solution space. In the case of CBCT, the statistical model behind PL is based on the Poisson statistics: $\mathbf{q}_\alpha \propto \mathbf{r}_0 e^{-[\mathbf{Af}]_\alpha}$ where $\mathbf{q}_\alpha$ is the projection under angle α in the intensity space, and $\mathbf{r}_0$ is the corresponding air norm, i.e. acquisition without any object in the scanner. As shown in [**?**], the PL cost function can be derived using the maximum log-likelihood, resulting in:

$$\mathbf{f}^*_{\text{PL}} = \arg\min_{\mathbf{f}} \sum_\alpha \left\{ \mathbf{q}_\alpha \cdot [\mathbf{Af}]_\alpha + \mathbf{r}_0 e^{-[\mathbf{Af}]_\alpha} \right\} + \lambda_{\text{PL}} \|\mathbf{f}\|_{TV} \tag{2}$$

where the product $\mathbf{q}_\alpha \cdot [\mathbf{Af}]_\alpha$ is computed pixel-wise.

In the scenario considered in this paper, the iterative reconstruction employs both methods: First, ART is performed with a more aggressive regularization (i.e., higher value of λ_{ART}) in order to obtain a first-pass reconstructed volume $\mathbf{f}^*_{\text{ART}}$. This is then used as the initial estimation for the second-pass PL, the goal of which is to compute $\mathbf{f}^*_{\text{PL}}$ refining the smooth anatomies with higher-frequency information. To this end, the regularization parameter λ_{PL} is significantly smaller than λ_{ART}. Both optimizations require the evaluation of gradients of both ART and PL cost functions. In the following, we focus only on gradients of the data fidelity term.

Given the cost function penalizing the discrepancy between the re-projected volume and each projection in the attenuation space, the data-fidelity gradient corresponding to (re-)projection under angle α is given as:

$$\nabla^{\text{DF-ART}}_\alpha = \mathbf{A}^\intercal([\mathbf{Af}]_\alpha - \mathbf{p}_\alpha) \tag{3}$$

where $\mathbf{A}^\intercal$ represents the back-projection of the difference between the re-projected and input projection onto the reconstructed volume. As shown in [**?**], the derivation of PL data-fidelity gradients leads to following expression:

$$\nabla^{\text{DF-PL}}_\alpha = \mathbf{A}^\intercal(\mathbf{q}_\alpha - \mathbf{r}_0 e^{-[\mathbf{Af}]_\alpha}) \tag{4}$$

where $\mathbf{A}^\intercal$ is again the back-projection operator which transforms the difference between the re-projected volume and each projection in the transmission space to the volume space.

B. Accelerated reconstruction and motion-related artifacts

Beside the quality of the reconstructed image, the time necessary to perform the reconstruction is a key factor determining the clinical applicability of the method. To reduce the complexity of the optimization, the non-linear objective function is replaced with a *separable quadratic surrogate* (QS) [**?**]. Since the surrogate is not well-defined for negative values, it requires the application of non-negativity constraints (NNC): In the actual reconstruction, these are applied via simple projection of negative values to zero. To further accelerate the reconstruction, *ordered subsets* (OS) [**?**] and Nesterov's momentum method [**?**] are employed. While the former decreases the cost of one iteration which is performed using only one subset instead of the entire data set, the latter reduces the number of steps necessary to achieve the convergence.

During an evaluation of the iterative reconstruction of abdominal scans on more than 50 clinical data sets acquired during radiation therapy, it was observed that inconsistencies following from motion of the abdominal wall of a patient during the CBCT acquisition result in significant streak artifacts in the reconstructed images. Further investigation revealed that the motion-related streak artifacts gradually appear during the PL refinement while the first-pass ART reconstruction does not suffer from this issue. Next, iterative reconstructions without the acceleration techniques (i.e. ordered subsets and momentum) were performed, however, neither of these was found responsible for the artifacts. Finally, the reconstructions were executed without applying NNC in PL. In this case, the artifacts were significantly reduced, however, the PL optimization became unstable and often resulted in divergence.

In the case of half-fan beam geometry, the inconsistency between the projections acquired under similar or same angles at the and the end of the arc is especially strong. Without loss of generality, let us assume that the abdominal wall moves outwards during the acquisition due to involuntary relaxation of abdominal muscles. In this case, the same projected area at the beginning of the acquisition is perceived as empty (i.e., filled with the air), while it is occupied by the tissue at the end of the acquisition. In the case when NNC are disabled, this critical region in the reconstructed image is compensated by an accumulation of negative attenuation values in the adjacent areas. If these values are constantly re-projected to zero, PL compensates for the tissue-air inconsistency by artificially modifying the attenuation values in the direction of the inconsistency which results in observed artifacts.

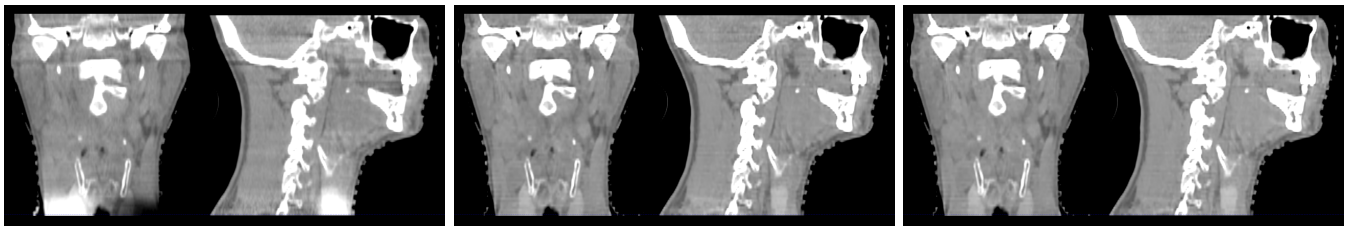
C. Gradient Weighting in Statistical Reconstruction

Based on these findings, we modified the update of reconstructed volumes during the second-pass PL: Instead of computing the gradients of data-fidelity globally for all pixels on each projection, we down-weight the values of the gradient corresponding to pixels receiving signal which completely avoids the tissue, thus is transmitted exclusively through the air. In general, the artifacts related to the low photon statistics do not occur close to the body surface. Thus, we still retain the benefits of the statistical approach (e.g. suppressing the photon starvation deeper inside body), addressing the shortcomings of PL which occur close to the body surface.

We first compute the weighing projections $\mathbf{w}_\alpha$ using the input transmission data and second, we apply these weights in each sub-iteration of PL. To perform the first step, each transmission projection $\mathbf{q}$ is normalized through the pixel-wise division by the air-norm $\mathbf{r}_0$. Threshold τ is applied to flatten the area capturing the signal transmitted only through the air and final mask is inverted and re-normalized with τ to contain only values between 0 and 1:

$$\mathbf{w}_\alpha = \frac{\tau - \lceil \mathbf{q}_\alpha / \mathbf{r}_0 \rceil^\tau}{\tau} \tag{5}$$

The resulting mask is strictly zero in the area where the pixels capture the air-only signal. As soon as the signals are attenuated by a tissue, the photon counts sharply decrease when compared to the air-only areas. Therefore, the values of mask pixels in these areas quickly approach 1 which means that the PL gradients in this pixels remain unmodified w.r.t. the conventional PL.



FDK iCBCT iCBCT-GW

Fig. 1: The artifacts in FDK volume completely suppressed by both iCBCT and iCBCT-GW methods. Head-and-neck scan acquired on TrueBeam: Half-fan beam geometry, 100 kV, W/L = 800/0 HU. Courtesy of Henry Ford Health System.

Given the weighting mask $\mathbf{w}_\alpha$ corresponding to each input projection, the gradient of PL data fidelity term in Eq. 4 is modified so that before the application of the back-projection operator $\mathbf{A}^\intercal$, the gradient in the projection space is weighted as follows:

$$\nabla_\alpha^{\text{DF-WPL}} = \mathbf{A}^\intercal \left\{ \mathbf{w}_\alpha \cdot \left(\mathbf{q}_\alpha - \mathbf{r}_0 e^{-[\mathbf{A}\mathbf{f}]_\alpha} \right) \right\} \tag{6}$$

where the multiplication between the mask and the difference of the right-hand side is performed pixel-wise.

III. Results

The modified iCBCT with gradient weighting has been tested on about 20 real patient data sets acquired under breath hold (when applicable) with Varian Halcyon™ and Varian TrueBeam™ treatment devices. We have selected four data sets which where reconstructed with analytical FDK, iterative CBCT reconstruction described in section II-A both without (abbr. as *iCBCT*) and with (*iCBCT-GW*) PL gradient weighting. In the latter case, a threshold $\tau = 0.9$ was applied. The typical resolution of the reconstructed volume was $512 \times 512 \times 100$ voxels.

Fig. 1 confirms that the modification does not compromise the benefits of IR methods: The artifacts clearly visible in FDK image both in the neck and the intracranial areas do not appear in neither iCBCT nor iCBCW-GW images. Fig. 2 demonstrates the advantage of iCBCW-GW over both iCBCW and FDK: Both the shading artifacts (visible in the FDK image) and tissue-air inconsistency artifact (visible in iCBCT) are significantly reduced. The following scan (Fig. 3) shows an extreme scenario where almost the entire abdominal wall was displaced by more then 30 mm due to the muscle relaxation that occurred during the acquisition. In this case, the gradient weighting significantly reduces the global impact of the artifact which in the case of iCBCT compromises the image quality as far as around the isocenter. Although the artifact is not completely suppressed by iCBCT-GW, it remains locally isolated close to the body edge. The reconstructions of a pelvis scan depicted in Fig. 4 confirm that in the case without motion-related artifacts, gradient weighting maintains the quality of image when compared to FDK which suffers from low contrast in the region around the isocenter. We executed repetitive time measurements to determine the increase of computational cost due to the modification. In all reconstructions, the impact of the gradient weighing was negligible, adding about 2% of additional time.

IV. Conclusions

We have presented a modified iterative method for CBCT image reconstruction. Combining ART and PL techniques, the proposed algorithm improves the image quality in data suffering from motion during acquisitions, while it maintains the benefits of iterative methods The method have been evaluated on more than 20 data sets, out of which four illustrative examples are presented. Since the gradient weighting adds only a negligible computational cost when compared to the original method already employed clinically, the proposed algorithm is suitable for integration into treatment protocols.

References

[1] X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" *Inverse Problems*, vol. 25, no. 12, p. 123009, 2009.

[2] G. N. Hounsfield, "A method of and apparatus for examination of a body by radiation such X or gamma radiation," uK 1283915.

[3] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *Journal of Theoretical Biology*, vol. 36, no. 1, pp. 105 – 117, 1972.

[4] K. Grant and R. Raupach, *SAFIRE: Sinogram affirmed iterative reconstruction*, Siemens Healthcare, Erlangen, Germany, 2012.

[5] J.-B. Thibault, *Veo: CT model-based iterative reconstruction*, GE Healthcare, Wiscons, USA, 2010.

[6] D. Mehta, R. Thompson *et al.*, "Iterative model reconstruction: Simultaneously lowered computed tomography dose and improved image quality," *Medical Physics International Journal*, vol. 1, no. 2, pp. 147 – 155, 2013.

[7] R. Irwan, S. Nakanishi, and A. Blum, *AIDR 3D–Reduces dose and simultaneously improves image quality*, Toshiba Medical Systems, Otawara, Japan, 2011.

[8] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, Jun 1984.

[9] S. J. Gardner, W. Mao *et al.*, "Improvements in CBCT image quality using a novel iterative reconstruction algorithm: a clinical evaluation," *Advances in radiation oncology*, vol. 4, no. 2, pp. 390–400, 2019.

[10] B. Cai, E. Laugeman *et al.*, "Characterization of a prototype rapid kilovoltage x-ray image guidance system designed for a ring shape radiation therapy unit," *Medical physics*, vol. 46, no. 3, pp. 1355–1370, 2019.

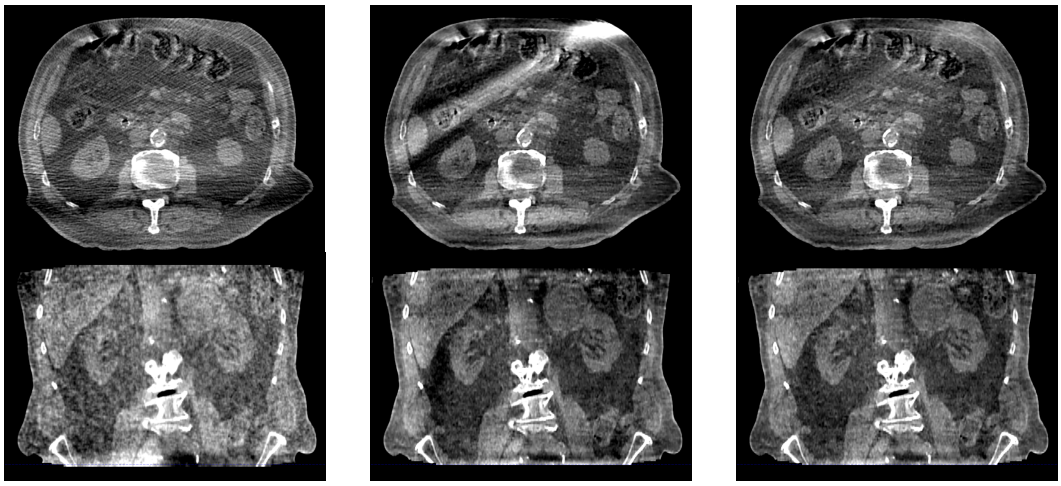
[11] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[12] T. Buzug, *Computed Tomography*. Springer Verlag, 2008.

[13] H. Erdogan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Transactions on Medical Imaging*, vol. 18, no. 9, pp. 801–814, Sept 1999.

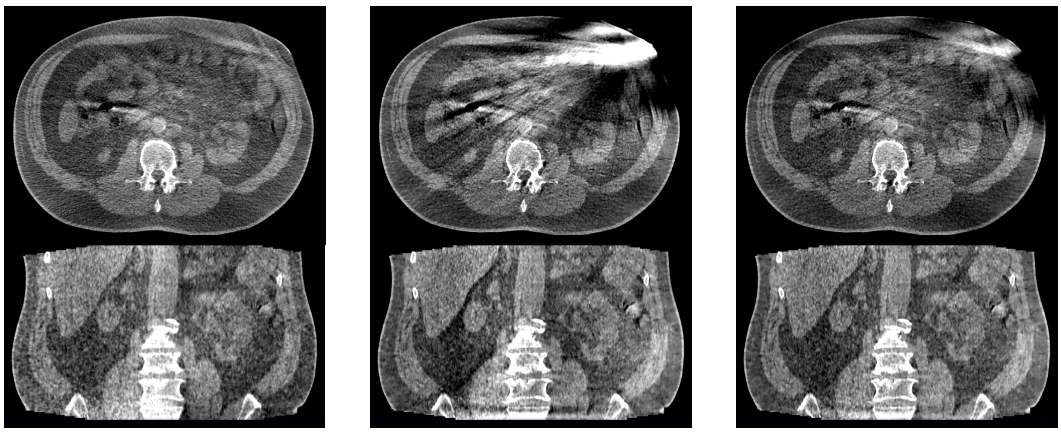
[14] ——, "Ordered subsets algorithms for transmission tomography," *Physics in Medicine & Biology*, vol. 44, no. 11, p. 2835, 1999.

[15] A. S. Wang, J. W. Stayman *et al.*, "Accelerated statistical reconstruction for C-arm cone-beam CT using Nesterov's method," *Medical Physics*, vol. 42, no. 5, pp. 2699 – 2708, 2015.



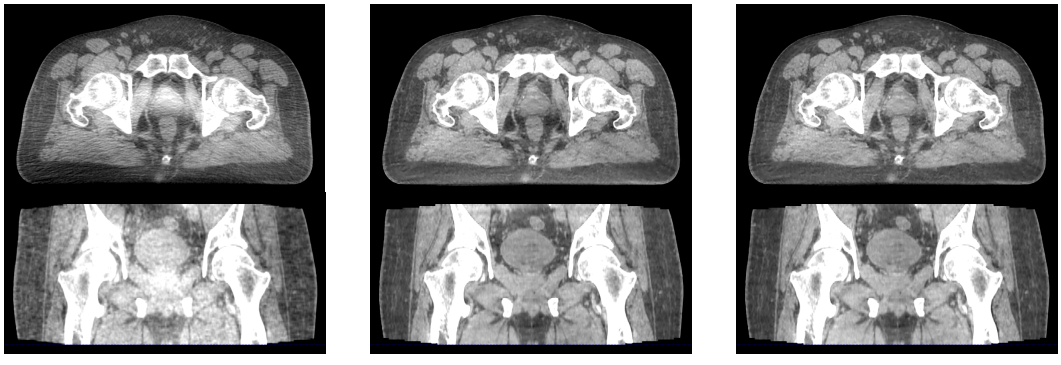
FDK iCBCT iCBCT-GW

Fig. 2: Improved quality of image produced by iCBCT-GW compared to both FDK and CBCT. Abdominal scan acquired on Halcyon: Half-fan geometry, 140 kV, W/L = 400/75 HU. Courtesy of Washington University (WU) in St. Louis.



FDK iCBCT iCBCT-GW

Fig. 3: Improved quality of image produced by iCBCT-GW compared to both FDK and CBCT. Abdominal scan with important air-tissue inconsistency acquired on Halcyon: half-fan geometry, 140 kV, W/L = 400/75 HU. Courtesy of WU in St. Louis.



FDK iCBCT iCBCT-GW

Fig. 4: Iterative reconstructions improving the quality in the region around the isocenter. iCBCT-GW further slightly improves the image close to the upper edge (coronal view). Pelvis scan acquired on TrueBeam: half-fan geometry, 125 kV, W/L = 400/75 HU. Courtesy of Henry Ford Health System.

Motion correction for 4D CBCT reconstruction with TV regularization

Rob Heylen, Johan Nuyts

***Abstract*—Recently, we proposed a 4D CBCT reconstruction algorithm with spatial and temporal total variation regularization, and employed this algorithm for dual-energy dynamic angiography of the brain on a software phantom. This approach is very sensitive to movement, as misalignment between time frames will interfere with the temporal regularization. In this work, we integrate an iterative motion correction algorithm with the 4D CBCT reconstruction algorithm. Several strategies can be followed, and we assess which strategy is most successful. Experiments are performed on a dynamical software phantom.**

I. Introduction

With 4D cone-beam CT (CBCT) one can reconstruct multiple 3D time frames from a single CBCT acquisition. As the angular interval for each time frame is relatively small, prior information is typically employed to reduce artefacts and improve image quality. Many authors have shown that total variation (TV) regularization in both the time and spatial domain is very useful for 4D CBCT reconstruction, with applications in cardiac, pulmonary and brain imaging [1], [2]. Computationally efficient implementations based on the primal-dual optimization algorithm of Chambolle and Pock [3] have recently been presented as well [4], [5], [6].

Furthermore, dual-energy (DE) CBCT systems are being deployed more frequently, which increases the diagnostic capabilities of CBCT examinations. A popular application is material decomposition, where the dual-energy image is transformed into two material images. This decomposition can be performed in image space, but also in data space when the DE views are matched. An interesting decomposition basis for dynamic angiography of the brain would be a water and iodine component, where the iodine image visualizes the injected contrast. Applying 4D CBCT reconstruction to such an iodine image can generate several time frames of the contrast flow, corresponding to different phases of the bolus traversal, and can provide time-resolved dynamic collateral imaging with a single CBCT acquisition.

A TV prior penalizes differences between neighbouring voxels, either in the spatial or temporal direction. Such a prior is sensitive to movement, as voxels should stay registered between different temporal frames. As the delay between these time frames is in the order of several seconds, one cannot assume that these frames stay registered with sub-voxel accuracy in practical use, and a registration method or motion correction needs to be applied.

In this work, we investigate the influence of significant movement on 4DTV CBCT reconstruction methods, and the ability of a motion correction method to alleviate the resulting issues. Experiments are performed on a dynamical software phantom, so that ground truth is always available.

Manuscript received January 17, 2020.

R.H. and J.N. are with the Department of Imaging and Pathology, Division of Nuclear Medicine, KULeuven, Belgium. This work was done under the NEXIS project, that has received funding from the European Union's Horizon 2020 Research and Innovations Programme (Grant Agreement #780026).

II. 4D CBCT reconstruction

The algorithm for 4D CBCT reconstruction combines several data fidelity and prior information terms into a single optimization equation, which is solved via the Chambolle-Pock primal-dual optimization algorithm [3]. The algorithm is very similar to the ROOSTER algorithm proposed in [4], but has slight differences in implementation and interpretation.

We consider a single circular DE CBCT scan, typically over 360 degrees and with an acquisition time of around 25s. We focus on reconstructing a 4D reconstruction of an iodine sinogram, obtained by material decomposition of a cranial DE CBCT scan. We assume that the image intensities change along linear curves between different time frames, and that these changes are due to changes of the iodine contrast concentrations. See also [7] for a more detailed explanation of the algorithm.

We consider several terms in the optimization equation:

- A data fidelity term that penalizes differences between the forward projection of the reconstruction and the iodine sinogram. The penalty is implemented as the L2 norm of the difference. The forward and backward projectors employ linear interpolation between the different time frames.
- A binary mask indicating static areas, such as the skull. The masked voxels are averaged over all time frames. This significantly improves performance, especially when highly attenuating static areas are masked. A heuristic method to obtain a skull mask is to threshold a static water image, obtained by FDK reconstruction of the water sinogram.
- Non-negativity, as the amount of iodine in a voxel cannot be negative.
- TV regularization, both in the spatial direction (3D) and the temporal direction (1D). Both regularizers can be weighted differently, and we choose the temporal TV regularizer 10 to 100 times stronger than the spatial regularizer.

This leads to the optimization equation

$$\underset{x\in X}{\operatorname{argmin}} \left(\|Ax - s\|_2^2 + \alpha \|x\|_{\text{TV4}} + \mathrm{I}_{\mathbb{R}^{\mathbb{N}+}}(x) \right) \tag{1}$$

with

$$\|x\|_{\text{TV4}} = \sum_{ijkt} |(\Delta x)_{ijkt}| \tag{2}$$

$$|(\Delta x)_{ijkt}| = \left(\sum_{d=1}^{3} (\Delta_d x)_{ijkt}^2 + (\gamma \Delta_4 x)_{ijkt}^2 \right)^{1/2} \tag{3}$$

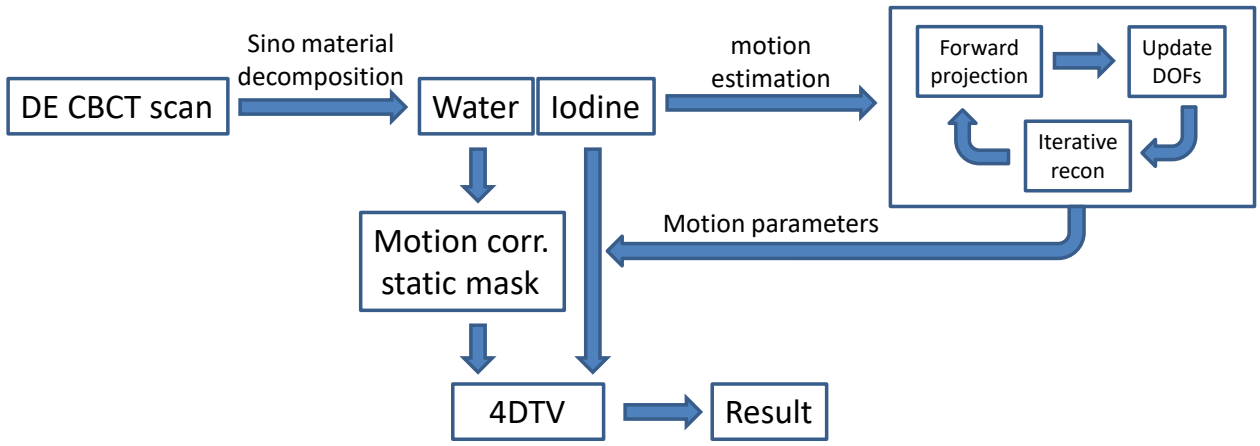


Fig. 1: Diagram representing the motion-corrected DE CBCT data processing chain for dynamic collateral imaging.

and $\mathrm{I}_{\mathbb{R}^{N+}}(.)$ the characteristic function of the positive orthant. The γ factor in the last term weighs the spatial and temporal directions differently. This equation can be written in terms of the Chambolle and Pock framework, yielding algorithm 1 to solve (1). See [3], [4], [6], [7] for more information.

Algorithm 1: The 4DTV algorithm

Initialize all variables, choose τ, $\sigma > 0$, $\theta \in [0, 1]$;
for $n \geq 0$ **do**
  $y_1^{n+1} = \frac{2}{2+\sigma}\left(y_1^n + \sigma(A\overline{x} - s)\right)$;
  $x_1^{n+1} = A^* y_1^{n+1}$;
  $y_2^{n+1} = \mathrm{proj}_{\alpha P}\left(y_2^n + \sigma\Delta\overline{x}\right)$;
  $x_2^{n+1} = -\mathrm{div}\left(y_2^{n+1}\right)$;
  $x^{n+1} = \mathrm{proj}_{\mathbb{R}^{N+}}\left(x^n - \tau x_1^{n+1} - \tau x_2^{n+1}\right)$;
  $\overline{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n)$;

III. Rigid motion correction and 4DTV

We assume that rigid motion might be present during the CBCT acquisition. This rigid motion can be described by 6 degrees of freedom (DOF), corresponding to the Euler angles and a 3D translation, for each view/angle in the sinogram. We propose a data-driven method for estimating this motion, adapted from the recently proposed method for rigid motion correction in helical CT [8], [9].

The rigid motion correction method is an iterative procedure, consisting of the following steps. See [8], [9] for more details:

1) Determine an initial static reconstruction, e.g. with the FDK algorithm. Set $k = 1$.
2) Downscale/Upscale the sinogram and the current reconstruction by a factor D_k. Set $it = 1$.
3) For each DOF:
   a) Calculate the numerical derivative of the L2 reconstruction error wrt. this DOF with finite differences.
   b) Update this DOF for each view/angle by minimizing the L2 error. We employ a linear extrapolation method.
4) Use an iterative reconstruction method to improve the current reconstruction employing the updated DOF estimates.
5) Increase it. If $it \geq N_k$ increase k and go to 2, else go to 3.

The employed iterative reconstruction method is the proposed 4DTV algorithm, with motion-aware projectors and backprojectors. Note that a static version of this algorithm can be easily obtained by choosing the number of time frames equal to one, and that the TV terms can be deactivated as well, leading to a classical iterative reconstruction algorithm.

Calculation of the numerical derivative in step 3(a) only requires an additional forward projection for 3 of the 6 DOFs: Translations parallel to the sensor plane and axial rotations do not require an additional forward projection, as these can be easily calculated as 2D geometrical operations in the sensor plane. We also noted experimentally that repeating step 3 a couple of times before moving to step 4 improves the convergence significantly, as each updated DOF influences the others in turn. Most of the computational time is spent in the iterative reconstruction in step 4.

By employing multiple scales, as implemented in step 2 of the algorithm, the convergence can be sped up in a major way: significant motion can be easily estimated on downsampled data sets, and the computational resources required decrease fast. We typically employed $D = [4, 2, 1]$ downsampling in our experiments, with $N = [5, 3, 1]$ iterations for each scale. The iterative reconstruction algorithm was run for 10 iterations each time.

We tested two strategies for combining the 4DTV reconstruction algorithm with the motion estimation: simultaneous estimation of rigid motion and multiple time frames by using the full implementation of 4DTV in step 4 of the algorithm, or cascade the two algorithms, where first the rigid motion is estimated (assuming a static reconstruction), followed by 4DTV reconstruction using the estimated motion and the static reconstruction as initialization. We found that the second strategy produces better results than the first, and has a lower computational cost.

A schematic representation of the entire data processing chain is given in Fig. 1

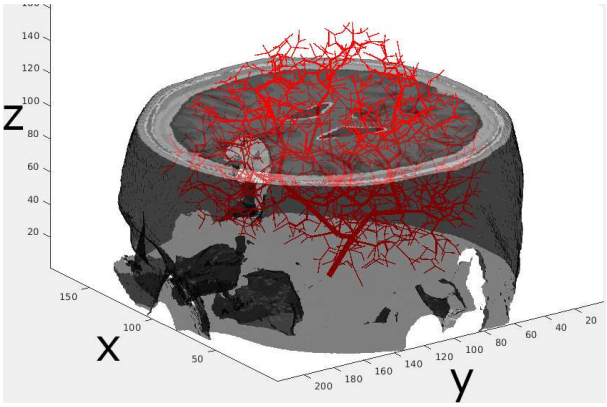


Fig. 2: The simulated vasculature inside the brainweb phantom.

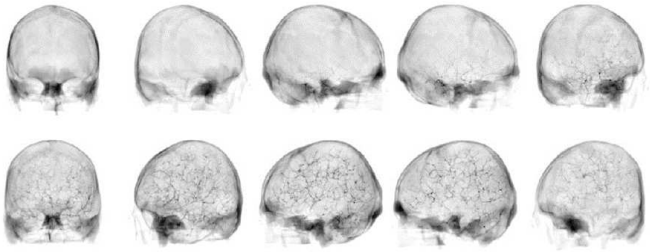


Fig. 3: 10 views of the iodine sinogram, evenly spread over 360 degrees, showing the evolution of the contrast through the vascular tree.

IV. Experiments

A. The dynamical phantom

We created a dynamic phantom for testing the capabilities of the presented algorithm on simulated DE CBCT scans (see also [7] for more information). This phantom is based on the brainweb phantom [10], where we added an additional dynamical iodine class consisting of a simulated arterial tree [11] inside the grey/white matter with contrast dynamics. The dynamics was obtained by considering the binary tree as a network of interconnected tubes, and calculating the hydrodynamics of such a system under laminar flow [12]. An input function was obtained from a voxel in the carotid area of a perfusion CT examination. The diffusion coefficient, the single free parameter in the dynamics, was modified such that the time for the bolus to reach the terminal points was comparable to the time delay between the arterial and venous phase in the perfusion data set.

This data set was forward projected with a DE CBCT simulator, and decomposed into water and iodine components via Newton optimization of the spectral detector model. A visualization of the arterial tree is shown in Fig. 2, and several views of the iodine sinogram in Fig. 3, showing the dynamics as the contrast bolus passes through the network. The total scan time was 25 seconds, and we simulated a full 360 rotation with 200 views of 300×300 pixels of size $(1\text{mm})^2$, with a focus height of 1000mm and a detector distance of 150mm. The reconstructed volume corresponds with the size of the brainweb phantom, and has size $(181, 217, 181)$, with $(1\text{mm})^3$ voxels.

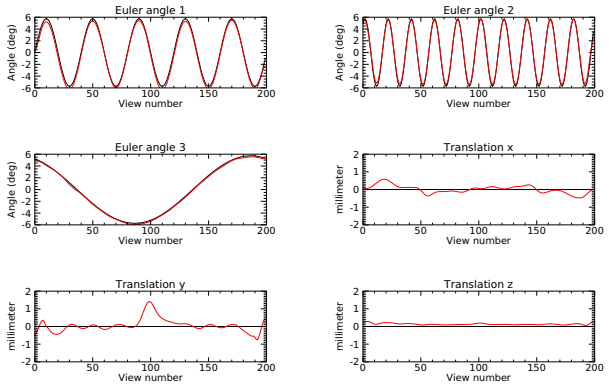


Fig. 4: The true movement (black) and the calculated movement by the rigid motion correction (red), for all 6 DOFs.

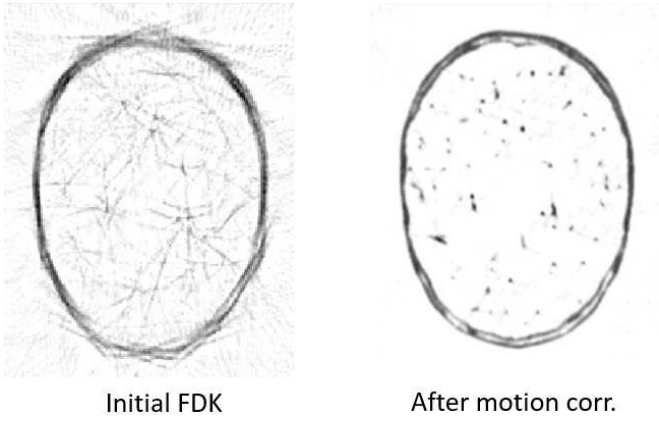


Fig. 5: Central slice of static reconstructions before (left) and after (right) motion correction.

B. Rigid motion and estimation

To add rigid motion to the data set, we modified each DOF as a function of view/angle as the CBCT acquisition takes place. These movement functions are displayed in Fig. 4, along with the estimated movement after rigid motion correction. We changed only rotational DOFs, keeping the translational DOFs equal to zero. Note that although these motion functions are artificial, they represent significant motion in all rotational directions.

The rigid motion estimation procedure succeeds in estimating most DOFs relatively well, with only some translational deviations for some views. Note that there are several invariances allowed in the estimated DOFs, as the estimation procedure might result in a translated or rotated coordinate frame which is also valid for reconstruction. This hampers quantitative assessment strategies based on voxel-wise comparison.

In Fig. 5, the input and output images of the motion correction procedure are shown. The input is an FDK reconstruction of the iodine sinogram including movement. It is clear that the motion corrected image is of much better quality.

C. 4DTV reconstruction

After motion correction, the estimated motion parameters are used in the 4DTV algorithm for reconstructing multiple time frames. Also a static motion-corrected version of the

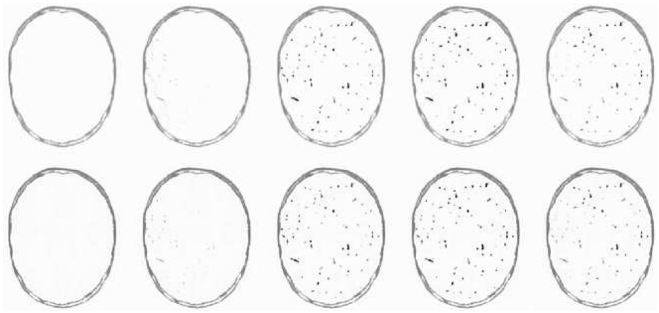


Fig. 6: Central slice of 5 dynamically reconstructed frames after motion correction (top row), and the corresponding ground truth (bottom row).

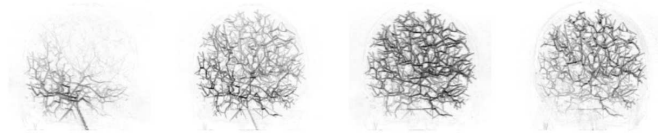


Fig. 7: MIPs of subtraction images with the first frame.

water image has to be reconstructed to determine the static skull mask.

We reconstructed 5 time frames with 100 iterations of the 4DTV algorithm. The central slice of these reconstructions is shown in Fig. 6, along with the ground truth obtained from the dynamical software phantom.

Because the first reconstructed time frame does not yet show any contrast dynamics, we can employ this frame to create digital subtraction images. Maximum intensity projections (MIP) of these subtraction images are shown in Fig. 7.

To stress the influence of the motion correction step in the reconstruction procedure, we have also recreated all results with the same 4DTV parameters, but without added motion or motion correction. The second MIP from Fig. 7 is repeated in Fig. 8, along with the version without motion and the MIP obtained from the ground truth. While the motion-corrected data set has more blur, it is clear that these results are qualitatively comparable.

V. Conclusions

We have shown that motion correction can be combined with 4DTV, significantly improving reconstruction results, and that using motion correction and 4D reconstruction successively is superior to integrating 4DTV and iterative motion correction into a single algorithm. We performed experiments on an artificial dynamical data set, obtained with a software phantom including contrast dynamics, and show that the effects of motion can be alleviated almost completely. In future work, we will apply these techniques to realistic Monte Carlo simulations and real data sets, aiming to obtain dynamic collateral imaging with DE CBCT devices.

Fig. 8: The motion-corrected 4DTV result of the data set including motion (left), the 4DTV result on the original data set without motion (middle), and the ground truth (right).

References

[1] D.C.Hansen and T. Sorensen, "Fast 4d cone-beam ct from 60 s acquisitions," *Phys. Im. in Rad. Oncology*, vol. 5, pp. 69–75, 2018.

[2] L. Ritschl, S. Sawall, M. Knaup, A. Hess, and M. Kachelriess, "Iterative 4d cardiac micro-ct image reconstruction using an adaptive spatio-temporal sparsity prior," *Phys. Med. Biol.*, vol. 57, pp. 1517–1525, 2012.

[3] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, May 2011.

[4] C. Mory and L. Jacques, "A modified 4d rooster method using the chambolle–pock algorithm," *Proc. 3rd Intl. Conf. on Image Formation in X-Ray CT*, pp. 191–193, 2014.

[5] O. Taubmann, V. Haase, G. Lauritsch, Y. Zheng, G. Krings, J. Hornegger, and A. Maier, "Assessing cardiac function from totalvariation-regularized 4d c-arm ct in the presence of angular undersampling," *Phys. Med. Biol.*, vol. 62, p. 2762–2777, 2017.

[6] V. Nikitin, M. Carlsson, F. Andersson, and R. Mokso, "Four-dimensional tomographic reconstruction by time domain decomposition," *IEEE Tran Comp. Im.*, pp. 1–1, Feb 2019.

[7] R. Heylen, G. Schramm, P. Suetens, and J. Nuyts, "4d cbct reconstruction with tv regularization on a dynamic software phantom," *IEEE Nuclear science symposium and medical imaging conference*, pp. 1–3, 2019.

[8] T. Sun, J.-H. Kim, R. Fulton, and J. Nuyts, "An iterative projection-based motion estimation and compensation scheme for head x-ray ct," *Medical Physics*, vol. 43, no. 10, pp. 5705–5716, 2016.

[9] J.-H. Kim, T. Sun, A. Alcheikh, Z. Kuncic, J. Nuyts, and R. Fulton, "Correction for human head motion in helical x-ray ct," *Phys Med Biol*, vol. 61, no. 4, pp. 1416–1438, 2016.

[10] C. Cocosco, V. Kollokian, R.-S. Kwan, and A. Evans, "Brainweb: Online interface to a 3d mri simulated brain database," *NeuroImage*, vol. 5, no. 4, pp. part 2/4, S425, 1997.

[11] R. Karch, F. Neumann, M. Neumann, and W. Schreiner, "A three-dimensional model for arterial tree representation, generated by constrained constructive optimization," *Computers in Biology and Medicine*, vol. 29, no. 1, pp. 19 – 38, 1999.

[12] G. I. Taylor, "Dispersion of soluble matter in solvent flowing slowly through a tube," *Proc. Royal Soc. of London. Ser. A. Math. Phys. Sci.*, vol. 219, no. 1137, pp. 186–203, 1953.

Iterative FDK reconstruction for helical CT of the head with rigid motion compensation

Johan Nuyts, Roger Fulton

***Abstract*—In previous work, we have developed a reconstruction algorithm for helical CT which compensates for a known rigid motion of the object being scanned [1]. This algorithm was used for CT imaging of the head, as part of a method which estimates possible rigid head motion on a view by view basis from the CT data themselves, and compensates for that motion during reconstruction of the clinical image. In this earlier work, we have used iterative maximum likelihood algorithms for the reconstruction. Compensating for motion in such algorithms is easy and effective, but the computation time is high. Here we propose an iterative FDK algorithm as an alternative. To compensate for rigid motion, we account for the motion during the backprojection step, while the filtering step is not modified. The resulting image suffers from low frequency artefacts. However, it is found that these artefacts are eliminated by applying a few iterations of the FDK reconstruction. The performance of the algorithm is demonstrated on a clinical CT scan which was affected by significant patient motion.**

I. Introduction

Compensating for rigid object motion in iterative reconstruction is fairly straightforward: the inverse motion can be assigned to the CT system by superimposing it on the system's helical trajectory. This results in an irregular trajectory of the source and detector in a coordinate system where the object is stationary. Iterative maximum likelihood (ML) algorithms use the original views, which is convenient because it is assumed that the pose of the object may be different for each view. In addition, they make use of all the measured data, which is beneficial for the signal to noise ratio of the reconstructed image. Very promising results have been obtained with this approach, and we are currently evaluating it for clinical application [1]. However, the long computation times of ML reconstruction hamper its introduction in the clinic. In this paper, we report on our attempt to use the FDK algorithm [2] as an alternative for ML reconstruction in this context. Advantages of the FDK algorithm are that it is fast, it uses the acquired views without the need for extensive interpolations (like e.g. fan- to parallel rebinning) and it can use (almost) all of the acquired data. Other choices are possible. Ye et al. proposed an analytical algorithm for cone beam reconstruction with an arbitrary trajectory, where the only requirement is that every point is located on a PI-line [3]. Jang et al. developed a motion correction approach with similarities to that of [4], where they replaced ML reconstruction with a faster analytical reconstruction. That algorithm involves fan- to parallel rebinning, and they managed to account for the motion in this rebinning step.

We are evaluating our methods on different Siemens CT systems, the most recent one being the Siemens Force. On this system, the head CT scans are typically acquired with a low pitch (0.55), one rotation per second, 4200 views per rotation and flying focal spot (i.e. increasing the transaxial and axial sampling by small motions of the focal spot). In addition, the Force detector has non-uniform angular sampling, because the detectors are combined in modules, and there are small gaps between these modules. In the next sections, the preprocessing of the data to account for these sampling issues is briefly explained. Then application of the FDK algorithm for reconstructing scans without motion correction and its extension for motion correction are discussed. Finally, results obtained on a motion corrupted clinical CT scan are shown

J. Nuyts is with the Department of Imaging and Pathology, Nuclear Medicine & Molecular Imaging and with the Medical Imaging Research Center, KU Leuven, B-3000, Leuven Belgium. R. Fulton is with the Faculty of Health Sciences, University of Sydney, and the Department of Medical Physics, Westmead Hospital, Australia.

II. Methods

A. FDK for helical CT reconstruction unaffected by motion

1) Interpolation of adjacent views: To exploit the sampling improvement provided by the transaxial flying focal spot, pairs of adjacent views are combined into a higher resolution view, creating a new sinogram with twice the number of detector pixels, half the number of angles and an unchanged number of detector rows. The focal spot motion is designed to create a uniform sampling pattern of double density near the center of the field of view in the transaxial plane. Because of the fan angle geometry, this ideal interference is only obtained in the center, the interference deviates more from the ideal double sample at locations which are at larger distance from the center. This is illustrated by the red and blue lines in figure 1, which represent the two projection lines associated with two adjacent views. In that same figure, the gray lines represent the geometry of a virtual view with double resolution, which is to be computed (approximately) from the two available views. This virtual view is computed by interpolating the detector values of the two available views, according to the 1D coordinates on a line through the center and perpendicular to the central virtual projection (the dashed black line in fig. 1). This 1D coordinate assigned to each detector value is the position of the intersection point of the projection line associated with that detector and the central line. In other words, we assume that the available detector values correspond to the intersection of the red and blue lines with the black line in figure 1, and the target virtual detector values correspond to the intersection of the gray lines with the black line. For head

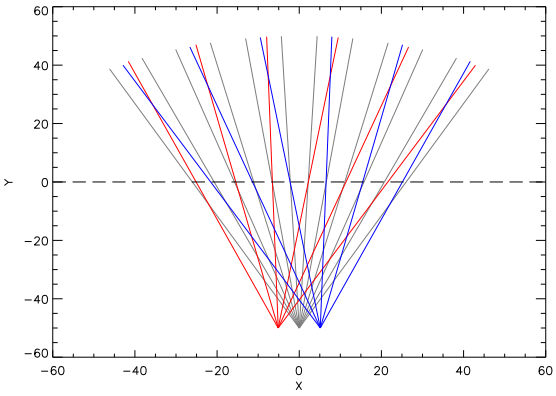


Fig. 1. The red and blue lines represent the geometry of a pair of adjacent views, which not only differ by a small rotation angle, but also by an additional opposite motion of the focal spot. The gray lines represent the geometry of a virtual detector with double detector sampling. The interpolation is done as if the detector values correspond to the intersection of the projection lines with the central dashed black line.

imaging, this approximation should be acceptable because the head should be positioned near the center of the field of view. If the CT system has in addition a non-uniform detector sampling, then that sampling is accounted for; for the virtual detector, always a uniform sampling is used.

2) Tam-Danielsson window: By masking the detector values with the Tam-Danielsson window, one can ensure that the remaining data provide projection values over exactly 180 degrees for each voxel (see e.g. [5]). The boundaries of the window are obtained by computing the projection of the locus of the source trajectory on a particular view. Figure 2 shows a view from a patient scan, together with a virtual view computed as outlined above and the Tam-Danielsson window for that view.

Fig. 2. Top row: A single view from the Siemens Force CT image. Middle row: a high resolution virtual view computed from two adjacent views. Bottom row: the Tam-Danielsson window for the virtual view (black values are pixels that are masked to zero).

Because of the low pitch, the view contains enough data for the application of a larger Tam-Danielsson window covering 3π of data for every voxel. Because with this window more of the data are used, the reconstructed image will have a better signal to noise ratio than when the π window were used. This is illustrated in fig. 3, which also shows that for a pitch of 0.55, the 3π window covers nearly the entire detector and therefore causes only a small loss of data.

3) FDK reconstruction: The FDK algorithm has been developed for circular cone beam CT, in which case it is exact for the central axial slice, and produces excellent images for the other slices, where exact reconstruction is not possible due to data insufficiency. Introducing some approximations, the FDK algorithm can also be applied to helical CT data.

Fig. 3. The Tam-Danielsson window of fig. 2 is shown in panel A, after changing the dimensions to make it more visible. Panel B shows the corresponding boundaries, obtained by projecting the source trajectory. Panel D shows that also the previous and next turns of the helical trajectory project on the detector, producing a Tam-Danielsson window that provides 3π of data for every voxel (panel C).

For CT reconstruction from scans acquired with a cylindrical detector, the FDK algorithm involves

1) multiplying the views with a position dependent weight,
2) filtering the projections with a modified ramp filter along a line tangent to the source trajectory,
3) computing a weighted backprojection of the filtered projections, where the weight decreases quadratically with the distance to the source,
4) zeroing the negative values.

The first step is straightforward. Considering that the view is very large and the pitch very small, as illustrated in fig. 2, the line tangent to the source trajectory is almost horizontal. Therefore, the filter is simply applied along the rows. The third step would require a modification of the backprojection procedure, if a volume based backprojector was used. However, we use a ray tracing backprojector (the adjoint operator of a ray tracing forward projector), which inherently has the desired backprojection weighting. We included the fourth step to impose the known non-negativity of the reconstructed attenuation coefficients.

We find that applying our current implementation of this procedure to the views masked with the 3π Tam-Danielsson window produces images of high resolution and good visual quality, although they suffer from subtle position dependent scaling artefacts of low spatial frequency, causing smooth variations of up to 10 HU. The artefacts rotate as one scrolls through the transaxial slices. These artefacts may be due to inaccuracies in our discretisation of the Tam window, because as shown below, even small distortions of the window can lead to significant scaling artefacts.

4) Iterative FDK reconstruction: The FDK algorithm can be extended to an iterative algorithm with the following heuristic approach:

$$\begin{aligned} X^{(0)} &= \text{FDK}(Y) \\ X^{(i)} &= X^{(i-1)} + \text{FDK}(Y - \hat{Y}(X^{(i-1)})), \end{aligned} \qquad (1)$$

where $X^{(i)}$ is the reconstruction at iteration i, FDK represents the non-iterative reconstruction described above, Y is the CT sinogram and $\hat{Y}(X)$ is the sinogram computed by forward projection of the image X. Applying one iteration eliminates the low frequency artefacts. Iterative FDK does not have proven convergence, but in simulations using noise-free data and applying 5 iterations, the sum of squared differences

between the consecutive reconstruction images and the ground truth image decreased rapidly and monotonically.

B. FDK reconstruction with motion correction

To incorporate the motion in the FDK reconstruction, we simply model it during the backprojection step, without modifying the filtering.

The FDK algorithm is obtained by discretization of a backprojection integral, i.e. an integral over the rotation angle. As a result, every view has to be multiplied with the angular increment associated with each view. For a regular trajectory, this value is a constant, but for an irregular trajectory, this angular increment is view dependent. FDK was derived for a circular orbit and therefore it is not obvious how to deal with rotations in 3D. We solve this by projecting the motions to the transaxial plane. A 3D vector pointing from the source to the center of the detector is considered. This vector is projected to the transaxial plane and we compute how much incremental 2D rotation it undergoes from one view to the next. The angular increment assigned to each view is computed as the mean of the angular increment between the previous view to the current one and the increment between the current view and the next.

The Tam-Danielsson window is valid for a helical trajectory, but not for a distorted trajectory. Due to the motion, the window no longer ensures that exactly 3π of tomographic data are used for every voxel. The resulting use of insufficient or excessive redundant data produces a very significant position dependent scaling, even for small motions. In addition, the Tam-Danielsson window has sharp boundaries, which can produce sudden changes in the local scale factors, in particular in the direction of the rotation axis. This is illustrated in figure 4 for a clinical CT study affected by patient motion. The patient moved significantly when the lower part of the head was scanned, but motion correction was done for the entire scan. Fig. 4 shows that also the very small motion corrections for the other parts of the scan resulted in very disturbing artefacts.

As a partial remedy, the 3π window was replaced with a smoothed version of the π window, illustrated in figure 5. The smoothing was done in the vertical direction with a Gaussian convolution mask, designed to ensure that the resulting window covers all detector rows. The integral of the convolution mask equals 1, which ensures at least that the decreased weight of the central π data is compensated by the use of redundant data outside the π window. This window does not provide exact weights and results in a degradation of the reconstruction in the motion-free case. On the other hand, in the motion case it may be beneficial, because it eliminates the high frequency streaks that were due to the sharp transitions in the unsmoothed Tam-Danielsson window.

To eliminate the scaling artefacts, a few iterations of the FDK algorithm are applied (eq (1)). The original FDK images and the first two FDK iterations are shown in figure 6 for the clinical motion corrupted study, using a narrow display window of about 70 HU to clearly show the scaling artefacts. The first iteration strongly suppresses and the second iteration eliminates the scaling artefacts.

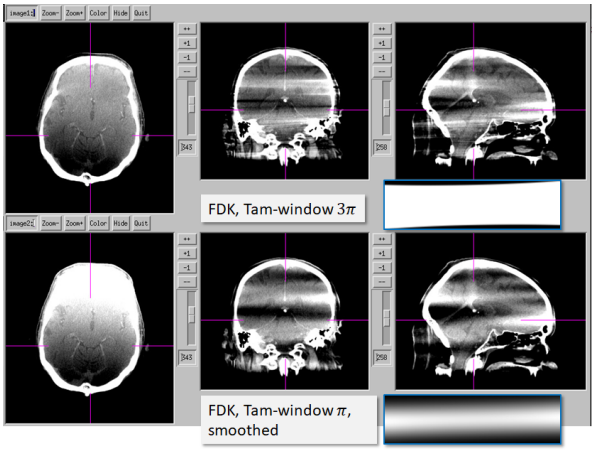


Fig. 4. When the 3π Tam-Danielsson window is used for reconstruction with motion correction, the reconstructed image suffers from artefacts, which are low frequency in the transaxial plane but can have high frequencies in coronal or sagittal slices (first row). Using instead a smoothed version of the π window makes the artefacts low frequency in all dimensions.

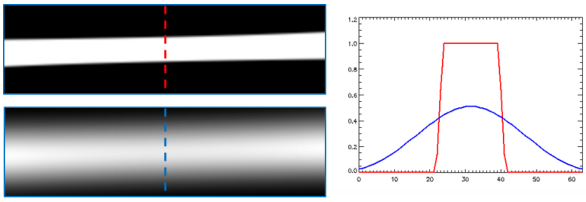


Fig. 5. The images show (resized versions of) the π Tam-Danielsson window and a Gaussian smoothed version of that. Profiles along the dashed lines are shown in the plot.

III. Results

Figures 7 and 8 show selected slices from two different scans, comparing the vendor reconstruction to the reconstruction with the proposed iterative FDK algorithm (2 iterations).

The slice of the first scan, shown in fig. 7, was affected by motion. The reconstructed images are shown with a bone window of [-250, 1250] HU. The vendor reconstruction suffers from significant patient motion artefacts, which are eliminated by FDK with motion compensation.

The slice selected from the second scan, shown in fig. 8, was not affected by motion. The reconstructed images are shown with a brain window of [-40, 110] HU. For brain imaging, the vendor software applies a dedicated bone beam hardening reconstruction which involves the forward projection of a bone only image. This correction is applied on top of the standard water-based correction. Therefore, we have implemented a similar method, using the first FDK reconstruction (iteration 0) to extract the bone image. Then the bone beam hardening correction is applied to the sinogram, which is used in the calculation of the first and second iteration of the FDK reconstruction. This means that the first FDK iteration is initialized with the FDK image obtained without bone beam hardening correction, but it is found that the images are similar enough for this to be an effective initialization. Figure 8 compares the vendor and the FDK reconstruction images. The pixel size of the slices is 0.47 mm x 0.47 mm,

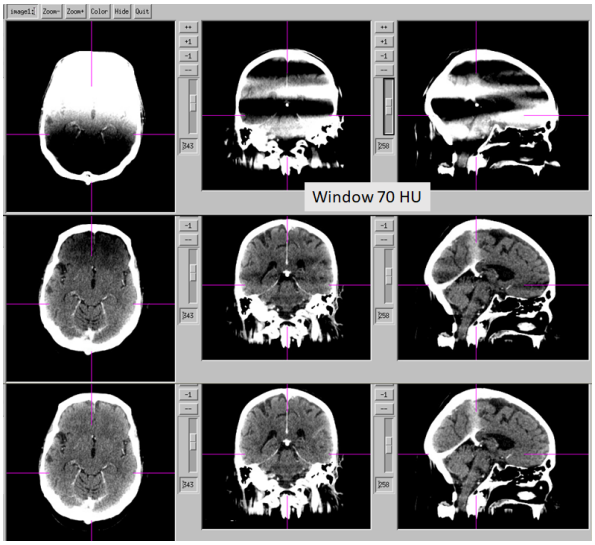


Fig. 6. The original FDK reconstruction with motion correction is shown in the first row, using a narrow display window of about 70 HU. The next rows show the first and second FDK reconstructions using the same display window.

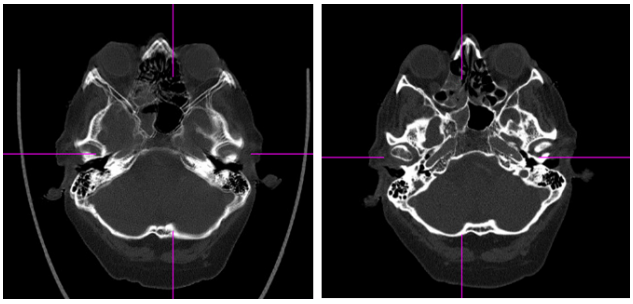


Fig. 7. The vendor reconstruction (left) and the iterative FDK reconstruction (right) from a motion corrupted CT scan. This is the same scan as the one used in figures 2, 4 and 6. A bone display of [-250, 1250] HU was used.

with a slice separation of 0.75 mm. There is a small axial mismatch between the images of less than half the slice separation. The images are not identical, which may due to small differences in the beam hardening correction and the use of a different post-reconstruction filter. We have post-filtered the FDK reconstruction with a Butterworth filter, trying to match the features of the vendor image. Nevertheless, we think that the images are rather similar and probably have very similar diagnostic value.

IV. Discussion

FDK is an effective reconstruction algorithm developed for circular cone beam CT. We have implemented an iterative version of the algorithm for helical CT with motion correction. Our results show that good images are obtained, both for bone and brain imaging. Even with 2 iterations, FDK is still much faster than ML-reconstruction for CT. When bone beam hardening correction is applied, the vendor reconstruction uses two backprojections and one forward projection, whereas the proposed approach needs three backprojections and three forward projections. If code with similar efficiency would be used, the computation time should only be about twice as long. Therefore, we believe this will help in introducing data driven motion correction into the clinical routine.

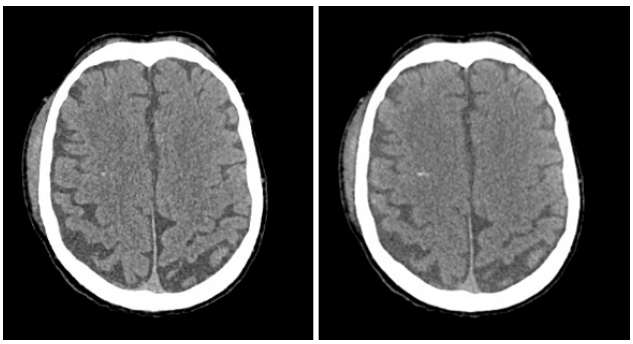


Fig. 8. The vendor reconstruction (left) and the iterative FDK reconstruction (right) obtained from a scan portion unaffected by motion. Both reconstruction are done with an additional bone beam hardening correction, which is required to obtain a uniform HU distribution inside the skull. The slices are approximately matched (there can be a mismatch of about 0.35 mm in axial position).

Acknowledgment

The authors would like to thank Frédéric Noo, Michel Defrise and Karl Stierstorfer for helpful discussions, and the latter also for his help with reading and understanding the Siemens CT data.

References

[1] T Sun, J-H Kim, R Fulton, J Nuyts, “An iterative projection-based motion estimation and compensation scheme for head X-ray CT”, *Medical Physics*, 2016, 43(10): 5705-5716.

[2] LA Feldkamp, LC Davis, JW Kress. “Practical cone-beam algorithm.” *Josa a* 1984, 1 (6): 612-619.

[3] Y Ye, S Zhao, H Yu, G Wang. “A general exact reconstruction for cone-beam CT via backprojection-filtration.” *IEEE Trans Med Imaging* 2005, 24(9): 1190-1198.

[4] S Jang, S Kim, M Kim, K Son, K-Y Lee, JB Ra, “Head motion correction based on filtered backprojection in helical CT scanning”, *IEEE Trans Med Imaging*, 2019.

[5] F Noo, J Pack, D Heuscher. “Exact helical reconstruction using native cone-beam geometries”. *Phys Med Biol*, 2003, 48.23: 3787.

Coronary Artery Motion Compensation for Short-Scan Cardiac CT Using a Spatial Transformer Network

Joscha Maier, Sergej Lebedev, Elias Eulig, Stefan Sawall, Eric Fournié, Karl Stierstorfer, and Marc Kachelrieß

***Abstract*—During a typical cardiac short scan the heart may move several millimeters. Therefore, the corresponding CT images may be corrupted by motion artifacts. Especially the assessment of small structures, such as the coronary arteries, is potentially impaired by the presence of these artifacts. In order to compensate for coronary artery motion, the so-called partial angle-based motion compensation (PAMoCo) has been proposed recently. It divides the short scan data into a given number of consecutive subphases, reconstructs them individually, and applies motion vector fields (MVFs) to match their motion states. Since the PAMoCo relies on a time consuming optimization, we propose the deep PAMoCo. It follows the basic strategy of the PAMoCo but replaces the determination as well as the application of the MVFs by a deep neural network. The proposed approach was evaluated for simulated and measured data. In any case motion artifacts could be removed efficiently from a given CT image in about 5 ms.**

***Index Terms*—Cardiac CT, coronary arteries, motion compensation, neural network, spatial transformer network.**

I. Introduction

DEVELOPMENTS such as the increase of rotation speed, multirow detectors, dual-source CT systems or high-pitch scan protocols promoted the clinical relevance of cardiac CT. Today, it finds several applications mostly related to postsurgical follow-up or diagnosis of coronary artery diseases via coronary CT angiography [1]. However, despite recent improvements in temporal resolution, cardiac motion still poses a major challenge to cardiac CT. Coronary arteries, for instance, may move up to 10 mm during a short-scan acquisition of a typical single source CT (rotation time $\approx 0.3\ s$) [2]. Since organ motion cannot be handled appropriately by conventional reconstruction approaches, artifacts are introduced to the corresponding CT images impairing their diagnostic value [3].

Consequently, several approaches have been proposed to increase the temporal resolution, and therefore, the image quality. Besides hardware-based solutions such as dual-source CT that inherently increases the temporal resolution by a factor of two [4], they can be divided into iterative approaches and motion compensation (MoCo) approaches. The former usually use projections acquired from an angular range less than 180°. The then ill-posed reconstruction problem is solved iteratively using additional prior knowledge to prevent the occurrence of limited angle artifacts [5]. MoCo algorithms, in contrast, aim to increase the temporal resolution by applying motion vector fields (MVF) that account for the cardiac motion. One possible realization of cardiac MoCo is to acquire and to reconstruct multiple phases of the cardiac cycle. Using an appropriate registration algorithm these phases are then mapped to a single reference phase, preferably one with minor motion artifacts [6]. A disadvantage of these approaches is that they strongly depend on the quality of the reference phase and might not use x-ray dose optimally. A possible alternative is to use data only from a single short scan acquisition of a certain cardiac phase [7]. The basic idea of this so-called partial angle-based motion compensation (PAMoCo) is to divide the short scan data into several subphases, each covering only a small angular range of a few degrees. Subsequently, PAMoCo determines an MVF to shift the partial angle images, i.e. the individual reconstructions of each subphase, such that their sum minimizes a metric which is sensitive to motion artifacts. However, since the motion artifact metric (MAM) is typically a strongly non-convex function its minimization is challenging and requires several reinitializations of the optimizer to escape local minima [7]. Although the PAMoCo algorithm has proven to be robust, there may be some cases where the minimum of the MAM does not correspond to the image with the least motion artifacts.

In order to increase the speed of the PAMoCo approach and to eliminate the need for a proper choice of the MAM, this study introduces the Deep PAMoCo. It follows the basic strategy of PAMoCo but replaces the determination as well as the application of the MVF by a deep neural network.

II. Material and Methods

A. Deep PAMoCo

1) Basic Principle: The basic principle of the deep PAMoCo approach is depicted in figure 1 and consists of

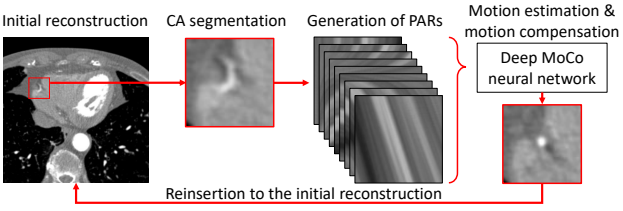


Fig. 1. Basic principle of the deep PAMoCo.

Dr. J. Maier, S. Lebedev, E. Eulig, Dr. S. Sawall and Prof. Dr. M. Kachelrieß are with the German Cancer Research Center, Heidelberg, Germany.

E. Fournié and Dr. K. Stierstorfer are with Siemens Healthineers, Forchheim, Germany.

Corresponding author: Dr. Joscha Maier (joscha.maier@dkfz.de)

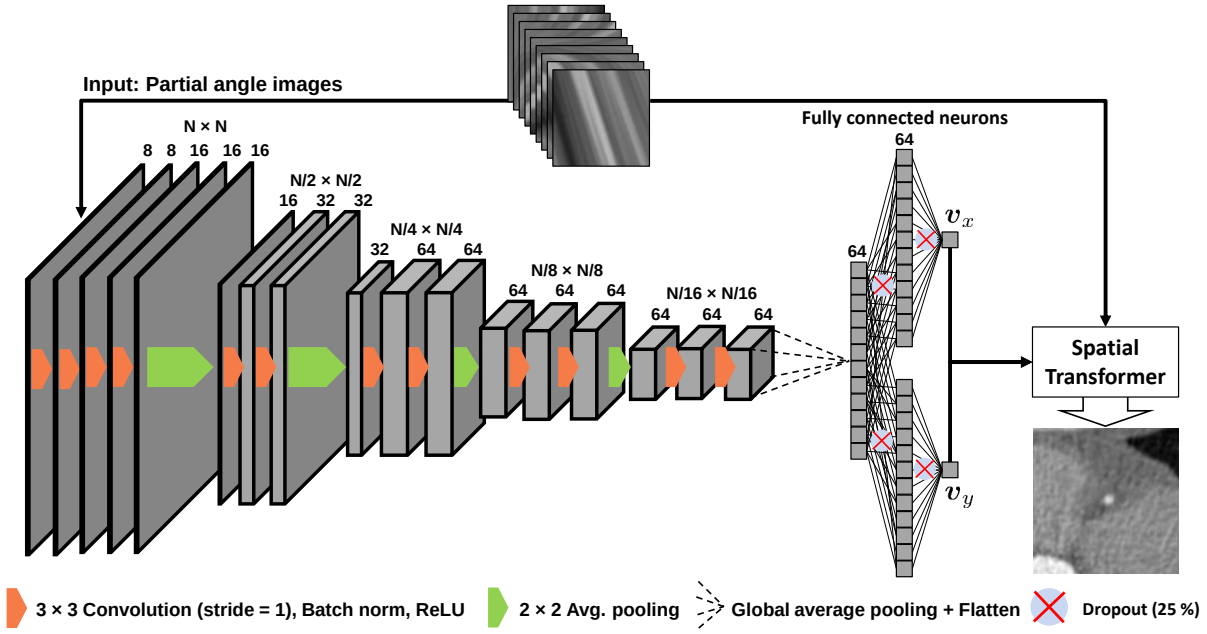


Fig. 2. Architecture of the deep PAMoCo neural network. The first part of the network estimates the motion vector while the spatial transformer applies it to the partial angle images.

the following steps. First, an initial short scan reconstruction f is performed to determine the position of the coronary arteries. Subsequently, $2K+1$ partial angle images f_k, with $-K \leq k \leq K$, are reconstructed for a patch centered at the coronary artery. Here, each partial angle image corresponds to a non-overlapping segment covering the angular interval $[\alpha_k - \Delta\alpha/2, \alpha_k + \Delta\alpha/2]$, with $\alpha_k = \alpha_0 + k \cdot \Delta\alpha$ and $\Delta\alpha = \pi/(2K+1)$. Thus, a single partial angle image has approximately a temporal resolution of $t_{\text{res}} = 0.5 \cdot t_{\text{rot}}/(2K+1)$. An appropriate choice of K allows to interpret the partial angle image as quasi-static representation of a certain subphase. To match the motion state of each subphase an MVF $\boldsymbol{s}(\boldsymbol{r}, t)$ is applied according to a predefined motion model (see section II-A2). Thus, the motion-compensated CT reconstruction is given by:

$$f_{\text{MoCo}}(\boldsymbol{r}) = \sum_{k=-K}^{K} f_k(\boldsymbol{r} + \boldsymbol{s}(\boldsymbol{r}, t_k)) \tag{1}$$

The corresponding neural network that estimates and applies the MVF is described in section II-A3.

2) Motion Model: In order to keep the number of parameters to be estimated low, Hahn et al. proposed to model the MVF by a low order polynomial [8]:

$$\boldsymbol{s}(\boldsymbol{r}, t) = w(\boldsymbol{r}) \sum_{i=1}^{I} \boldsymbol{v}_i (t - t_0)^i, \tag{2}$$

where $\boldsymbol{v}_i$ represents the i-th order motion vector and $w(\boldsymbol{r})$ is a weighting function that restricts the MVF to the region of the coronary artery. For the first proof of principle, and in contrast to conventional PAMoCo, this study only models motion up to the linear order ($I = 1$). Furthermore, we restrict the MVF to the axial plane, i.e. $\boldsymbol{v} = (v_x, v_y)^T$.

3) Network Architecture: The deep PAMoCo network architecture is depicted in figure 2. Since motion is compensated only in the axial plane, it is designed to process a given CT reconstruction slice by slice. Here, the partial angles of a 64×64 patch containing the coronary artery are used as 2D $(2K+1)$-channel input to an encoder-like architecture. It repeatedly applies convolution (3×3 kernel, batch normalization, ReLU activation) and pooling (2×2 average poling) operations to extract features related to the coronary artery motion. A successive set of fully connected neurons is used to predict the corresponding motion vector, i.e. v_x and v_y. In order to enable an end-to-end training, the application of the MVF is integrated into the neural network via a spatial transformer [9]. It uses the partial angle images as well as the motion vector as input and transforms each image according to equation (1) and (2).

4) Data Generation and Training: The deep PAMoCo network was trained using simulated data. The simulations are based on 250 clinical CT images (Siemens SOMATOM Definition AS+) that only show minor motion artifacts. Using random samples of these images as prior, CT scans were simulated by a forward projection. To account for coronary artery motion, the prior was deformed during the forward projection according to the model described in section II-A2. To cover different motion patterns, each simulation used a velocity vector $\boldsymbol{v} = v \cdot (\cos\vartheta, \sin\vartheta)^T$ with a random amplitude v between 0 mm/s and 70 mm/s and a random motion direction ϑ. Since the appearance of motion artifacts also depends on the source trajectory, the start angle of the simulated CT scan was

sampled randomly as well. Finally, a filtered backprojection of the projection data led to CT images with motion artifacts. In total 150 000 CT images, as well as the corresponding partial angle images, were generated. 200 of the 250 prior images were used to generate training data (120 000 samples) while the remaining 50 were used to generate testing data (30 000 samples). The network was trained using a supervised setting. Labels were generated by an ideal motion compensation that relies on the simulated motion vector. The training was performed on an NVIDIA GeForce RTX 2080 Ti for 100 epochs using an Adam optimizer, a batch size of 16 and the mean squared error as loss function.

B. Evaluation

The performance of the proposed approach was evaluated for simulated and measured data. Here, the simulated data correspond to the testing data described in the previous section. Patient scans were acquired with a Siemens SOMATOM Definition AS+. Both, simulations and measurements were reconstructed on a 512×512 grid with a pixel spacing of 0.34 mm. Partial angle images ($K = 12$) were generated as described in section II-A1 and used as input to the deep PAMoCo network. Simulations were evaluated quantitatively by analyzing the accuracy of the estimated motion velocity vectors as well as qualitatively by a comparison to an ideal motion compensation. Due to the absence of a ground truth, the measured data were compared to the results of the conventional PAMoCo approach.

III. Results

A. Simulation Study

The encoder-like structure of the deep PAMoCo network yields a motion velocity vector that is used as input to the spatial transformer (see figure 2). Here, we analyzed the quality of the estimated motion velocity vectors by comparing them to the ground truth that was used for simulation. Results of the x-component are shown in figure 3. A quantitative evaluation of the x- and y-component of the motion vector for all 30000 testing samples yields a root mean squared error of 3.8 mm/s with respect to the ground truth. CT reconstruction with and without motion compensation are shown in figure 4. Here, the deep PAMoCo is able to remove the motion artifacts and yields images that are almost equal to the ground truth.

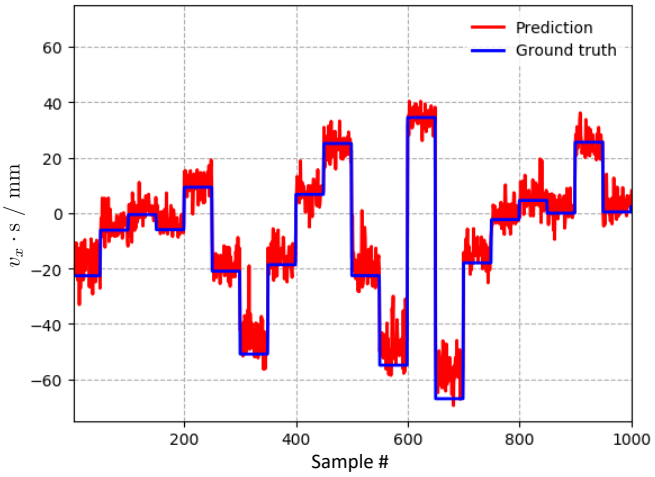


Fig. 3. X-component of the motion vectors that were used for simulation (ground truth) as well as the corresponding predictions of the deep PAMoCo. For better visibility 50 consecutive samples (from different priors) were simulated with the same motion vector.

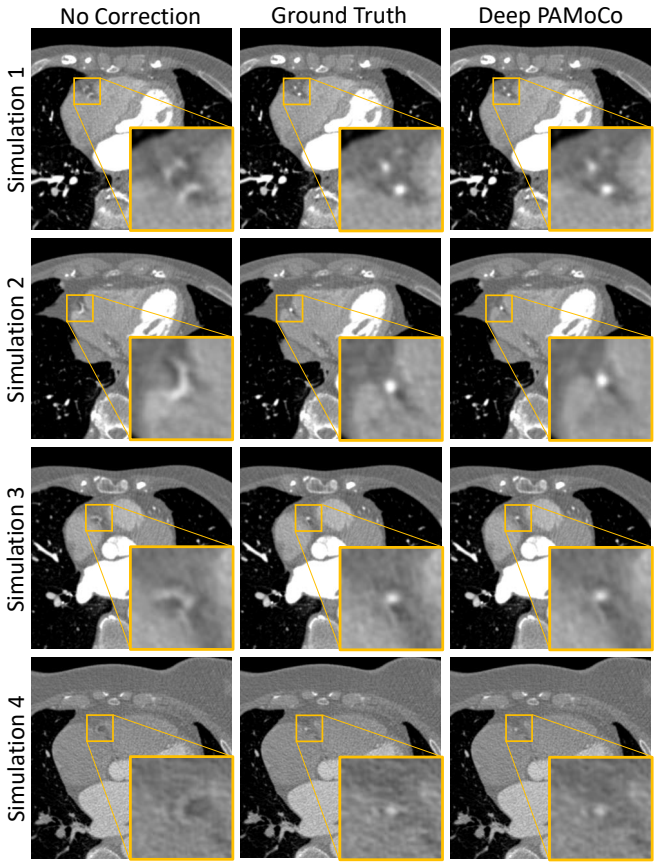


Fig. 4. Results of the simulation study without correction (left), an ideal correction (middle) as well as with the deep PAMoCo correction (right).

B. Measurements

Real patient data were evaluated by a comparison to motion compensated reconstructions using the PAMoCo approach. Exemplary results are shown in figure 5. Compared to the simulation study, there is a slightly inferior performance of the deep PAMoCo for measured data. While the conventional PAMoCo algorithm is able to remove the present motion artifacts almost completely, some of them are remaining when deep PAMoCo is applied. Possible reasons are discussed in the next section.

IV. Discussion

This study introduced the deep PAMoCo, a spatial transformer network for coronary artery motion compensation of short scan CT data. For the first proof of principle, the

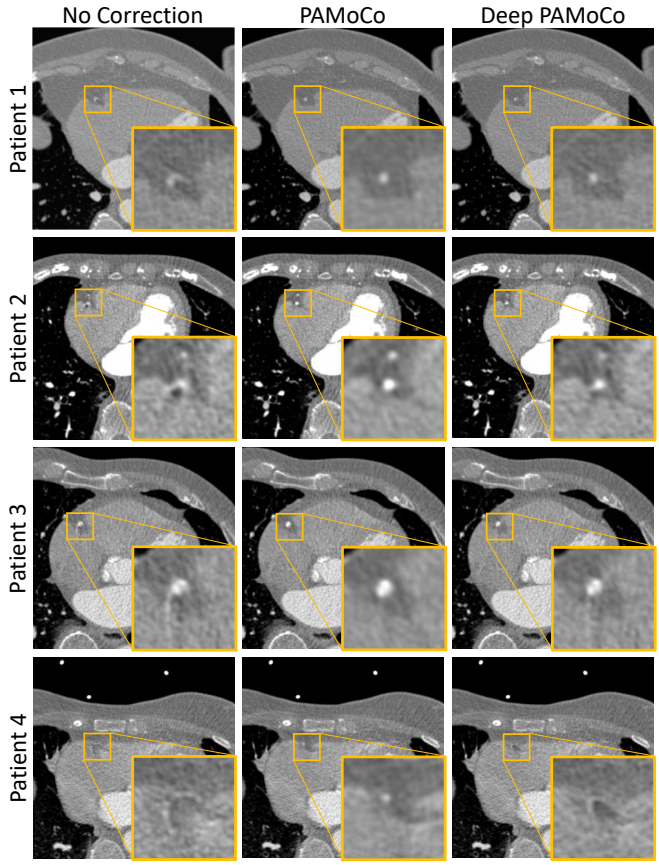


Fig. 5. Exemplary results of the patient measurements without correction (left), the PAMoCo correction (middle) as well as the deep PAMoCo correction (right). In contrast to the simulation study there are some cases (e.g. case 4) where the PAMoCo is not able to compensate the coronary artery motion completely.

proposed approach was restricted to a 2D linear motion model. Applied to simulated data that were generated according to the same motion model, the deep PAMoCo was able to provide accurate estimates of the motion velocity vector. In any case, there were only minor differences between the motion-compensated images calculated with the deep PAMoCo and the ground truth. However, considering measurements an inferior performance, especially in comparison to the conventional PAMoCo results, could be observed. This might be explained by several reasons. First of all, PAMoCo uses a more complex motion model that includes higher order motion patterns (see equation (2)) and applies the MVFs in 3D. Thus, the motion model may better fit the actual motion of the coronary arteries. In general none of the components of the deep PAMoCo is restricted to a 2D linear motion. It will be subject of further research to use 3D convolutions, additional fully connected neurons to predict higher orders of motion as well as a 3D spatial transformer.

Another current limitation is the low number of independent training samples. In this study each training sample is based on one of 200 prior images without motion artifacts. Potentially, the distribution of these images is not representative for the distribution of the patient measurements e.g. in terms of the background anatomy or the coronary artery contrast. Further work may address this issue by using unsupervised training approaches that do not require artifact-free prior images, and thus, can be trained directly on measurements.

V. Conclusions

We demonstrated the potential of using a deep neural network for coronary artery motion estimation. The use of the spatial transformer within the network enables an efficient end-to-end training and allows to remove motion artifacts from a given CT image in about 5 ms. Although the proposed approach clearly outperforms the conventional PAMoCo in terms of speed, further optimization is necessary to reach a similar image quality.

Acknowledgment

Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.

References

[1] L. J. Shaw and J. Narula, "Coronary CT angiography: An established, not emerging, basis of diagnosis and risk stratification," *JACC: Cardiovascular Imaging*, vol. 4, no. 5, pp. 565–566, 2011.

[2] S. Achenbach, D. Ropers, J. Holle, G. Muschiol, W. G. Daniel, and W. Moshage, "In-Plane Coronary Arterial Motion Velocity: Measurement with Electron-Beam CT," *Radiology*, vol. 216, no. 2, pp. 457–463, 2000.

[3] G. Pontone, J. R. Weir-McCall, A. Baggiano, A. Del Torto, L. Fusini, M. Guglielmo, G. Muscogiuri, A. I. Guaricci, D. Andreini, M. Patel, K. Nieman, T. Akasaka, C. Rogers, B. L. Nørgaard, J. Bax, G. L. Raff, K. Chinnaiyan, D. Berman, T. Fairbairn, L. H. Koweek, and J. Leipsic, "Determinants of Rejection Rate for Coronary CT Angiography Fractional Flow Reserve Analysis," *Radiology*, vol. 292, no. 3, pp. 597–605, 2019.

[4] M. Petersilka, H. Bruder, B. Krauss, K. Stierstorfer, and T. G. Flohr, "Technical principles of dual source CT," *European Journal of Radiology*, vol. 68, no. 3, pp. 362–368, 2008.

[5] M. Kachelrieß, "Iterative Reconstruction Techniques: What do they Mean for Cardiac CT?" *Current Cardiovascular Imaging Reports*, vol. 6, no. 3, pp. 268–281, 2013.

[6] R. Bhagalia, J. D. Pack, J. V. Miller, and M. Iatrou, "Nonrigid registration-based coronary artery motion correction for cardiac computed tomography," *Medical Physics*, vol. 39, no. 7, pp. 4245–4254, 2012.

[7] J. Hahn, H. Bruder, C. Rohkohl, T. Allmendinger, K. Stierstorfer, T. Flohr, and M. Kachelrieß, "Motion compensation in the region of the coronary arteries based on partial angle reconstructions from short-scan CT data," *Medical Physics*, vol. 44, no. 11, pp. 5795–5813, 2017.

[8] J. Hahn, H. Bruder, T. Allmendinger, K. Stierstorfer, T. Flohr, and M. Kachelrieß, "Reduction of motion artifacts in cardiac CT based on partial angle reconstructions from short scan data," in *Proceedings of the SPIE Medical Imaging Conference*, D. Kontos, T. G. Flohr, and J. Y. Lo, Eds., 2016.

[9] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," *Advances in Neural Information Processing Systems*, vol. 2015-Janua, pp. 2017–2025, 2015.

Metal Artifact Reduction in Photon Counting CT using Pseudo-Monochromatic Images

Achim Byl, Laura Klein, Jennifer Hardt, Stefan Sawall, Heinz-Peter Schlemmer, Sarah Heinze, and Marc Kachelrieß

***Abstract*—Photon counting CT (PCCT) inherently provides spectral information, as measured photons are separated into several energy bins. The resulting bin images can be linearly combined to produce pseudo-monochromatic images (PMIs). For dual energy CT, these images have been shown be useful to reduce moderate metal artifacts. This study investigates the usefulness of PMIs for metal artifact reduction in photon counting CT with either two or four energy bins. The linear weights are chosen by minimizing a cost function that combines an energy and a total variation term, and indicates artifact extent. The method is applied to multiple data sets of forensic specimen containing metal artifacts of varying severity. PMIs are found to reduce metal artifacts in PCCT, for both two and four bins. However, they cannot fully remove them and have low contrast. PMIs generally perform worse than frequency split normalized metal artifact reduction (FSNMAR), an inpainting based metal artifact reduction method. Applying FSNMAR to PMIs yields the best results.**

I. Introduction

Metal artifacts in CT are a combination of severe beam hardening, scatter, and photon starvation, and occur when a highly attenuating object is present in the patient, e.g. metal implants. CT images that suffer from metal artifacts exhibit thin streaks and broader hypo- or hyperdense bands. The latter predominately occur close to the metal, and can drastically lower the diagnostic value of the images, as even different types of tissue can no longer be differentiated. Therefore, metal artifact reduction (MAR) has become an important area of study.

Photon counting CT (PCCT) is an emerging technology that sorts the measured photons into multiple energy bins, yielding spectral information with every measurement. Analogously to dual energy CT (DECT), this extra information can be used for various applications, including material decomposition and contrast enhancement [1]–[5]. In this work, the spectral information will be exploited for MAR through linearly combining the bin images.

In DECT, these so-called pseudo-monochromatic images (PMI) have already been shown to reduce metal artifacts [1], [6]–[9]. A monochromatic image is favorable, as beam hardening artifacts originate from the polychromaticity of the x-ray beam. In contrast to virtual monochromatic images which are generated from the rawdata, PMIs are linear combinations of the low and the high kV images. Because they take polychromatic images as input, they can only approximate monochromatic images, and are hence called pseudo-monochromatic [1]. PCCT provides spectral information similar to DECT, such that PMIs from two energy bins should reduce metal artifacts also in PCCT. The technique has not yet been applied to more than two basis images. This study investigates the efficacy of PMIs to reduce metal artifacts in PCCT with two and four bins. The PMIs with minimal metal artifact content are compared to conventional frequency split normalized metal artifact reduction (FSNMAR) [10], [11] applied to the lowest threshold image. This threshold image contains all measured photon counts, and thus most closely resembles a conventional CT image. In addition, the PMIs are also processed with FSNMAR.

Achim Byl is with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with the University of Heidelberg, Germany. Email: achim.byl@dkfz-heidelberg.de. Laura Klein, Jennifer Hardt, Dr. Stefan Sawall, Prof. Dr. Heinz-Peter Schlemmer, and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with the University of Heidelberg, Germany. Dr. Sarah Heinze is with the University of Heidelberg, Germany.

II. Methods and Materials

PMIs are obtained by a linear combination of the bin images f_i. For the two bin case, this gives the equation

$$f_\alpha = \alpha f_1 + (1-\alpha) f_2, \tag{1}$$

where α is the linear combination weight. With knowledge of the detected x-ray spectra, this weight can be converted into an energy. However, if α gets too large it cannot be associated with an energy anymore [1] which is why this discussion focuses on α rather than on energies. For four energy bins, the equation has to be expanded to include three weighting factors:

$$f_{\alpha,\beta,\gamma} = \alpha f_1 + \beta f_2 + \gamma f_3 + (1-\alpha-\beta-\gamma) f_4. \tag{2}$$

A. Selection of Coefficients

When there is only one coefficient, it is trivial to manually select an optimal combination weight for each slice that yields the best metal artifact reduction, as done in commercially available products [6], [7], [9]. For three coefficients, this is no longer the case. To automate the process, a Nelder-Mead algorithm minimizes a cost function C with respect to the coefficients, given by

$$\begin{aligned} C(\alpha,\beta,\gamma) =& L(\alpha,\beta,\gamma) + \lambda\, TV(\alpha,\beta,\gamma) \\ C(\alpha,\beta,\gamma) =& \sum_{(x,y)\in \mathrm{ROI}} \left(f_{\alpha,\beta,\gamma}(x,y)\right)^2 \\ &+ \lambda \sum_{(x,y)\in \mathrm{ROI}} \left|\nabla\left(f_{\alpha,\beta,\gamma}(x,y)\right)\right|, \end{aligned} \tag{3}$$

for the four bin case. Here, $f_{\alpha,\beta,\gamma}(x,y)$ is again the PMI as defined in 2. The two bin case follows analogously. C

consists of an energy term L and a total variation term TV, weighted with λ, applied to an ROI of soft tissue close to the metal. Both terms will be low when there are only weak artifacts and high when there are severe artifacts. The energy term additionally penalizes large hyper- or hypodense artifacts that are homogeneous in CT-value, whose influence on TV is disproportionately small compared to the artifact severity. Without the energy term, the cost function might be lower for an image with little noise and strong artifacts than for an image with more noise but fewer artifacts, although the latter is preferable. The cost function is not applied to bone, metal, or air as the artifact reduction is most important in the soft tissue. If material with very high or low CT value was in the ROI, it would also increase the energy term.

To decide which regions consist of soft tissue, the highest energy bin is corrected with the normalized metal artifact reduction (NMAR) [10], and subsequently thresholded such that pixels between -500 and 500 HU are classified as soft tissue. The highest energy bin is chosen because it generally suffers the least from metal artifacts.

B. CT Data

All data sets were obtained with a prototype PC CT system, the SOMATOM CounT (Siemens Healthineers, Forchheim, Germany). Images of forensic specimens were acquired at 140 kV with an effective tube current of 300 mAs. The cadaver measurements are approved by the local ethics board review (S-388/2014). For two bins, the energy thresholds were set to 25 and 90 keV, and for four bins to 25, 45, 75, and 90 keV. The raw data were reconstructed at the same spatial resolution on a 512×512 matrix with 0.6 mm slice thickness and 0.5 mm pixel size using filtered backprojection. The specimens featured different metal implants, including total hip endoprostheses, and dental fillings. Likewise, the magnitude of metal artifacts ranged from low to severe.

C. FSNMAR

To illustrate how well the PMIs perform, they will be compared to a conventional inpainting based MAR algorithm, FSNMAR [10], [11]. FSNMAR can take up to three input images: the image that is to be corrected, the prior image for the initial NMAR step, and an image to replace the high frequencies in the NMAR result, i.e. the frequency split (FS) partner. In this work, the prior images and the FS partner are both generated from the original image. For the conventional FSNMAR image we apply FSNMAR to the lowest threshold image T1, since it contains all photon counts and resembles a single energy CT.

D. Analysis

To quantify the impact on soft tissue contrast, we calculate the contrast-to-noise ratio (CNR). Here, the mean values μ_k and standard deviations σ_k are measured in soft tissue and fat ROIs. Then, the CNR is computed as

$$\mathrm{CNR} = \frac{|\mu_2 - \mu_1|}{\sqrt{\sigma_2^2 + \sigma_1^2}}. \tag{4}$$

To evaluate artifact content in the images, the standard deviation σ_A is measured in a third ROI that is placed inside a region with metal artifacts. As mentioned above, PMIs are compared to the threshold image T1, conventional FSNMAR(T1) and to PMIs with additional FSNMAR processing.

To validate the performance of the coefficient selection algorithm, the automatically obtained values of α are compared to manually chosen values for the data with two energy bins.

III. Results

A. Two Bins

Figure 1 shows an example from a cadaver measurement with dental fillings, before and after processing. The original measurements ($\sigma_A = 271.8,\ 114.5,\ 237.9$ HU for bin 1, bin 2, and T1, respectively) suffer from hyperdense artifacts around the teeth and strong streaking. The PMI produced by the optimization algorithm has a blending coefficient of $\alpha = -0.25$, which does not perfectly agree with the manually chosen value of $\alpha = -0.59$. However, the actual difference is marginal, as evident in the figure. The PMI reduces the metal artifacts ($\sigma_A = 96.2$ HU), but cannot fully remove them. In comparison, FSNMAR applied to the threshold image ($\sigma_A = 43.7$ HU) removes most artifacts at the cost of introducing some new streaks. Applying FSNMAR to the PMI ($\sigma_A = 44.1$ HU) yields an even better result as artifacts close to the teeth are also removed. However, the PMI exhibits a drastically lower CNR than the threshold image with 1.1 compared to 5.8. Likewise, conventional FSNMAR has a CNR of 4.6 whereas PMI with FSNMAR has a CNR of only 1.4.

B. Four Bins

Figure 2 demonstrates pseudo-monochromatic imaging with four bins applied to the same cadaver as in Figure 1. The coefficients for the PMI are $\{0.51, -0.69, -0.56, 1.74\}$. As for the two bin case, the PMI ($\sigma_A = 94.5$ HU) reduces the metal artifacts in the image but is outperformed by conventional FSNMAR ($\sigma_A = 48.0$ HU). Likewise, applying FSNMAR to the PMI ($\sigma_A = 56.9$ HU) produces the best artifact reduction as it enhances details close to the teeth. However, the PMI has a CNR of 0.6 and 0.9 after processing with FSNMAR. This is extremely low compared to the threshold image and conventional FSNMAR, which have CNRs of 4.5 and 4.8.

Figure 3 shows another specimen, also with dental fillings. The PMI has coefficients of $\{-0.16, -0.10, 0.48, 0.79\}$. Unlike the previous case, the PMI ($\sigma_A = 116.8$ HU) does not visibly reduce artifacts. In contrast, FSNMAR ($\sigma_A = 56.9$ HU) effectively removes most artifacts but introduces minor streaks. The combination of PMI and FSNMAR ($\sigma_A = 50.7$ HU) removes the artifacts almost completely while introducing practically no new artifacts. As in the previous cases, the CNR of the PMI (2.6, 2.9 with FSNMAR) is significantly lower than the CNR of the threshold image (5.7, 5.5 with FSNMAR).

IV. Discussion and Conclusion

Pseudo-monochromatic images are able to reduce metal artifacts in photon counting CT, both with two and with four

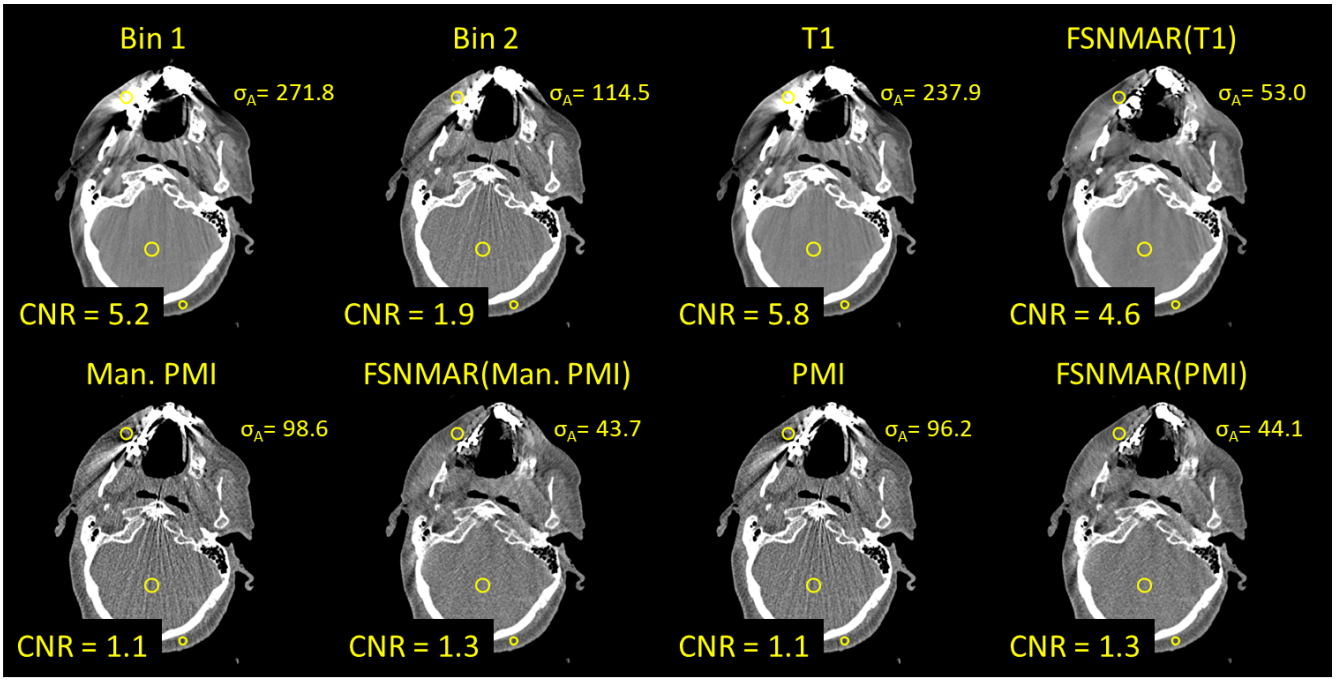


Fig. 1. Slice from a PC scan with two bins. From top left to bottom right: bin 1, bin 2, threshold image, conventional FSNMAR, PMI with manually chosen coefficient, PMI with manually chosen coefficients and FSNMAR, PMI with automatically generated coefficient, FSNMAR applied to PMI with automatically generated coefficients. Artifacts are reduced in the PMI ($\sigma_A = 96.2$ HU)and almost completely removed in the FSNMAR image ($\sigma_A = 43.7$ HU). PMI with FSNMAR ($\sigma_A = 44.1$ HU) yields the best result in terms of artifact reduction as details close to the teeth are visible. However, soft tissue contrast is strongly decreased for PMIs. $C = 50$ HU, $W = 700$ HU.

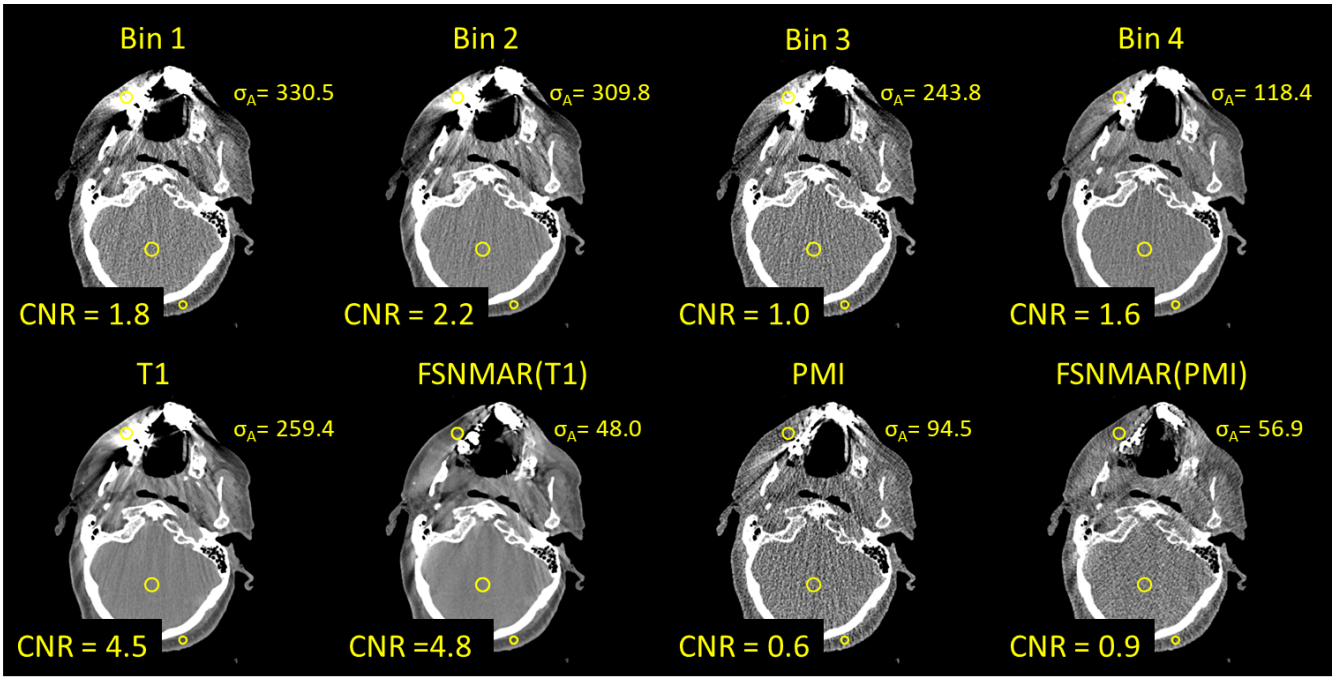


Fig. 2. Slice from a PC scan with four bins; same cadaver as in Figure 1. Top left to bottom right: bin 1, bin 2, bin 3, bin 4, threshold image, conventional FSNMAR, PMI, PMI with FSNMAR. The PMI ($\sigma_A = 94.5$ HU) is able to reduce metal artifacts and shows details close to the teeth. However, it has little soft tissue contrast (CNR = 0.6). Conventional FSNMAR ($\sigma_A = 48.0$ HU) removes most metal artifacts without increasing noise levels (CNR = 4.8), but there is little detail close to the teeth. The combination of both techniques removes practically all artifacts at the cost of low contrast ($\sigma_A = 56.9$ HU, CNR = 0.9). $C = 50$ HU, $W = 700$ HU.

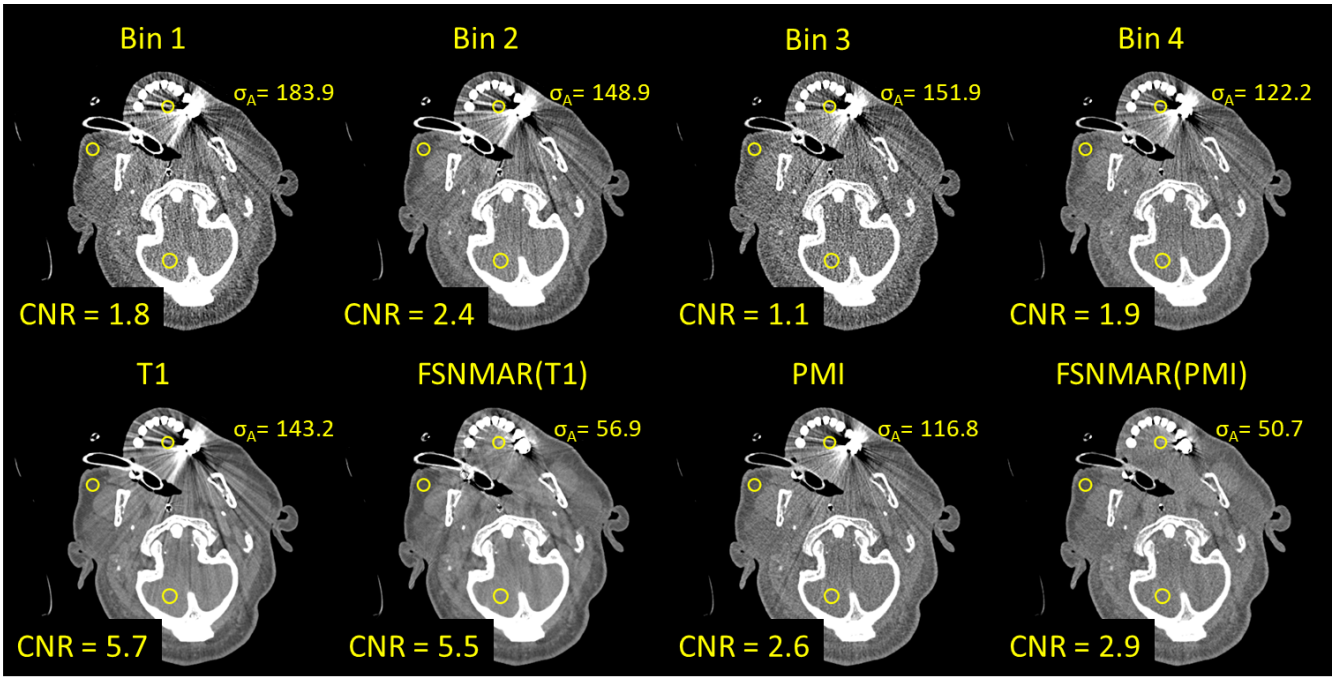


Fig. 3. Slice from a PC scan with four bins. Top left to bottom right: bin 1, bin 2, bin 3, bin 4, threshold image, conventional FSNMAR, PMI, PMI with FSNMAR. The PMI ($\sigma_A = 116.8$ HU) is not able to noticeably reduce metal artifacts compared to the threshold image ($\sigma_A = 143.2$ HU). Conventional FSNMAR removes most metal artifacts ($\sigma_A = 56.9$ HU), although some bright streaks close to the teeth and a dark patch below the metal remain. Applying FSNMAR to the PMI ($\sigma_A = 50.7$ HU) removes most artifacts with less residue artifacts than conventional FSNMAR. The CNR in PMIs (2.6 and 2.9) is significantly lower than in the conventional images (5.5 and 5.7). $C = 50$ HU, $W = 700$ HU.

energy bins. However, the artifact reduction is often marginal. Furthermore, PMIs exhibit a considerably lower CNR than the threshold image. In contrast, FSNMAR is capable of removing most metal artifacts without affecting image noise or contrast while only introducing minor streaks. Therefore, a dedicated MAR technique should be preferred over standalone PMIs.

The best results were achieved when combining PMIs with FSNMAR, as structures close to the metal or teeth retain detail, and fewer artifacts were introduced in the FSNMAR step. As FSNMAR combines a low-frequency and a high-frequency image, it is possible to use a PMI that reduces artifacts to reinsert detail into a NMAR corrected PMI that minimizes noise. This way, the CNR in the PMI can be maximized without compromising artifact reduction.

Acknowledgment

Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.

References

[1] S. Kuchenbecker, S. Faby, S. Sawall, M. Lell, and M. Kachelrieß, "Dual energy CT: How well can pseudo-monochromatic imaging reduce metal artifacts?" *Med. Phys.*, vol. 42, no. 2, pp. 1023–1036, 2015.

[2] S. Faby, S. Kuchenbecker, S. Sawall, D. Simons, H.-P. Schlemmer, M. Lell, and M. Kachelrieß, "Performance of today's dual energy CT and future multi energy CT in virtual non-contrast imaging and in iodine quantification: a simulation study," *Med. Phys.*, vol. 42, no. 7, pp. 4349–4366, 2015.

[3] X. Wang, D. Meier, K. Taguchi, D. J. Wagenaar, B. E. Patt, and E. C. Frey, "Material separation in x-ray ct with energy resolved photon-counting detectors," *Medical physics*, vol. 38, no. 3, pp. 1534–1546, 2011.

[4] S. Si-Mohamed, V. Tatard-Leitman, A. Laugerette, M. Sigovan, D. Pfeiffer, E. J. Rummeny, P. Coulon, Y. Yagil, P. Douek, L. Boussel *et al.*, "Spectral photon-counting computed tomography (spcct): in-vivo single-acquisition multi-phase liver imaging with a dual contrast agent protocol," *Scientific reports*, vol. 9, no. 1, p. 8458, 2019.

[5] R. P. Marcus, J. G. Fletcher, A. Ferrero, S. Leng, A. F. Halaweish, R. Gutjahr, T. J. Vrtiska, M. L. Wells, F. T. Enders, and C. H. McCollough, "Detection and characterization of renal stones by using photon-counting–based ct," *Radiology*, vol. 289, no. 2, pp. 436–442, 2018.

[6] L. Yu, S. Leng, and C. H. McCollough, "Dual-energy CT–based monochromatic imaging," *American journal of Roentgenology*, vol. 199, no. 5_supplement, pp. S9–S15, 2012.

[7] F. Bamberg, A. Dierks, K. Nikolaou, M. F. Reiser, C. R. Becker, and T. R. Johnson, "Metal artifact reduction by dual energy computed tomography using monoenergetic extrapolation," *European radiology*, vol. 21, no. 7, pp. 1424–1429, 2011.

[8] L. Filograna, N. Magarelli, A. Leone, R. Guggenberger, S. Winklhofer, M. J. Thali, and L. Bonomo, "Value of monoenergetic dual-energy CT (DECT) for artefact reduction from metallic orthopedic implants in post-mortem studies," *Skeletal radiology*, vol. 44, no. 9, pp. 1287–1294, 2015.

[9] T. M. Coupal, P. I. Mallinson, P. McLaughlin, S. Nicolaou, P. L. Munk, and H. Ouellette, "Peering through the glare: using dual-energy CT to overcome the problem of metal artefacts in bone radiology," *Skeletal radiology*, vol. 43, no. 5, pp. 567–575, 2014.

[10] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Normalized metal artifact reduction (NMAR) in computed tomography," *Med. Phys.*, vol. 37, no. 10, pp. 5482–5493, 2010.

[11] ——, "Frequency split metal artifact reduction (FSMAR) in computed tomography," *Med. Phys.*, vol. 39, no. 4, pp. 1904–1916, 2012.

Slow triple kVp switching CT with convolutional neural network based sinogram completion and material decomposition

Wenchao Cao and Peter B. Noël

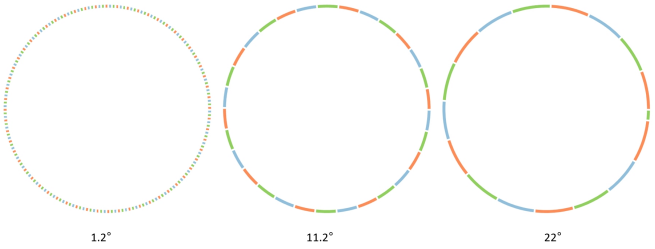
Abstract

Multi-Energy Computed Tomography (CT) has been intensively studied as it has the potential to improve diagnostics and therapy assessment. In this work, we investigate a slow triple kVp switching configuration. The kVp slowly alternates among three kVp levels and thus can be implemented on a standard single-source CT system. This scheme presents several processing challenges. First, the data is under-sampled in each energy level. Second, material decomposition in the sinogram domain is impractical as the projection images of the three energy levels are not angularly aligned. To overcome these issues, we develop a convolutional neural network (CNN) framework for sparse sinogram completion. Numerical simulations are conducted to validate the proposed scheme. Results illustrate that with the sinogram completion method, the sparsely sampled spectral CT data can be reconstructed with minimal artifact at a 22° kVp switching rate. The material decomposition network allows us to differentiate two types of contrast agents (*e.g.*, iodine and gadolinium) in the sinogram domain. Results demonstrate that 10mg/ml of iodine and gadolinium contrast agents can be separated from the water-based phantom with errors below 3%. In conclusion, a slow triple kVp switching data acquisition scheme is proposed. With the aid of the purposed CNN based data processing, it may, in the future, be used for dual-contrast agent protocols on a single-energy CT system.

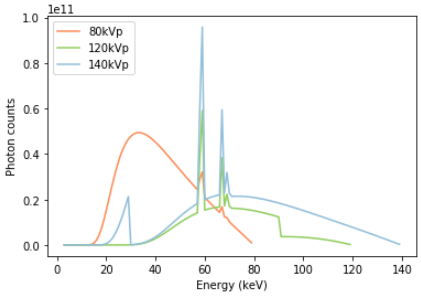
***Index Terms*—Spectral CT, slow triple kVp switching, deep learning, sinogram completion, material decomposition**

I. Introduction

Dual-contrast protocols [1; 2; 3; 4; 5] are a promising clinical Multi-Energy Computed Tomography (CT) application among other for focal liver lesion characterization [6]. One technology to enable Multi-Energy CT are photon-counting detectors (PCD) [7]. While the clinical transition is highly anticipated because of the diagnostic benefits of PCDs, the actual translation is progressing slowly due to certain shortcomings. Rapid kVp switching is a source sided spectral CT approach wherein the energy alternates between two tube voltage settings [8]. It is potentially eligible for dual-contrast material discrimination if the kVp switches among three voltages settings. However, a dedicated CT system, including a special x-ray tube, enabling fast kVp switching, and a detector with short response time, is required. More recently, other investigators have presented promising results towards slowly switch the tube voltage on a conventional CT system [9; 10]. The tube voltage switches gradually between two kVp settings. Thus the acquired dataset is a mixture of projections under two energies, with each energy level dominating a specific angular range. One of the main difficulties faced by this approach is under-sampled data. Unlike the sparse angular sampling in conventional cone-beam CT, where the sampling rate is even, the slow kVp switching data is densely sampled in an angular range but completely missing in another. As a result, interpolation is inappropriate for sinogram completion. On the other hand, the advantage of the slow kVp switching acquisition lies in the fact that the full structural information is provided by stitching projections from different energies together. The under-sampling problem can be solved if there exists a mapping function that maps the mixed sinogram to



(a)



(b)

Fig. 1: (a) An illustration of the three slow kVp switching patterns, with the corresponding spectra (b) at 80 (orange), 120 (green) and 140 kVp (blue).

Wenchao Cao and Peter B. Noël are with the Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104 USA. E-mail: wenchao.cao@pennmedicine.upenn.edu.

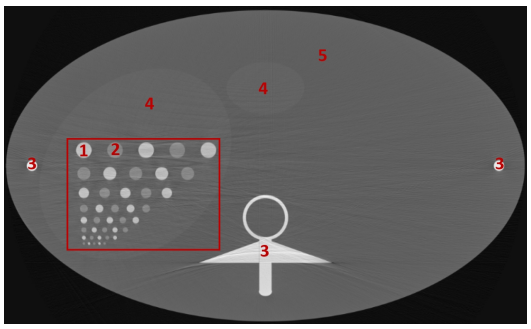


Fig. 2: The modified abdominal CT phantom for dual-contrast applications. The materials are: 1. iodine solution; 2. gadolinium solution; 3. bone; 4. soft tissue; 5. water. Both contrast agent solutions are 10mg/ml. Notice that the spheres in the red rectangle are alternately composed of iodine solution and gadolinium solution.

the individual kVps.

The main contributions of this work are summarized as follows: (1) A slow kVp switching scheme which allows triple-energy CT data acquisition on a single-energy system is proposed (Fig. 1); (2) A CNN based sinogram completion method is investigated; and (3) A CNN based material decomposition method is introduced.

II. Methods and Materials

A. Digital dual-contrast phantom

The proposed framework is applied to a simulated FORBILD abdomen phantom [11]. The phantom is modified as follows: (1) the materials are limited to soft tissue/water, iodine-based, gadolinium-based contrast material and bone as labeled in Fig. 2, and (2) the structure is simplified by excluding the high-resolution targets.

B. Sinogram completion and material decomposition networks

The network structures are shown in Fig. 3. The sinogram completion network is a modified U-Net [12], which efficiency has been proven in many medical imaging applications [13; 14; 15]. The key feature is the skip connection. U-Net links the feature maps from the encoder with the up-sampled ones, through which the full spatial resolution can be recovered. In spectral CT, the projection images under different energies are similar in the structure but different in the gray value due to the spectral deviation. Therefore, for the proposed slow kVp switching scheme, projections under different energies can be aligned together corresponding to the angular position. This structurally completed sinogram entails all geometry information but only sparse spectral data. Essentially, the network learns to map this sinogram to different gray values corresponding to individual spectra. There are two stages of sinogram completion (Fig. 4). In the first stage, the three spectral CT datasets are down-sampled and aligned together as the sparsely sampled input of the training. Then these three datasets are averaged to be used as the fully sampled output. In the second stage, the outcome

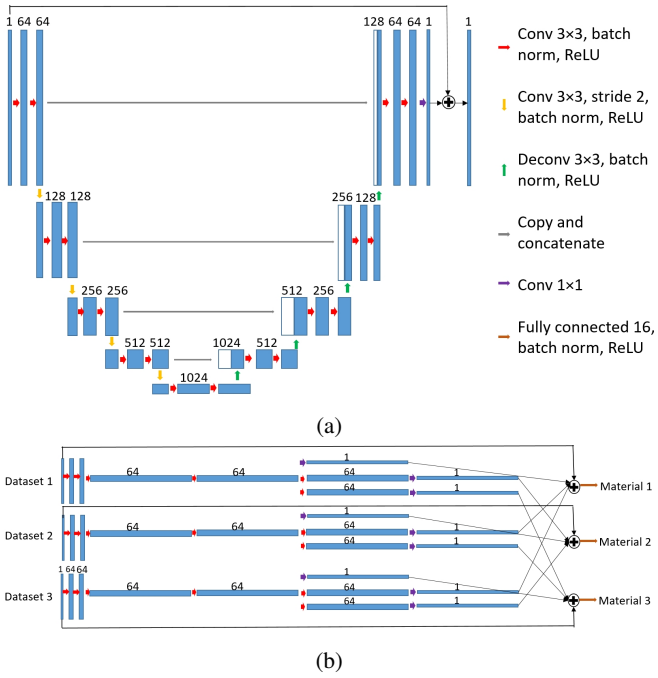


Fig. 3: (a) Sinogram completion network; (b) Material decomposition network.

of the first stage is used as the input and each spectral CT dataset is utilized as an output.

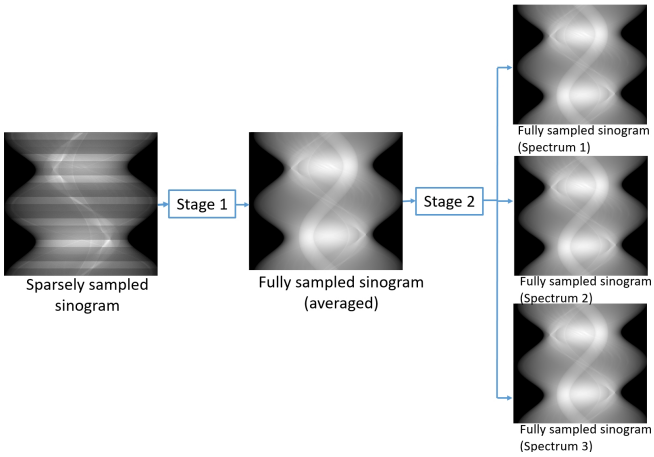
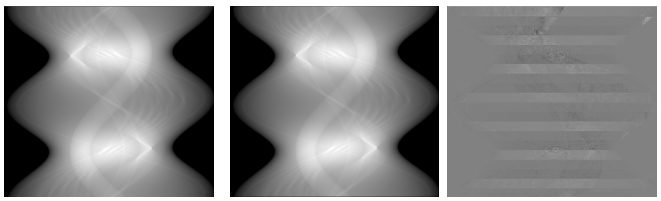


Fig. 4: An illustration of the sinogram completion network training.

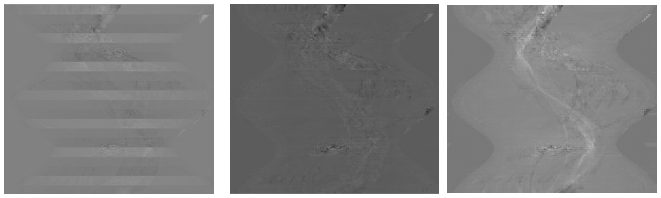
The material decomposition network involves a three-in-three-out structure that contains three streams of independent kernels. Three material maps - iodine, gadolinium, and water - are estimated separately. The key idea of the material decomposition network is to train the three streams of networks corresponding to the three materials separately.

For both sinogram completion and material decomposition training processes, the sinogram images are segmented into patches of size 64×64 pixels to reduce memory consumption.



(a) (b) (c)

Fig. 5: Results of the first stage of the sinogram completion. (a) Ground truth sinogram. (b) Completed sinogram. (c) Residual error (display window [-0.2, 0.2]).



(a) (b) (c)

Fig. 6: Residual error results of the second stage (a) 80kVp. (b) 120kVp. (c) 140kVp (display window [-0.5, 0.5]).

C. Data preparation and network training

The method for slow kVp switching simulation is shown in Fig. 1a. It is assumed that the kVp varies as a step function of the rotational angle. The number of projections for a fully sampled dataset is set as 900, in which case the angular step size is 0.4° for a 360° rotation. Three different kVp switching rates - 1.2°, 11.2° and 22° - are evaluated. The kVp switches among 80, 120 and 140kVp. Under a 1.2° switching rate, for example, 900 total views yield 300 views at each energy level, and the kVp switches every three views.

Six datasets of the FORBILD phantom are simulated. They are three spectral datasets and three material (iodine, gadolinium and water) datasets. The spectra are plotted in Fig. 1b. They are 80kVp with 2.5mm aluminum filter; 120kVp with 2.5mm aluminum and 0.225mm bismuth filters; and 140kVp with 2.5mm aluminum and 0.225mm tin filters.

III. Results

Preliminary results illustrate that the proposed CNN based method can effectively complete the sparse data. Fig. 5 and Fig. 6 show the 22° switching rate results at the two stages separately. All absolute residual errors are smaller than 0.5. In Fig. 5, although the completed sinogram is visually similar to the ground truth, gray value transitions due to sinogram stitching can be observed in the residual error image. A similar pattern can be seen in Fig. 6a. The residual error images look more uniform at 120 and 140kVp datasets.

The reconstruction results in Fig. 7 further show the efficiency of the proposed method. With sinogram completion, the FORBILD abdomen phantom can be reconstructed with minimal streaks or deformation for the three kVp settings. Further, low contrast components (water and soft tissue) are visible. Without sinogram completion, the phantom can be reconstructed with acceptable image quality at a switching rate of 1.2°, which is close in speed to current rapid kVp switching. However, visual deformation and streaks are observed for switching rates of 11.2° and 22°. The images are in general darker and the visibility of the low contrast components is further weakened due to streaks.

The material decomposition results are presented in Fig. 8. While accuracy is not degraded with a descending switching rate, streaks can be observed. A potential reason for these artifacts could be the residual error in the sinogram completion. Normalized root-mean-square error (NRMSE) [16] between material maps from the CNN decomposition and the ground truth is calculated to quantify the error. For our results, the NRMSE values are all under 3%. The method yields the best result in water decomposition and the worst result in the iodine decomposition, possibly because contrast material attenuates nonlinear compared to water.

IV. Discussion and conclusions

The present research aimed to examine the feasibility of a slow triple kVp switching data acquisition scheme. This concept can be potentially introduced to enable triple-energy spectral CT on a single-source system. A CNN framework for sinogram completion and material decomposition was developed to facilitate this scheme. A modified U-Net was first applied to complete the sparsely sampled data. Reconstruction results with and without sinogram completion were compared. We were able to illustrate with the proposed method, that artifacts caused by sparse sampling can be significantly reduced. Next, the completed datasets were utilized as input to a material decomposition CNN. Results illustrated that for a digital phantom, three materials could be discriminated. For the future, it will be essential to evaluate our concept with experimental datasets, including multiple contrast-agents.

In conclusion, a slow triple kVp switching data acquisition scheme was proposed. The findings may aid the clinical translation of dual-contrast agent protocols into the clinical day-to-day routine.

Acknowledgment

We acknowledge support through the University of Pennsylvania Research Foundation (URF) and Philips Healthcare.

References

[1] Daniela Muenzel et al. Spectral photon-counting ct: Initial experience with dual–contrast agent k-edge colonography. *Radiology*, 283(3):723–728, 2016.

[2] David P Cormode et al. Multicolor spectral photon-counting computed tomography: in vivo dual contrast imaging with a high count rate scanner. *Scientific reports*, 7(1):4784, 2017.

[3] Rolf Symons et al. Dual-contrast agent photon-counting computed tomography of the heart: initial experience. *The international journal of cardiovascular imaging*, 33 (8):1253–1261, 2017.

[4] Salim Si-Mohamed et al. Spectral photon-counting computed tomography (spcct): in-vivo single-acquisition

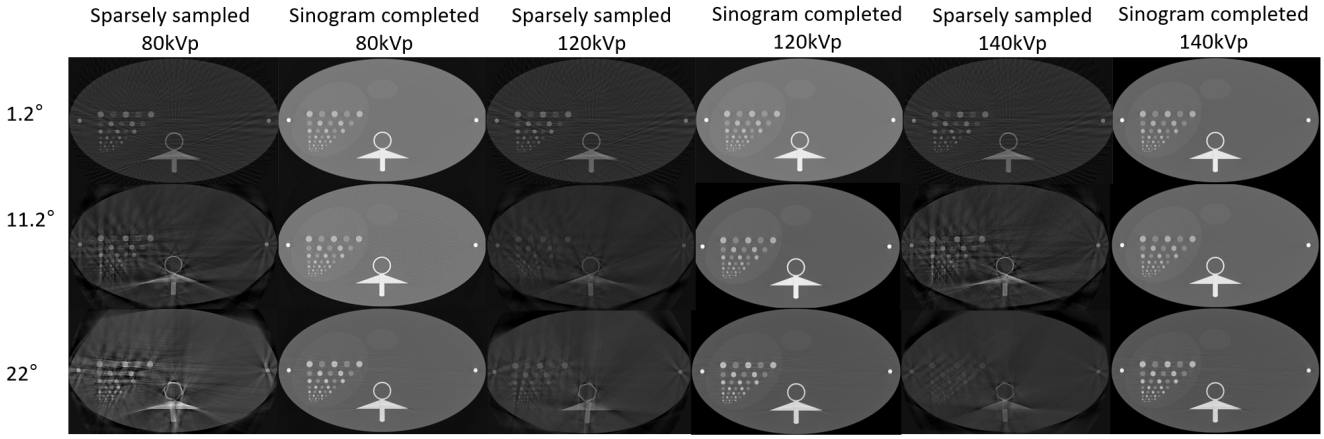


Fig. 7: Reconstructed slices (display window [0, 46.99]).

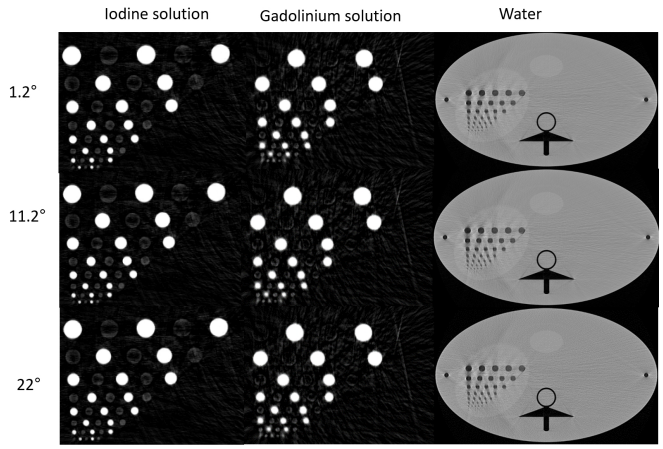


Fig. 8: Results of the material decomposition (display window [0, 46.99]).

multi-phase liver imaging with a dual contrast agent protocol. *Scientific reports*, 9(1):8458, 2019.

[5] Shengzhen Tao et al. Feasibility of multi-contrast imaging on dual-source photon counting detector (pcd) ct: An initial phantom study. *Medical Physics*, 2019.

[6] Daniela Muenzel et al. Simultaneous dual-contrast multi-phase liver imaging using spectral photon-counting computed tomography: a proof-of-concept study. *European radiology experimental*, 1(1):25, 2017.

[7] Martin J Willemink et al. Photon-counting ct: technical principles and clinical prospects. *Radiology*, 289(2):293–312, 2018.

[8] Yu Zou and Michael D Silver. Analysis of fast kv-switching in dual energy ct using a pre-reconstruction decomposition technique. In *Medical Imaging 2008: Physics of Medical Imaging*, volume 6913, page 691313. International Society for Optics and Photonics, 2008.

[9] Timothy P Szczykutowicz and Guang-Hong Chen. Dual energy ct using slow kvp switching acquisition and prior image constrained compressed sensing. *Physics in Medicine & Biology*, 55(21):6411, 2010.

[10] Gilad Shechter. Under-sampled, multi-energy computed tomography (ct) data acquisition data processing, May 12 2015. US Patent 9,031,298.

[11] M Kachelrieß. The forbild ct–simulation phantoms. In *Proc. of the 1999 Int. Meeting on Fully 3D Image Reconstruction*, volume 383, 1999.

[12] Hoyeon Lee et al. Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2):109–119, 2018.

[13] Olaf Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[14] Hao Dong et al. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *annual conference on medical image understanding and analysis*, pages 506–517. Springer, 2017.

[15] Ozan Oktay et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[16] Paul J Schleyer et al. Retrospective data-driven respiratory gating for pet/ct. *Physics in Medicine & Biology*, 54(7):1935, 2009.

Projective Reconstruction of Acquisition Geometry without Point Correspondences

Tobias Würfl, Nicole Maaß, Frank Dennerlein, Andreas K. Maier

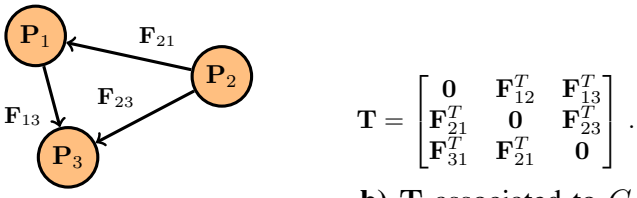
***Abstract*—Despite some fundamental differences in the image formation model 3D Reconstruction in the fields of Computed Tomography and Computer Vision share some similar concepts. In Computed Tomography, reconstruction refers to the problem of estimating 3D structure from 2D projection images assuming that accurate information about the camera geometry is available while in Computer Vision reconstruction typically refers to the task of simultaneously estimating the 3D structure and the camera motion solely from 2D projection images. In Computed Tomography accurate information about the camera geometry or motion is usually acquired by calibration and scanners are manufactured to support reproducible trajectories. Being able to reconstruct without requiring calibration and a reproducible trajectory would solve the problem of rigid motion estimation and enable new types of scanners. However, it is difficult to transfer Computer Vision techniques to Computed Tomography because point correspondences are hard to establish since all structures are superimposed over each other in projection images. Recently we introduced a new algorithm two estimate pairwise relative geometry based solely on the 2D projection images without point correspondences. In this work, we present the adaptation of a structure from motion algorithm to obtain a projective reconstruction of the whole scene from pairwise geometry estimates. In a simulation experiment, we show that we are able to estimate the camera geometry only based on projections.**

I. Introduction

RECONSTRUCTION of the three-dimensional structure of a scene has become a mature task both in Computer Vision (CV), where color images of reflected light are used and in Computed Tomography (CT), where gray value images of transmitted X-ray light are used. Despite those differences in the image formation model those two reconstruction tasks share some concepts. In the following, we will consistently use the wording of the Computer Vision field and refer to a 3D reconstruction result as *structure* instead of *volume* and we will refer to the *camera geometry* expressed as projection matrices as *motion*.

In CV it is nowadays conventional to reconstruct not only the structure but the motion simultaneously only based on the acquired images. In CT the structure reconstruction is more advanced but no system exists which simultaneously reconstructs the structure and the motion. This disparity is on one hand caused by the fact that reconstruction in CT is much more demanding because single pixels in input data do not represent a single point in space but an integral over the attenuation along a line and therefore one has to reconstruct the motion fully before being able to reconstruct the structure. On the other hand, existing techniques from Computer Vision estimating the motion between two cameras are not applicable to CT because algorithms to identify corresponding points fail. For these reasons motion reconstruction has never been approached in full generality in CT. Instead prior work has concentrated on only estimating slight deviations from the calibrated motion which has been studied under the terms of rigid (patient) motion estimation [1], [2] and misalignment compensation [3] while in CV researchers have shown how to reconstruct city-scale scenes from unordered image collections [4].

T. Würfl and A. K. Maier are with the Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany.
N. Maaß and F. Dennerlein are with Siemens Healthcare GmbH, Erlangen, Germany.



a) A triplet G **b)** $\mathbf{T}$ associated to G

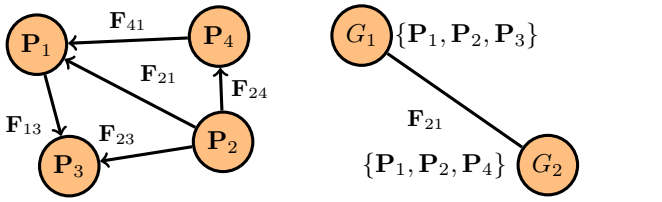
Fig. 1: Fully connected triplet and 3-view fundamental matrix.

Reconstructing both structure and motion simultaneously would solve both misalignment caused by slight deviations from the calibrated motion and patient motion compensation. Being able to deal with uncalibrated or non-reproducible trajectories could enable high quality reconstructions on new types of CT scanners which consist of flexibly mounted detectors and sources which can move freely and independently like robotic X-ray systems. Currently such systems are investigated both in medical and industrial imaging [5] but are difficult to calibrate and require specific procedures to deal with those problems.

In CV many algorithms referred to as *structure from motion (SfM)* exist to obtain a reconstruction of the motion in the form of projection matrices while simultaneously reconstructing the structure in form of 3D point coordinates. The central component in SfM algorithms are methods to identify corresponding points. Those algorithms typically consist of calculating descriptors of keypoints in projections [6] which are then matched in multiple projections. From those corresponding points most methods compute a minimal parametrization of the relative geometry of pairs of images known as the fundamental matrix [7]. The fundamental matrix $\mathbf{F}_{ij}$ is defined by the equation

$$\mathbf{x}_j^T \mathbf{F}_{ij} \mathbf{x}_i = \mathbf{0}\,, \tag{1}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are corresponding projection coordinates of the same 3D point in projections i and j in homo-



a) Viewing graph G **b)** Triplet cover G_τ of G

Fig. 2: A viewing graph and a triplet cover for it.

geneous coordinates. The fundamental matrix can be used to obtain a projective reconstruction of the two associated projection matrices $\mathbf{P}_i$ and $\mathbf{P}_j$. A *projective reconstruction* of N projection matrices denotes a non-unique recovery up to a projective transformation $\mathbf{H}$. For more than two projections different algorithms exist to obtain a projective reconstruction of all projection matrices. The central method in all SfM methods for this problem is known as bundle adjustment. We denote M estimated 3D points as $\hat{\mathbf{X}}^m$ and their measured projections by N estimated projection matrices $\hat{\mathbf{P}}_i$ as $\mathbf{x}_i^m$ so that: $\hat{\mathbf{P}}_i\hat{\mathbf{X}}^m = \mathbf{x}_i^m$. This allows us to formulate projective reconstruction as an optimization problem jointly optimizing for all $\hat{\mathbf{P}}_i$ and $\hat{\mathbf{X}}^m$ by minimizing:

$$\underset{\hat{\mathbf{P}}_i,\hat{\mathbf{X}}^m}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{m=1}^{M} \|\hat{\mathbf{P}}_i\hat{\mathbf{X}}^m - \mathbf{x}_i^m\|_2^2 \,. \tag{2}$$

This construction

$$\|\hat{\mathbf{P}}_i\hat{\mathbf{X}}^m - \mathbf{x}_i^m\|_2 \,, \tag{3}$$

is called the reprojection error. Once this projective reconstruction of structure and motion has been computed the matrix $\mathbf{H}$ has to be estimated to upgrade the projective reconstruction to a metric reconstruction. The input to those methods is a projective reconstruction and an assumption about the intrinsic camera parameters such as rectangular pixels. A popular method for this by Chandraker et al. [8] formulates this as rank constrained semidefinite optimization problem.

The key challenge in transferring this pipeline to CT is that because of the line-integral projection model the identification of corresponding points fails [9]. To overcome this challenge we recently proposed a method to estimate fundamental matrices without the use of corresponding points [10]. The method works by formulating fundamental matrix estimation as a non-convex minimization of a inconsistency metric between two projections.

However, since bundle adjustment also requires corresponding points, we can not use it to obtain a projective reconstruction. In this work we adapt a recently proposed method by Kasten et al. [11] for CT calculating a projective reconstruction solely from collections of fundamental matrices.

II. Method

The fundamental matrices define a *viewing graph* [12]. A viewing graph $G(\mathcal{V}, \mathcal{E})$ consists of nodes $v_i \in \mathcal{V}$ representing the unknown projection matrices $\mathbf{P}_i$ and directed edges $e_{ij} \in \mathcal{E}$ which represent the fundamental matrices $\mathbf{F}_{ij}$ relating them. The edges are directed because the edge could either represent $\mathbf{F}_{ij}$ or $\mathbf{F}_{ji} = \mathbf{F}_{ij}^T$ if no direction was indicated. A viewing graph can at most contain $\binom{n}{2} = \frac{n^2-n}{2}$ distinct fundamental matrices because $\mathbf{F}_{ij} = \mathbf{F}_{ji}^T$ and $\mathbf{F}_{ii} = \mathbf{0}$. The viewing graph can be weighted using metrics to quantify the certainty in the estimate of the associated fundamental matrices.

For some special cases of such a viewing graph, linear solutions for the $\mathbf{P}_i$ exist which are compatible with the measured fundamental matrices. The simplest case is given by two nodes connected by a single fundamental matrix. This case can be extended to a viewing graph consisting of three nodes fully connected by three fundamental matrices by first obtaining a projective reconstruction for two nodes and calculating the missing projection matrix $\hat{\mathbf{P}}_3$ from those previously computed projection matrices $\hat{\mathbf{P}}_1$, $\hat{\mathbf{P}}_2$ and two Fundamental matrices $\mathbf{F}_{13}$, $\mathbf{F}_{23}$. The algorithm is briefly described by Hartley and Zisserman [13, p. 386]. They show that $\mathbf{P}_3^T\mathbf{F}_{13}\mathbf{P}_1$ and $\mathbf{P}_3^T\mathbf{F}_{23}\mathbf{P}_2$ are skew-symmetric. This gives rise to 20 equations because each of the two skew-symmetric 4×4 matrices has six independent entries and four entries equal to zero. A different reconstruction formula for this case was given by Kasten et al. [11]. Their formula uses a so called 3-view fundamental matrix $\mathbf{T}$ which assembles the three fundamental matrices of this special viewing graph to a 9×9 matrix as seen from the matrix in Fig. 1 with a depiction of the graph.

They give an explicit formula how to obtain a projective reconstruction from $\mathbf{T}$. These methods can be extended to solve viewing graphs with more than three projections by first using the three-view reconstruction on an arbitrary subgraph and adding more projections to the projective reconstruction by using two fundamental matrices to calculate it. We will denote such a sequence of solvable three-view reconstruction problems as *triplet cover*. Figure 2 shows an example of such graphs.

More formally a triplet cover $G_\tau(\mathcal{V}_\tau, \mathcal{E}_\tau)$ of $G(\mathcal{V}, \mathcal{E})$ is a connected, undirected graph with nodes $v_k^\tau \in \mathcal{V}_\tau$ representing viewing graphs G_k. These G_k contain a set $\mathcal{V}_k$ of exactly three nodes v_i from $\mathcal{V}$ and are fully connected implying that they have a 3-view fundamental matrix $\mathbf{T}$. Edges $e_{kl}^\tau \in \mathcal{E}_\tau$ of a triplet cover denote that two nodes v_k^τ and v_l^τ contain two identical projections: $|\mathcal{V}_k \cap \mathcal{V}_l| = 2$. This is equivalent to stating that each edge denotes a shared fundamental matrix $\mathbf{F}_{kl}$. Additionally every node $v_i \in \mathcal{V}$ needs to be contained in at least one viewing graph of the triplet cover.

As we have outlined before a triplet cover can be traversed and a projective reconstruction of the geometry can be calculated. This shows that a triplet cover fully specifies the geometry. However, many more fundamental matrices can be calculated which express this geometry redundantly. The collection of all of those can be represented as a N-view fundamental matrix $\mathbf{N}$. It takes the form of a symmetric matrix compiling all possible fundamental matrices:

$$\mathbf{N} = \begin{bmatrix} \mathbf{0} & \mathbf{F}_{12}^T & \dots & \mathbf{F}_{1N}^T \\ \mathbf{F}_{21}^T & \mathbf{0} & \dots & \mathbf{F}_{2N}^T \\ \vdots & & \ddots & \\ \mathbf{F}_{N1}^T & \dots & \mathbf{F}_{NN-1}^T & \mathbf{0} \end{bmatrix} . \tag{4}$$

Note that we use the transpose operator here: $\mathbf{F}_{ij}^T$ which is different from the definition of Kasten et al. because they introduce this construction using a atypical definition of $\mathbf{F}_{ij}$ with swapped indices ij. Identical to the viewing graph, $\mathbf{F}_{ij} = \mathbf{F}_{ji}^T$ and $\mathbf{F}_{ii} = \mathbf{0}$. Additionally we need rows of $\mathbf{N}$ denoted as $\mathbf{N}_i$ with a single index e.g.

$$\mathbf{N}_1 = \begin{bmatrix} \mathbf{0} & \mathbf{F}_{12}^T & \dots & \mathbf{F}_{1N}^T \end{bmatrix} . \tag{5}$$

Because a triplet cover and its collection of 3-view fundamental matrices already suffices to obtain a projective reconstruction this representation is redundant. Reconstruction from such a redundant representation can easily be achieved by the same method as for the triplet cover, if the redundant projection matrices are *consistent* with each other. Consistent means that no matter what triplet we use to reconstruct a projection matrix $\mathbf{P}_i$ the result is identical. However, in practical situations where we only have noisy estimates of fundamental matrices this consistency is not guaranteed.

In their work Global Projective Structure from Motion [11] Kasten et al. prove that a set of necessary and sufficient conditions for consistency of a N-view fundamental matrix $\mathbf{N}$ with a set of N projection matrices $\mathbf{P}_i$ which do not only have collinear centers is:

1) $\text{rank}(\mathbf{N}) = 6$ and $\mathbf{N}$ has exactly 3 positive and 3 negative eigenvalues,
2) $\text{rank}(\mathbf{N}_i) = 3$ for all $i = 1, \cdots, n$.

These conditions can be used to formulate an optimization problem finding a consistent N-view fundamental matrix $\mathbf{N}$ which is closest to the measured noisy data.

We now introduced all concepts for the projective reconstruction method. It consists of four steps:

Algorithm 1 Projective Reconstruction

1: $G_\tau \leftarrow construct_triplet_cover(G)$
2: $G_\tau \leftarrow optimize_N(G_\tau)$
3: $\mathbf{P}_i \leftarrow reconstruct(G_\tau)$

We will briefly describe each step and outline how we adapted them to CT.

A. Construct Triplet Cover

The triplet cover is generated from an input viewing graph by first enumerating sets of possible triplets by constructing maximum spanning trees T_i from a viewing graph G in a loop, each time removing the edges which are contained in T. Kasten et al. use the number of matches to weight the viewing graph. In our case we can not use this and therefore currently use a weight of one reducing this step to a random selection. We loop for 5 iterations as reported by Kasten et al. This loop produces 5 edge-disjoint maximal spanning trees. From those all possible triplets are enumerated and an initial triplet cover G_τ is created.

Since not all triplets are equally well suited for reconstruction we prune G_τ to contain only suitable triplets for reconstruction. A necessary assumption for reconstruction is that the three $\mathbf{P}_i$ are not all collinear. Kasten et al. [11] define a measure for collinearity which we use identically. Additionally if the 3-view fundamental matrix of a triplet is inconsistent even after optimization of its matrices it is heuristically favorable to swap it for a more consistent triplet. For measuring consistency the algorithm of Section II-B is used. Assume for now it produces a projection of a measured, possibly noisy, 3-view fundamental matrix $\hat{\mathbf{T}}_k$ to a consistent $\mathbf{T}_k$. Inconsistency of a triplet is now defined as residual inconsistency of its 3-view fundamental matrix:

$$c_k = \|\mathbf{T}_k - \hat{\mathbf{T}}_k\|_F \,, \tag{6}$$

where $\|\dots\|_F$ denotes the Frobenius-norm. Those two measures are calculated for all v_k of G_τ and used to prune G_τ.

B. Optimize N-view Fundamental Matrix

After a suitable triplet cover G_τ has been constructed the fundamental matrices of each triplet will in general still be inconsistent because each fundamental matrix estimate is noisy. If this inconsistency is not handled properly, subsequent reconstruction algorithms will produce projections to consistent pairs which are optimal according to unsuitable non-geometric cost functions. This usually leads to unsatisfying results. Therefore a general strategy is to formulate such projection problems explicitly making use of all available geometric constraints on a consistent solution. This prior knowledge helps finding more favorable projections to consistent sets.

Because a triplet cover is a solvable viewing graph, it uniquely defines a projective reconstruction which in turn uniquely defines a N-view fundamental matrix. This allows to formulate an optimization problem for a consistent N-view fundamental matrix $\mathbf{N}$ from an inconsistent $\hat{\mathbf{N}}$ only specified by a triplet cover:

$$\begin{aligned} &\underset{\mathbf{N}}{\text{argmin}} \left\{ \sum_{k=1}^{|\mathcal{V}_\tau|} \|\mathbf{T}_k - \hat{\mathbf{T}}_k\|_F^2 \right\} \\ &\text{s.t.} \quad \mathbf{N} = \mathbf{N}^T \,, \mathbf{F}_{ii} = \mathbf{0} \,, \text{rank}(\mathbf{T}_k) = 6 \,. \end{aligned} \tag{7}$$

This optimization is in general a non-convex problem despite the convex objective because of the rank constraint. However, for the rank constraint a proximal operator is available and those problems are often solvable with practical accuracy. Therefore the problem can be tackled using an Alternating Direction Method of Multipliers (ADMM) method. ADMM leads to a iterative solver with closed form update formulae and two auxiliary variables. For the derivation and the details we refer to Kasten et al. [11].

A difference to Kasten et al. in our implementation is that they normalize the Fundamental matrix using the location of interest points in each image to zero mean and unit variance similar to the normalized 8-point algorithm [14]. After optimization the fundamental matrices are denormalized. Since we do not have interest points we do not perform this normalization.

C. Projective Reconstruction Using a Triplet Cover

After optimization of the triplet cover we assume it is consistent and perform the actual projective reconstruction. Kasten

Method	Sum of error	Conclusion
GT-Jittered	9.132088	Error introduced by jitter.
Proposed	6.849069	Image-based geometry.
Ideal $\mathbf{F}_{ij}$	0.000068	Fundamental matrix estimation determines the quality.

TABLE I: Results of our experiment.

et al. propose to traverse the triplet cover and reconstruct each triplet independently using their reconstruction formula. They follow this up by traversing the triplet cover a second time unifying the projective frame of all reconstructions to be identical. We found that this method produced large errors in some geometric setups. Instead we only use their formula for the first triplet and for all other triplets we use the reconstruction formula for the missing projection matrix of Hartley and Zisserman [13] mentioned before.

III. Experiment

We present a proof-of-concept experiment on simulated data. As simulated dataset we use a phantom described in [10] consisting of a spherical acquisition of homogeneous beads of different densities. In our experiment we use 25 projections. We use our previously published method [10] to estimate fundamental matrices from 2D projection images. We initialize the fundamental matrices from jittered projection matrices with uniform random offsets of the principal point coordinates u_0 and v_0. As range we use shifts in pixels from $[-5, 15]$. Our experiment is evaluated by reporting the reprojection error as defined in Equation 3 of the origin against the groundtruth geometry. To enable this comparison we need to upgrade our projective reconstruction to a metric one. We do this by assuming knowledge of two ideal, unjittered projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$. Using those a simple strategy to estimate the homography $\mathbf{H}$ is to apply the method for unification of two triplets of Kasten et al. Since the two known projection matrices are in the Euclidean frame of reference a transformation $\mathbf{H}$ which maps the two estimated corresponding matrices $\hat{\mathbf{P}}_1$ and $\hat{\mathbf{P}}_2$ of the projective reconstruction to their known counterparts is the desired homography. We used this method in our experiments to assess the performance of the projective reconstruction algorithm. We also report the error of our projective reconstruction algorithm if ideal fundamental matrices are reported. The numeric results are summarized in Table I.

IV. Discussion and Conclusion

We presented a method for projective reconstruction of the acquisition geometry from projections for Computed Tomography. To this end, we used a previously developed method to estimate fundamental matrices without point correspondences and adapted a recently proposed method for projective reconstruction from sets of fundamental matrices to our setting.

We showed that the results of our pipeline strongly depends on the accuracy of the fundamental matrix estimation method. This suggests that the most difficult problem in this line of research is the accurate estimation of those. Future work could combine our fundamental matrix estimation method e.g. with algorithms based on landmarks such as the method of Bier et al. [15] to obtain point correspondences in this difficult setting. This would enable to use a geometric error which is known to improve results in Computer Vision.

A shortcoming of our work is that the construction of maximum spanning trees currently uses uniform weights. Subsequent work should define a sensible cost based on uncertainty of the fundamental matrix estimation algorithm. Similarly the measures for collinearity and accuracy of Kasten et al. should be investigated for their effectiveness in our setting.

Disclaimer

The concepts and information presented in this paper are based on research and are not commercially available.

References

[1] A. Preuhs, N. Ravikumar, M. Manhart, B. Stimpel, E. Hoppe, C. Syben, M. Kowarschik, and A. Maier, "Maximum likelihood estimation of head motion using epipolar consistency," in *Bildverarbeitung für die Medizin 2019*. Wiesbaden: Springer Fachmedien Wiesbaden, 2019, pp. 134–139.

[2] J. Wicklein, H. Kunze, W. A. Kalender, and Y. Kyriakou, "Image features for misalignment correction in medical flat-detector ct," *Medical physics*, vol. 39, no. 8, pp. 4918–4931, 2012.

[3] L. Von Smekal, M. Kachelrieß, E. Stepina, and W. A. Kalender, "Geometric misalignment and calibration in cone-beam tomography: Geometric misalignment and calibration in cone-beam tomography," *Medical physics*, vol. 31, no. 12, pp. 3242–3266, 2004.

[4] J.-M. Frahm, P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik *et al.*, "Building rome on a cloudless day," in *European Conference on Computer Vision*. Springer, 2010, pp. 368–381.

[5] P. Landstorfer, J. Hiller, and M. Herbst, "Investigation of positioning accuracy of industrial robots for robotic-based x-ray computed tomography," in *Proceedings of the 9th Conference on Industrial Computed Tomography (iCT 2019)*, 2019.

[6] D. G. Lowe *et al.*, "Object recognition from local scale-invariant features." in *iccv*, vol. 99, no. 2, 1999, pp. 1150–1157.

[7] Q.-T. Luong and O. Faugeras, "The fundamental matrix: Theory, algorithms, and stability analysis," *International Journal of Computer Vision*, vol. 17, pp. 43–75, 01 1996.

[8] M. Chandraker, S. Agarwal, F. Kahl, D. Nistér, and D. Kriegman, "Autocalibration via rank-constrained estimation of the absolute quadric," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[9] M. Klüppel, J. Wang, D. Bernecker, P. Fischer, and J. Hornegger, "On Feature Tracking in X-Ray Images," in *Bildverarbeitung für die Medizin 2014*, ser. Informatik aktuell, T. Deserno, H. Handels, H.-P. Meinzer, and T. Tolxdorff, Eds., Berlin Heidelberg, 2014, pp. 132–137.

[10] T. Würfl, A. Aichert, N. Maaß, F. Dennerlein, and A. Maier, "Estimating the fundamental matrix without point correspondences with application to transmission imaging," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 1072–1081.

[11] Y. Kasten, A. Geifman, M. Galun, and R. Basri, "Gpsfm: Global projective sfm using algebraic constraints on multi-view fundamental matrices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3264–3272.

[12] N. Levi and M. Werman, "The viewing graph," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 1. IEEE, 2003, pp. I–I.

[13] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[14] R. I. Hartley, "In defense of the eight-point algorithm," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 6, pp. 580–593, 1997.

[15] B. Bier, M. Unberath, J.-N. Zaech, J. Fotouhi, M. Armand, G. Osgood, N. Navab, and A. K. Maier, "X-ray-transform Invariant Anatomical Landmark Detection for Pelvic Trauma Surgery," in *Medical Image Computing and Computer-Assisted Intervention*, Springer, Ed., 2018, pp. 55–63.

Multi-Energy Monte Carlo Simulation of the Multi-Sources Static CT System

Yidi Yao, Liang Li, Zhiqiang Chen, Wuyang Liang, Xin Jin

***Abstract*— In this study, the multi-energy Monte Carlo simulation results based on the multi-sources static CT system are presented. Compared with current CT systems, the static CT system doesn't have the rotation gantry. Hence the artifacts introduced by the gantry vibration are reduced. In the simulation, six-energy projections are obtained using three different tube voltages and the 'sandwich'-structure detectors. An 1D-UNet neural network are applied to reduces scatter artifacts. Hybrid reconstruction and image-domain decomposition of a thorax CT phantom is presented. The final high-quality reconstruction and decomposition results show that it's feasible and effective for multi-energy imaging using the proposed static CT system.**

***Key Words*— Beam hardening, iterative reconstruction, material decomposition, multi-energy, neural network, 'sandwich' detector, scatter correction, static CT**

I. Introduction

Current CTs obtain projection data at different angles by rotating the gantry. With the development of the X-ray tube, detector material, electronics and electromechanical technologies, the scanning time for the medical imaging can be reduced to 1 second[1]. Therefore, the artifacts introduced by the movement of human body can be reduced largely. However, due to the mechanical structure of the gantry, it is difficult to further shorten the scanning time, which is necessary for clear cardiac imaging. On the other hand, the vibration accompanied with the high-speed rotation can introduce extra artifacts to imaging.

The complex mechanical structure can themselves cause problems. Because of the gantry, the volume of the CT is not suitable for hand luggage inspection, in which there is usually not enough room to place it. It needs to take great effort to ensure that the complex mechanical structures work properly.

Multi-energy imaging is important to material identification in both medical application and security application. However, conventional CT generates monochromatic images. Double-energy images can be reconstructed using the 'sandwich' detector. Spectral CT based on photon-counting detector could be a viable option in the future. But now the detector is too small for practical application[2].

Inspired by the drawbacks of the current CT system, our team propose a new CT system, the multi-sources static CT system. This system uses three line-array sources instead of the gantry to generate projection data from different views. Therefore, both the sources and detectors are static during the scanning, and the vibrations are eliminated to the greatest extent. Scatter correction based on neural network is applied in the reconstruction in order to correct the artifacts caused by the special geometric design. Three different X-ray tube voltages, 140kV, 110kV and 80kV, are set during the Monte Carlo simulation scanning. Hybrid reconstruction and image-domain decomposition results of a thorax CT phantom are presented. The final high-quality reconstruction and decomposition results show that it's feasible for multi-energy imaging using the proposed static CT system.

In the following, detailed algorithms, including CT system design, artifacts correction, reconstruction and decomposition, are described in Section 2. Monte Carlo simulation, reconstruction, and decomposition results are showed in Section 3. Analysis and further outlook are discussed in Section 4.

II. Methods

A. Static CT Design

In the static CT system, three sources-detectors planes in different directions are placed side by side. The diagrams of the system are showed in Figure 1.

In each sources-detectors plane, one line-array sources and three flat panel double-layer detectors are set. The length of the source array is 120cm. 128 sources are arranged in the array with equal distance. A 0.2 mm iron filter is placed in front of each source. The detector panels that opposite to the source array has 362×4 pixel, and the detector panels on the side have 315×4 pixels. The size of each pixel is 3 mm × 3 mm. Detector materials of both layers are GOS, and the density is 7.32 g/cm^3. The thickness of the first layer, or the low-energy layer, is 0.25 mm, and the thickness of the second layer is 1.5 mm. Between the two layers is a 0.6 mm copper filter.

The line-array sources and three detector panels are placed along the four sides of a square as the diagram shows. The centers of all the planes coincide in the horizontal direction.

Manuscript received January 16, 2020. This work was supported by the grant from NSFC 11525521 and NKRDPC 2017YFC0109103

Yidi Yao, Liang Li and Zhiqiang Chen are with the Department of Engineering Physics, Tsinghua University, Beijing, 100084, China & Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education, 100084, China.

Wuyang Liang, Xin Jin are with Nuctech Company Limited, Beijing, China

Corresponding author: czq@tsinghua.edu.cn; lliang@tsinghua.edu.cn

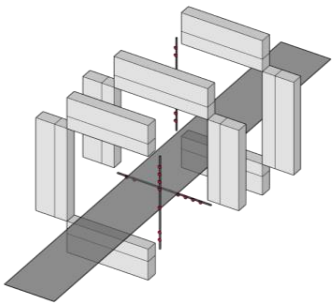


Fig. 1. The diagram of the static CT system

B. Monte Carlo Simulation

A Monte Carlo simulation for the static CT system is performed. A modified Forbid thorax phantom is applied in the simulation. The length of the long axis is 50 cm. The materials for different part of the phantom are chose from ICRU 46[3], and the cross section of the phantom at position where the ordinate is zero is shown in Figure 2. Different kinds of body tissue materials with extremely similar components, including adipose, bone, muscle and blood, are used. The detailed material types, their location and their densities are shown in Table 1.

The X-ray tube voltage is set to 140 kV, 110 kV, and 80 kV for three line-array sources respectively. The current of the tube is 25 mA. The exposure time is 1.5 ms for 140 kV and 110 kV, and 4.5 ms for 80kV. The Geant4 toolkit[4] is used to execute the Monte Carlo simulation. 28 servers run the simulation program in parallel. Each server has two Intel Xeon Gold 6135 CPU, and each CPU has 24 cores. The total running time is about 58 hours.

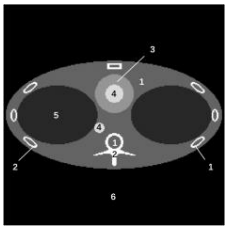


Fig. 2. The thorax phantom used in simulation

TABLE I
TYPES AND DENSITIES OF THE SELECTED MATERIALS

Number	Material Types	Density
1	Adipose	0.95g/cm^3
2	Bone	1.92g/cm^3
3	Muscle	1.06g/cm^3
4	Iodized Blood（95% Blood + 5% Iodine）	1.05g/cm^3
5	Lung Tissue	0.26g/cm^3
6	Air	1.3×10–3g/cm^3

C. Scatter Correction Based on 1D U-Net convolution network

The special CT design increases X-ray angle distribution, which aggravates the effects of scatter. Conventional collimator-based methods are invalid for this CT system.

In recent years, deep learning has become a powerful tool in imaging processing. In field of CT reconstruction, ScatterNet[5] and DSE[6] have been proposed to correct scatter artifacts. Inspired by them, we proposed the 1D U-Net convolution network for scatter correction in projection domain. The network architecture is showed in Figure 3.

Unlike common projection-domain network which inputs and outputs are 2D data, our proposed network uses 1D data. Each sample is a vector of the 2D projection, which represents the data emitted from one X-ray source. Projection data without scatter, which can be obtained from Monte Carlo simulation, are outputs of the network. Projection data considering scatter effect are inputs. Because different sources don't emit photons at the same time, the 1D network model is reasonable. Compared to 2D network, the 1D network uses 1D convolution instead of 2D convolution, which dramatically reduces the computing time. Moreover, replacing 2D data with 1D increases the number training samples.

Balancing the calculation complexity and the generation ability of the network, the depth is set to 7. After each down-sampling layer, two convolution layers follow. The size of the convolution kernel is 1×3. The activation function is ReLU. Batch normalization is applied. One eighth of the complete data is used for training to avoid overfitting. Adam algorithm is used for backpropagation. The start learning rate is 0.01 and exponential decay is applied. Batch size is 4. The total training epoch is 300, which is enough for convergence. Low-energy projection and high-energy projection are trained separately.

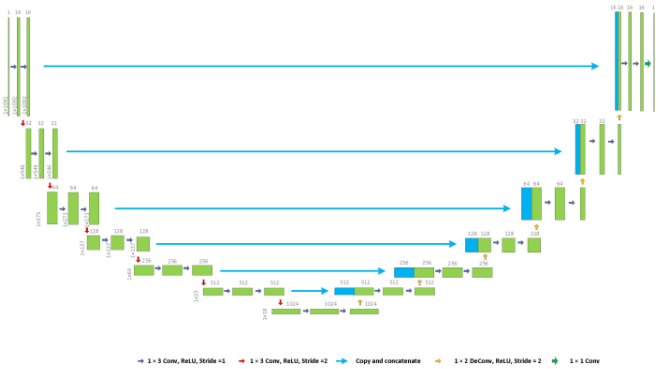


Fig. 3. The architecture of the neural network

D. Hybrid Reconstruction

In the reconstruction, SART algorithms and statistical weights are applied. Ten total variation (TV), and one low-rank constraints in the spectrum domain are performed after each iterative update. The total iteration time is 30 for each reconstruction.

As stated in Section 2B, the projection data of each plane corresponds to only one spectrum. Due to the requirement of data completeness for reconstruction, it is not feasible to reconstruct high-quality images using existing data. The data of the other two planes are needed.

Here we propose a method based on a pre-calibrated line integral curve. A phantom including similar material types are simulated for all three spectrums and all three planes. All the

points of the projection data are mapped onto a two-dimensional coordinate system. The abscissa and ordinate correspond to the line integral of each projection data point under different spectrum. A curve is fitted based on these points, and the corresponding relationship between the values of each projection data point under different spectrum is obtained. The example of the curves is shown in Figure 4. Both the low-energy and the high-energy curves fitted well. By finding the corresponding data points from projection of other spectrums, the missing projection data points can be obtained from the curves.

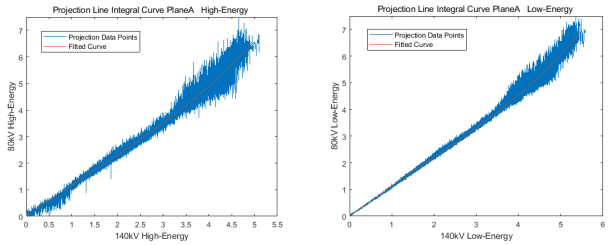


Fig. 4. The projection line integral curve of Plane A. The left image is the high-energy, and the right one is the low-energy. The abscissa is of 140kV, and the ordinate is of 80kV.

E. Image-domain Material Decomposition

The image-domain decomposition can be expressed in the form of matrix:

$$X = M \cdot f \tag{1}$$

X is the reconstructed spectral CT images. f is the decomposition result. M is the matrix of basis materials. Affected by noise, the pseudo-inverse is not an ideal solution. Here we use optimization method to solve it. Equation (1) can be rewritten as:

$$f = \arg\min_{f} \left(M \cdot f - X\right)^T V^{-1} \left(M \cdot f - X\right) \tag{2}$$

The volume conservation constraint and the non-negative constraint are added. The interior-point algorithm is applied to sovle the optimization problem.

III. Results

A. Reconstruction Results

In order to illustrate the effectiveness of our proposed methods, the reconstruction results without scatter correction and that using full-view projection data are presented for comparison. The reconstruction results are shown in Figure 5. The low-energy reconstruction results without scatter correction show obvious cup artifacts. Influenced by that, the values of adipose area are not the same, and the values of the iodine blood are lower and inaccurate. The image quality of the reconstructions with scatter correction is much better.

Comparing reconstructions of different spectrums, the values of the attenuation coefficient increases, and the artifacts are more evident as the maximum energy of the spectrum increases. Comparing the reconstructions of the same spectrum, the low-energy reconstructions have higher attenuation coefficient

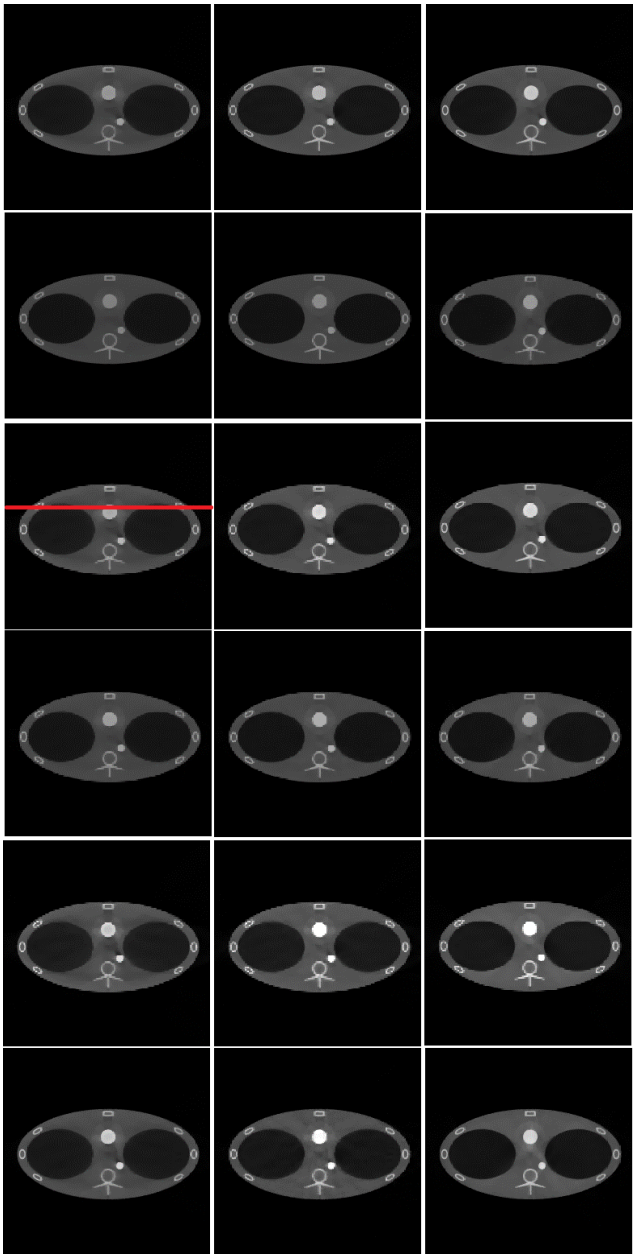


Fig. 5. The reconstruction results of different algorithms and spectrums. Columns, from left to right: Complete data without scatter correction, complete data with scatter correction, and incomplete data with scatter correction using hybrid reconstruction. Rows, from top to bottom: 140kV low-energy, 140kV high-energy, 110kV low-energy, 110kV high-energy, 80kV low-energy, 80kV high-energy.

than the high-energy ones. The above results are consistent with the physical laws.

In theory, the reconstructions using the complete data with scatter correction have the best image quality. From the comparison of the images between the second column and the third column, we can see that they have similar image quality. This illustrate the effectiveness of the hybrid reconstruction, which are proved in the decomposition results, too.

B. Decomposition Results

The decomposition results are shown in Figure 6. The results

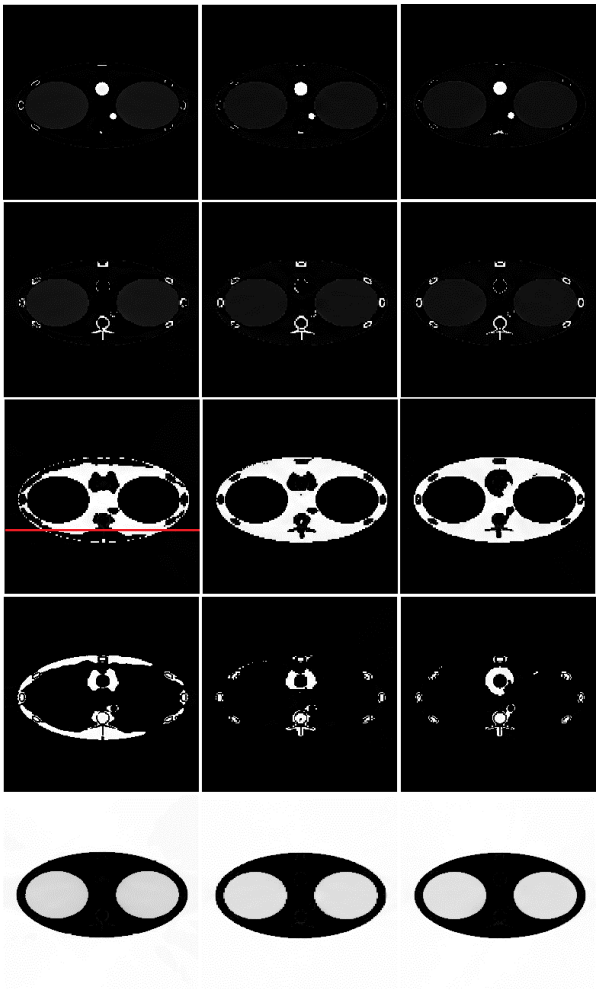


Fig. 6. The decomposition results of different algorithms and spectrums. Columns, from left to right: Complete data without scatter correction, complete data with scatter correction, and incomplete data with scatter correction using hybrid reconstruction. Rows, from top to bottom: adipose, bone, muscle, iodized blood and air.

using data without scatter correction have the worst image quality. Affected by scatter artifacts, part of the adipose area is misclassified to the muscle image, and more bone area is misclassified to the iodine blood image. These misclassifications influence the accuracy of the decomposition severely.

Both complete-data results and incomplete-data results using scatter correction are much better. All the materials are classified to their corresponding images approximately, though there is a little inaccurate area. The decompositions of the hybrid reconstruction have close image quality to that of the complete data, which illustrates the effectiveness of the hybrid reconstruction algorithm.

C. Quantitative Comparison

The quantitative comparisons of the reconstructions and decomposition results are shown in Figure 7. The values of the reconstruction results without scatter correction are much lower than the other two, caused by the cup artifacts. In the decomposition results, there are great differences between the values without scatter correction and that with scatter correction at specific pixels. This is caused by the misclassification of some adipose area in the non-scatter correction results. However, both the complete data results with scatter correction and the incomplete data results using hybrid reconstruction have similar section lines, which further proves the effectiveness of the hybrid reconstruction method.

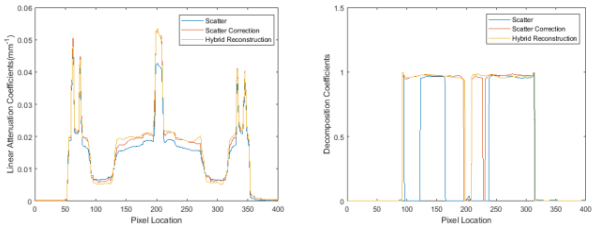


Fig. 7. Comparison of the results. The left image shows the section lines of the 110kV low-energy reconstructions. The right image shows the section lines of the adipose decompositions. The position of the section lines is shown in Figure 7 respectively. The blue line represents the complete data result without scatter correction. The red line represents the complete data result with scatter correction. The yellow line represents the incomplete data results with scatter correction using hybrid reconstruction.

IV. Discussion and Conclusion

In this paper, we proposed the multi-sources static CT design and presented its Monte Carlo simulation results. Scatter correction using 1D neural network and hybrid reconstruction algorithm are applied to obtain high-quality, multi-energy images. The presented reconstruction and decomposition results shows the feasibility of our proposed design and method.

However, there are some minor problems need to be addressed. The angular distribution of the emitted photons and the different incident angles of different ray cause inconsistency of the spectrums, which introduces extra artifacts in the reconstructions. The computing speed of Monte Carlo simulation is too slow, which affects the acquisition of the simulation data. The above problems will be solved next. The porotype of the static CT system is under test, and the experiment results will be showed in the future.

References

[1] J. Rydberg, K. A. Buckwalter, K. S. Caldemeyer, M. D. Phillips, D. J. Conces Jr, A. M. Aisen, *et al.*, "Multisection CT: scanning techniques and clinical applications," *Radiographics*, vol. 20, pp. 1787-1806, 2000.

[2] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: technical principles and clinical prospects," *Radiology*, vol. 289, pp. 293-312, 2018.

[3] D. R. White, R. V. Griffith, and I. J. Wilson, "Report 46," *Journal of the International Commission on Radiation Units and Measurements*, vol. os24, pp. NP-NP, 1992.

[4] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, *et al.*, "GEANT4—a simulation toolkit," *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, pp. 250-303, 2003.

[5] D. C. Hansen, G. Landry, F. Kamp, M. Li, C. Belka, K. Parodi, *et al.*, "ScatterNet: A convolutional neural network for cone-beam CT intensity correction," *Medical physics*, vol. 45, pp. 4916-4926, 2018.

[6] J. Maier, Y. Berker, S. Sawall, and M. Kachelrieß, "Deep scatter estimation (DSE): feasibility of using a deep convolutional neural network for real-time x-ray scatter prediction in cone-beam CT," in *Medical imaging 2018: physics of medical imaging*, 2018, p. 105731L.

A one-step algorithm for spectral CT with an application on multi-source inverse geometry

Frédéric Jolivet, Clarisse Fournier, Jérôme Lesaint and Andrea Brambilla.

***Abstract*—Spectral computerized tomography (Spectral CT) is an imaging technique which uses the spectral information of the attenuated X-ray beam. Energy-resolved photon-counting detector is a promising technique for improved spectral CT imaging and allows to obtain material selective images. While energy-resolved photon-counting detectors can have good spectral resolution (with many energy bins), the size of these detectors is often limited. For their part Multi-Source Inverse Geometry CT (MS-IGCT) architectures allow the use of a smaller detector than conventional Cone Beam CT architecture (CBCT). In a previous work, we have proposed a one-step reconstruction algorithm validated with spectral data from classical Cone Beam CT architecture. In this work we propose to adapt this one-step method on spectral data from MS-IGCT architecture. From noisy simulated data, we compare the proposed one-step method with two two-step decomposition methods.**

***Index Terms*—Multi-source Inverse Geometry CT, Spectral CT, Iterative image reconstruction, Material decomposition.**

I. Introduction

Spectral CT with energy-resolved photon-counting detectors gives the possibility to record the multi-energy data in a single acquisition [1]. One interest of spectral CT for medical x-ray imaging is to decompose the object onto some physical or materials basis [2]. Several approaches are proposed to tackle the decomposition problem. Two-step approaches first estimate multi-material decomposed sinograms from multi-energies sinograms. Then material specific images are reconstructed from the multi-material decomposed sinograms [3], [4]. One-step methods propose to tackle the decomposition problem in a one-step inversion, i.e. estimate multi-material reconstructions in the image domain from multi-energies sinograms [5], [6]. In a conventional CBCT architecture, the field of view of the reconstructed maps is directly limited by the size of the energy resolved photon counting detector. For their part multi-source inverse geometry CT (MS-IGCT) architectures (see Fig. 1 for example) allow the use of a smaller detector than conventional Cone Beam CT architecture [7]. Compared to conventional CBCT, MS-IGCT (with a small detector) also reduces scattered radiation, which should help improve material decomposition results. In previous works, we proposed an iterative tomographic reconstruction algorithm for MS-IGCT with an energy-integrating X-ray detector [8], and one-step method using photon-counting detectors for conventional CBCT architecture [9]. In this work, we adapt the one-step method [9] for multi-source inverse geometry CT and compare multi-material reconstructions obtained with the proposed method with those of two-step methods. The two-step methods estimate decomposed sinograms with a Maximum Likelihood Estimator (MLE), and from these decomposed sinograms the first two-step method uses Filtered Back-Projection (FBP) to obtain the multi-material reconstructions (in the image domain) whereas the second one uses a regularized tomographic reconstruction [8].

F. Jolivet is with the Univ. Grenoble Alpes, CEA, LETI, F-38000 Grenoble, France. E-mail: frederic.jolivet18@gmail.com .

J. Lesaint is with CEA, CTREG, DGDO, Bouguenais, France. (email : jerome.lesaint@cea.fr)

C. Fournier and A. Brambilla are with the Univ. Grenoble Alpes, CEA, LETI, F-38000 Grenoble, France. E-mail: firstname.lastname@cea.fr

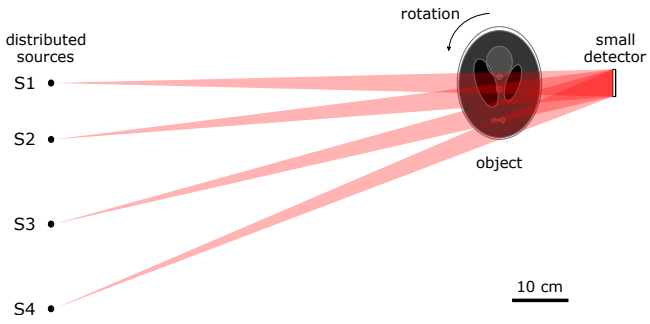


Fig. 1. Multi-Source Inverse Geometry setup.

II. An Empirical imaging Model

In inverse problems approaches, the image formation model links the observed data to the imaged object according to the properties of the acquisition system and the unknown object. In spectral CT the energy-dependant attenuation coefficient is classically expressed as a linear combination of a few basis materials [2], [10]. This model combined with the forward model of spectral tomographic reconstruction (including the incident spectrum and the spectral response of the detector) leads to a non-convex optimization problem (in the inverse problems framework) [3], [6] and requires an accurate knowledge of the system response. In this work, we choose to use an empirical imaging model based on a linear fit of attenuation calibration data [4] and adapt it for multi-source inverse geometry architectures.

A calibration basis of two materials is used to determine the coefficients p that best fits measurements obtained with N combinations of material specific thicknesses. With multi-source inverse geometry, N calibrated measurements are acquired for the energy-bin c of the pixel k when the q-th source is active. These measurements are collected in the vector $\boldsymbol{s}_{q,k,c}$,

$$\boldsymbol{s}_{q,k,c} = \boldsymbol{H}\boldsymbol{p}_{q,k,c} + \boldsymbol{e}^{calib}_{q,k,c} \tag{1}$$

where $\boldsymbol{s}_{q,k,c} \in \mathbb{R}^N$ and $\boldsymbol{e}^{calib}_{q,k,c} \in \mathbb{R}^N$ are respectively the calibration data vector and the error vector, whereas $\boldsymbol{p}_{q,k,c}$ is

a vector of two order-1-polynomial coefficients (in this work the number of basis materials is equal to 2) and $\boldsymbol{H} \in \mathbb{R}^{N\times 2}$ is a matrix defined by,

$$\boldsymbol{H} = \begin{pmatrix} l_{1,1} & \cdots & l_{1,N} \\ l_{2,1} & \cdots & l_{2,N} \end{pmatrix}^T .$$

Considering all vectors $\boldsymbol{e}_{q,k,c}^{calib}$ as white gaussian noises, we determine the coefficients $\boldsymbol{p}_{q,k,c}$ of these polynomial functions with,

$$\hat{\boldsymbol{p}}_{q,k,c} = \underset{\boldsymbol{p}\in\mathbb{R}^2}{\operatorname{argmin}} \|\boldsymbol{H}\boldsymbol{p} - \boldsymbol{s}_{q,k,c}\|_2^2 \tag{2}$$

The optimization problem (2) leads to the following closed-form solution (see chap.4 of the book [11]),

$$\hat{\boldsymbol{p}}_{q,k,c} = \left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{s}_{q,k,c} \quad . \tag{3}$$

We use the coefficients $\hat{\boldsymbol{p}}_{q,k,c}$ to model the C energy bins attenuation measurements at pixel k and object rotation angle θ when the q-th source is active,

$$m_{q,k,\theta}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{M}_{q,k}\boldsymbol{\chi}_{q,k,\theta} \tag{4}$$

where $\boldsymbol{x}_1 \in \mathbb{R}^{N_x}$ (respectively $\boldsymbol{x}_2 \in \mathbb{R}^{N_x}$) is the material-specific sinogram for the material 1 (respectively for the material 2). The matrix $\boldsymbol{M}_{q,k} \in \mathbb{R}^{C\times 2}$ and the vector $\boldsymbol{\chi}_{q,k,\theta} \in \mathbb{R}^2$ are defined by,

$$\boldsymbol{M}_{q,k} = \begin{pmatrix} \hat{\boldsymbol{p}}_{q,k,1} & \cdots & \hat{\boldsymbol{p}}_{q,k,C} \end{pmatrix}^T \quad \text{and} \quad \boldsymbol{\chi}_{q,k,\theta} = \begin{pmatrix} [\boldsymbol{x}_1]_{q,k,\theta} \\ [\boldsymbol{x}_2]_{q,k,\theta} \end{pmatrix},$$

where $[\boldsymbol{x}_\eta]_{q,k,\theta} = [\boldsymbol{A}\boldsymbol{z}_\eta]_{q,k,\theta}$ with $\boldsymbol{A}$ the forward projection matrix and $\boldsymbol{z}_\eta$ is the vector of N_z unknowns in the image domain for the material η.
This leads to the following data formation model,

$$\boldsymbol{d}_{q,k,\theta} = m_{q,k,\theta}(\boldsymbol{x}_1, \boldsymbol{x}_2) + \boldsymbol{e}_{q,k,\theta} \tag{5}$$

where $\boldsymbol{d}_{q,k,\theta} \in \mathbb{R}^C$ is the data vector for the k-th pixel of the θ-th object rotation angle when the q-th source is active and $\boldsymbol{e}_{q,k,\theta} \in \mathbb{R}^C$ is the error vector including different noises (detection noise, modeling noise...).
Note : In conventional CBCT, it is crucial to have pixel-specific model (and therefore a pixel-specific calibration) to take into account the non-uniform response between each pixels detector. In MS-IGCT, we must also take into account the variation between each sources.

III. The proposed one-step method

Considering all error vectors $\boldsymbol{e}_{q,k,\theta}$ in (5) are Gaussian noises, the data fidelity term can be formulated as the negative log-likelihood,

$$\mathcal{F}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{q=1}^{Q}\sum_{k=1}^{K}\sum_{\theta=1}^{\Theta} \|m_{q,k,\theta}(\boldsymbol{x}_1, \boldsymbol{x}_2) - \boldsymbol{d}_{q,k,\theta}\|^2_{\boldsymbol{W}_{q,k}} \tag{6}$$

with $\boldsymbol{W}_{q,k}$ the inverse of the covariance matrix of the noise for the k-th pixel when the q-th source is active (in our case $\boldsymbol{W}_{q,k}$ is diagonal because we assume a non correlated noise), K is the number of detector pixels, Θ is the number of object rotation angles and Q is the number of sources. This inverse problem is ill-posed and ill-conditioned. For this reason it is necessary to introduce prior information in order to restrict ambiguity for the inversion. In inverse problems framework, these prior information can be termed as a feasibility set (bound constraints,...) and/or a regularization function to favor specific properties of the spatial distribution of the reconstructed object. In our case we assume an isotropic total variation regularization which promotes piecewise-constant objects [12], [13] on reconstruction objects $\{\boldsymbol{z}_1, \boldsymbol{z}_2\}$, and a feasible set Ω. The proposed one-step method for MS-IGCT can be formulated as an optimization problem,

$$\hat{\boldsymbol{z}} \in \underset{\boldsymbol{z}}{\operatorname{argmin}}\ \mathcal{F}(\boldsymbol{A}\boldsymbol{z}_1, \boldsymbol{A}\boldsymbol{z}_2) + \mathcal{R}(\boldsymbol{z}_1, \boldsymbol{z}_2) \tag{7}$$

where $\boldsymbol{z} = \begin{pmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{pmatrix}$ and $\mathcal{R}$ is a regularization term which introduces an isotropic total variation constraint in the image domain and could restrict $\boldsymbol{z}$ values to a feasible set Ω (a non-negativity constraint for instance). Thus $\mathcal{R}$ is defined by

$$\mathcal{R}(\boldsymbol{z}_1, \boldsymbol{z}_2) = \lambda\sum_{\eta=1}^{2} \|\nabla\boldsymbol{z}_\eta\|_{2,1} + i_{\boldsymbol{z}\in\Omega}(\boldsymbol{z}_1, \boldsymbol{z}_2) \quad , \tag{8}$$

where $i_{\boldsymbol{z}\in\Omega}$ is the indicator function which restrict $\boldsymbol{z}$ values to the set Ω, $\lambda \in \mathbb{R}^+$ is an hyperparameter that balances the weight of the regularization term with the data fidelity term and $\|.\|_{2,1}$ is the (2,1)-mixed norm $\ell_{2,1}$.

IV. Proposed iterative reconstruction algorithm

In the previous section we have formulated the inverse problem as the large scale optimization problem (7) which is non-smooth (due to the regularization term). Furthermore, the data fidelity and the regularization term are not in the same domain, respectively in the projection domain and the image domain. For these reasons the optimization problem (7) is challenging. In order to limit the time complexity of the reconstruction, it is important to solve data fidelity term in the projection domain. We therefore choose to use a variable splitting strategy such that at convergence $\boldsymbol{x}_\eta = \boldsymbol{A}\boldsymbol{z}_\eta$. We reformulate the optimization problem (7) as the following constrained optimization problem,

$$\hat{\boldsymbol{z}} \in \underset{\substack{\boldsymbol{x},\boldsymbol{z} \\ \text{s.t. } \boldsymbol{x}=\tilde{\boldsymbol{A}}\boldsymbol{z}}}{\operatorname{argmin}}\ \mathcal{F}(\boldsymbol{x}_1, \boldsymbol{x}_2) + \mathcal{R}(\boldsymbol{z}_1, \boldsymbol{z}_2) \tag{9}$$

where $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}$ and $\tilde{\boldsymbol{A}} = \boldsymbol{I}_2 \otimes \boldsymbol{A}$ with $\boldsymbol{I}_2$ is the identity matrix and $\otimes$ is the Kronecker product. The function $\mathcal{F}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is smooth convex, and $\mathcal{R}(\boldsymbol{z}_1, \boldsymbol{z}_2)$ is a non-smooth and non-linear function. To handle the constrained optimization problem (9), we can form the augmented Lagrangian $\mathcal{L}_\rho$ and search for a saddle point,

$$\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) = \mathcal{F}(\boldsymbol{x}_1, \boldsymbol{x}_2) + \mathcal{R}(\boldsymbol{z}_1, \boldsymbol{z}_2) + \rho\sum_{\eta=1}^{2}\left(\|\boldsymbol{x}_\eta - \boldsymbol{A}\boldsymbol{z}_\eta + \boldsymbol{u}_\eta\|_2^2 - \|\boldsymbol{u}_\eta\|_2^2\right) \tag{10}$$

where $\boldsymbol{u}_{\eta=1,2} \in \mathbb{R}^{N_x}$ are the scaled dual variables, and $\rho \in \mathbb{R}^+$ is the augmented Lagrangian penalization parameter [14]. To find a saddle point we can resolve alternatively

a minimization with respect to the primal variables and a maximization with respect to the dual variables,

$$\left\{\boldsymbol{x}^{(j+1)}, \boldsymbol{z}^{(j+1)}\right\} \in \underset{\boldsymbol{x},\boldsymbol{z}}{\operatorname{argmin}}\ \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}^{(j)})$$
$$\text{dual update}: \boldsymbol{u}_\eta^{(j+1)} = \boldsymbol{u}_\eta^{(j)} + \boldsymbol{x}_\eta^{(j+1)} - \boldsymbol{A}\boldsymbol{z}_\eta^{(j+1)}$$

where $\eta \in \{1, 2\}$ and (j) corresponds to the iteration number. This optimization problem can be performed by the Alternating Direction Method of Multipliers (ADMM) [14],

$$\boldsymbol{x}^{(j+1)} = \underset{\boldsymbol{x}}{\operatorname{argmin}}\ \mathcal{L}_\rho\left(\boldsymbol{x}, \boldsymbol{z}^{(j)}, \boldsymbol{u}^{(j)}\right) \tag{11}$$
$$\boldsymbol{z}^{(j+1)} = \underset{\boldsymbol{z}}{\operatorname{argmin}}\ \mathcal{L}_\rho\left(\boldsymbol{x}^{(j+1)}, \boldsymbol{z}, \boldsymbol{u}^{(j)}\right) \tag{12}$$
$$\boldsymbol{u}_\eta^{(j+1)} = \boldsymbol{u}_\eta^{(j)} + \boldsymbol{x}_\eta^{(j+1)} - \boldsymbol{A}\boldsymbol{z}_\eta^{(j+1)}, \forall\ \eta \in \{1, 2\}\,.$$

Since the projection matrix $\boldsymbol{A}$ is a linear operator, the minimization problem (12) is a convex/non-smooth large-scale optimization problem solved with the MS-IGCT reconstruction method [8]. On the other hand, the minimization problem (11) is a convex/smooth minimization problem which is separable and can thus be solved independently pixel by pixel and source by source with the following closed-form expression,

$$\widehat{\boldsymbol{\chi}}_{q,k,\theta}^{(j+1)} = (\boldsymbol{M}_{q,k}^T \boldsymbol{W}_{q,k} \boldsymbol{M}_{q,k} + \rho \boldsymbol{I}_2)^{-1} (\boldsymbol{M}_{q,k}^T \boldsymbol{W}_{q,k} \boldsymbol{d}_{q,k,\theta} + \rho \boldsymbol{v}_{q,k,\theta}^{(j)}),$$

with $\boldsymbol{I}_2$ the identity matrix and $\boldsymbol{v}_{q,k,\theta} = \begin{pmatrix} [\boldsymbol{A}\boldsymbol{z}_1 - \boldsymbol{u}_1]_{q,k,\theta} \\ [\boldsymbol{A}\boldsymbol{z}_2 - \boldsymbol{u}_2]_{q,k,\theta} \end{pmatrix}$. The closed-form solution of (11) is the key point of the proposed one-step method.

V. Experiments & Results

A. Simulations

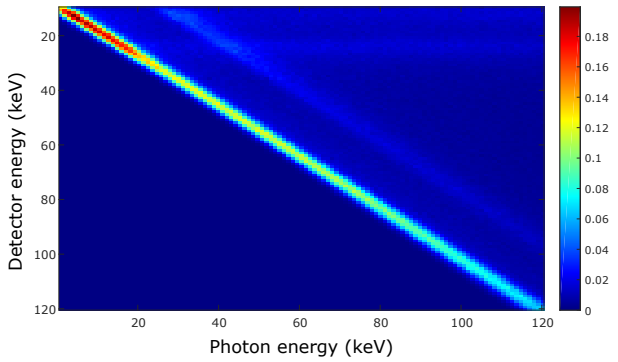


Fig. 2. Detector Response Matrix for the simulated CdTe detector.

Fig. 1 is a realistic illustration of the simulated MS-IGCT architecture. We simulate 4 distributed sources shifted from 0 to 40 cm (0/10/25/40cm). The source-to-detector distance is 150 cm whereas the source-to-isocenter distance is 125 cm. The object is a Shepp-Logan Phantom of 18.4×13.8 cm, containing different proportions of water and cortical bone materials. The simulated photon counting detector is a line of 50 pixels with a pitch of $1 \times 1\ mm^2$ (i.e. detector size ~ 5 cm). The detector is simulated with one hundred and eleven energy bins from 10keV to 120keV and the realistic Detector Response Matrix (DRM) shown in Fig. 2. The simulated source was set to 140kV with 0.1 cm aluminum filtration at 0.1 mA.s per projection which result of an x-ray exposure of 5×10^6 photons per pixel. Attenuation coefficients of water and cortical bone come from the computer program XMuDat [15]. Energy-resolved photon counting data are corrupted by a Poissonnian noise. These simulated data are represented Fig.3. A two basis material calibration of water and cortical bone was simulated with 15 thicknesses ranging from 0 to 17 cm for the water, and 15 thicknesses ranging from 0 to 6 cm for the cortical bone.

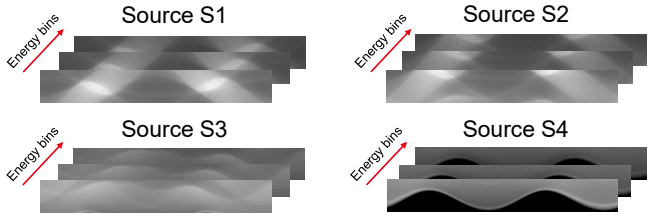


Fig. 3. Truncated simulated sinograms of the MS-IGCT architecture (Fig.1).

B. Results

Fig. 4 shows a comparison between 3 methods : the proposed one-step method (fourth column) with a non-negativity constraint. **Method 1** (second column) : A two-step method based on a Maximum Likelihood Estimation (MLE) of materials sinograms $\boldsymbol{x}$ (which consists in minimizing the data fidelity expressed in (6)) followed by a filtered back-projection step. **Method 2** (third column) : A two-step method based on the same first step but followed by a regularized tomographic reconstruction method with TV regularization adapted for MS-IGCT [8]. For the last method, regularization hyperparameters of the second step are tuned automatically to maximize reconstruction's Peak Signal-to-Noise Ratio (PSNR). All these reconstructions are obtained from simulated data described in Sec. V-A and can be compared with ground truth (second column). Results in Fig. 4 show that the proposed one-step method gives reconstructions with a better PSNR than the two-step methods (see Tab.I). Regarding the proposed one-step method, the number of iterations needed to solve (12) is another key parameter to maximize the ratio reconstruction image quality vs. computation time. In our case, 15 iterations of the regularized reconstruction for MS-IGCT [8] give satisfying reconstructions. To finish we applied 50 iterations of the ADMM method. The proposed method is implemented on GPU to drastically reduce the computational time.

PSNR	Method 1	Method 2	Proposed method
Bone	18.05 dB	23.31 dB	29.66 dB
Water	11.15 dB	22.25 dB	23.30 dB

TABLE I
Reconstruction's Peak-Signal-to-Noise Ratio.

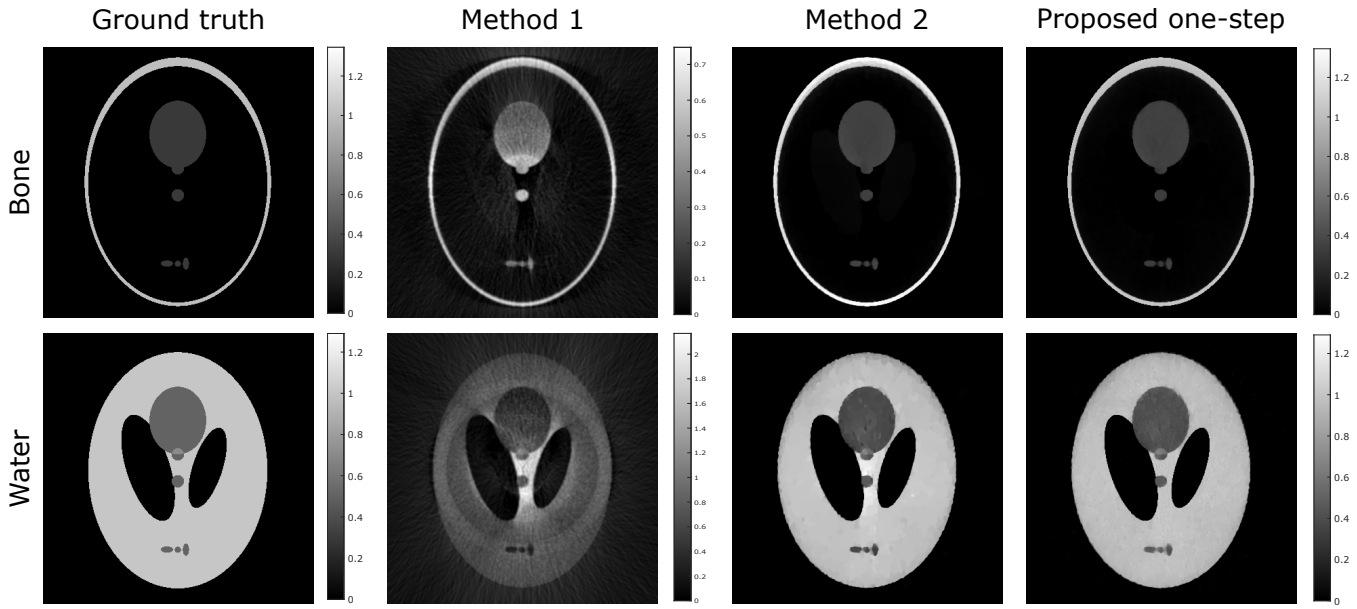


Fig. 4. Comparison of bone (top) and water (bottom) reconstructions in the image domain with spectral data from MS-IGCT architecture with 4 distributed sources and a small detector (5 cm). The ground truth (first column), and reconstructions obtained with the two-step method based on MLE decomposed sinograms followed by regularized iterative reconstructions (third column) and reconstructions obtained with the proposed one-step method (fourth column). With an other corlorbar, reconstructions obtained with the two-step based on MLE decomposed sinograms followed by filtered back-projections (second column).

VI. Conclusion & Perspectives

We have proposed a one-step method which has been applied on spectral data for a multi-source inverse geometry CT. This one step method used an empirical model based on calibration data, and an alternating optimization strategy. The data fidelity reconstruction step is solved with a closed-formed expression, which drastically reduces the computation time of the material decomposition. The total-variation constraint and bound constraints are introduced by the iterative regularized reconstruction method for multi-source inverse geometry [8]. The present work shows that the proposed method reduces the noise for a bi-material reconstruction. Finally, this work opens up interesting perspectives for spectral CT reconstruction with a small spectral detector and a flat panel of sources. These preliminary - but promising - results, obtained on noisy simulated data, must now be validated on experimental spectral data.

Acknowledgment

We acknowledge the financial support of the Cross-Disciplinary Program on Instrumentation and Detection of CEA, the French Alternative Energies and Atomic Energy Commission.

References

[1] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical physics*, vol. 40, no. 10, 2013.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerised tomography," *Physics in Medicine & Biology*, vol. 21, no. 5, p. 733, 1976.

[3] N. Ducros, J. F. P.-J. Abascal, B. Sixou, S. Rit, and F. Peyrin, "Regularization of nonlinear decomposition of spectral x-ray projection images," *Medical physics*, vol. 44, no. 9, pp. e174–e187, 2017.

[4] R. E. Alvarez, "Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis," *Medical physics*, vol. 38, no. 5, pp. 2324–2334, 2011.

[5] C. Mory, B. Sixou, S. Si-Mohamed, L. Boussel, and S. Rit, "Comparison of five one-step reconstruction algorithms for spectral ct," *Physics in Medicine & Biology*, vol. 63, no. 23, p. 235001, 2018.

[6] R. F. Barber, E. Y. Sidky, T. G. Schmidt, and X. Pan, "An algorithm for constrained one-step inversion of spectral ct data," *Physics in Medicine & Biology*, vol. 61, no. 10, p. 3784, 2016.

[7] B. De Man, S. Basu, P. Fitzgerald, D. Harrison, M. Iatrou, K. Khare, J. LeBlanc, B. Senzig, C. Wilson, Z. Yin *et al.*, "Inverse geometry ct: The next-generation ct architecture?" in *2007 IEEE Nuclear Science Symposium Conference Record*, vol. 4. IEEE, 2007, pp. 2715–2716.

[8] F. Jolivet, C. Fournier, and A. Brambilla, "A fast gradient-based algorithm for image reconstruction in inverse geometry ct architecture with sparse distributed sources," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 11072. International Society for Optics and Photonics, 2019, p. 110721N.

[9] F. Jolivet, C. Fournier, M. Garcin, L. Zdeborová, and A. Brambilla, "An alternating projection-image domains algorithm for spectral ct," in *IEEE 17th International Symposium on Biomedical Imaging (ISBI 2020)*.

[10] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler *et al.*, "Experimental feasibility of multi-energy photon-counting k-edge imaging in pre-clinical computed tomography," *Physics in Medicine & Biology*, vol. 53, no. 15, p. 4031, 2008.

[11] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

[12] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[13] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical imaging and vision*, vol. 20, no. 1-2, pp. 89–97, 2004.

[14] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[15] R. Nowotny, "Xmudat: Photon attenuation data on pc," *IAEA Report IAEA-NDS*, vol. 195, 1998.

Deep Learning based Metal Inpainting in the Projection Domain using additional Neighboring Projection Information

Tristan M. Gottschalk, Björn W. Kreher and Andreas Maier

***Abstract*—Orthopedic surgical interventions oftentimes require the evaluation of the correct positioning of e.g. inserted screws or metal implants. Therefore, typically mobile C-arms are used in order to acquire X-ray images directly inside the operation room. Besides 2D X-ray projections, new C-arm systems also provide 3D reconstructions which can be used as additional source of information in cases in which the 2D projection images do not suffice. However, this additional information is limited, mainly due to metal artifacts. Thus, metal artifact reduction methods were developed to resolve these problems, but no generally accepted approaches have been found yet. This paper can show that a projection domain based inpainting method significantly reduces metal artifacts inside the corresponding reconstructed 3D volumes and that using neighboring projections, as additional input to the inpainting network, help to further increase the performance, especially in cases in which extremely dense metal objects are present inside the X-ray acquisitions.**

I. Introduction

Apart from the possibility to acquire 2D X-ray projection images, modern mobile C-arms are also capable of acquiring 3D reconstructions. Particularly during surgical trauma interventions, 3D reconstructions can provide valuable additional information that help to assess the correct positioning of e.g. inserted screws or implants, when a sole evaluation based on 2D projections is not reliable. Nevertheless, different image artifacts alter the reconstruction quality and thus limit the benefits. Most of these artifacts are caused by metal objects which induce data inconsistencies between the metal corrupted projections. To address these problems, various metal artifact reduction (MAR) methods were developed, but no generally accepted approaches have been found yet. Besides the classic MAR methods [1]–[4], more recently deep learning based approaches show promising results. Latest methods either try to remove the metal artifacts in image-domain [5], [6] by learning a destreaking or by processing the data in projection domain [7]–[9]. Projection domain-based approaches typically try to inpaint empty regions without information inside the sinogram after removing the so called metal trace. Based on the fact that the sinogram representation of the projection data only holds for the central slice in cone-beam geometry and that tackling the artifacts solely in the reconstructed volume does not solve the underlying problem of data inconsistencies, we are convinced that inpainting the metal regions inside the 2D X-ray projections is the most promising approach.

T. M. Gottschalk and A. Maier are with the Pattern Recognition Lab, Department of Computer Science, Friedrich–Alexander–Universität Erlangen–Nürnberg, Erlangen, Germany. T. M. Gottschalk is funded by the Erlangen Graduate School in Advanced Optical Technology (SAOT), Friedrich–Alexander–Universität Erlangen–Nürnberg, Erlangen, Germany. A. Maier is principal investigator at the SAOT. B. W. Kreher is employee of Siemens Healthcare GmbH, Forchheim, Germany. Email: tristan.gottschalk@fau.de

As presented in [10], we have already been able to show that the proposed network architecture can produce meaningful inpainting results when applied directly to the 2D projection images and only receiving the corresponding metal mask as additional input. Nonetheless, the approach seems to reach a limit when the projections contain areas with extremely dense metal objects. The inpainting result in these areas becomes blurred and thus also the corresponding reconstructed 3D volumes contain blurry regions. To overcome that behavior, we propose an advanced inpainting approach that uses adjacent projections as additional input to the network. In this way, the network can potentially collect important information from neighboring projections, which may be overlaid at the current projection alone.

This paper evaluates whether the proposed projection domain based inpainting method not only creates meaningful inpainting results in the 2D projection images but rather helps to reduce metal artifacts within the resulting 3D reconstructions in general. Additionally, it is investigated whether the inclusion of additional information, in form of neighboring projections, further improves the image quality of the 3D reconstruction.

II. Proposed Method

A. Data

In order to be able to evaluate the influence of additional information provided in the form of neighboring projections compared with the initial inpainting approach presented in [10], the same dataset, which is also explained in greater detail in [10], must be used for training, testing and validating the further refined inpainting method. The data consists exclusively of real data that was acquired using a Siemens Cios Spin C-arm system. By randomly placing metal objects onto human knee cadavers, 12 corresponding 3D X-ray Cone-Beam scans were acquired, each containing of 400 2D projection images with 976x976x1 pixels covering an angular range of 180 degrees plus the C-arm's fan angle. In addition, the corresponding reconstructed 3D volumes of 512x512x512 voxels are stored for each of these scans. Each volume covers a cube with edge length of 160mm. The metal objects have been placed on the surface (skin) of the knees in order to be able to remove them without any motion of the knee or the scanner between the two corresponding scans. Thus, we end

up with 4800 pairs of matching 2D projection images – each one time with and one time without metal – that can be used for training, testing and validating the network.

B. Network Architecture

The proposed network architecture is a modified U-Net [11]. The network performs a projection-wise inpainting as shown in Figure 1. The network consists of eight contracting and eight expanding blocks, each containing two convolutional layers using 3x3 kernels and ReLUs as activation functions. The contraction blocks use max-pooling layers with a 2x2 kernel in order to halve the width and heights of feature maps at each block. In contrast, the expanding blocks use upsampling layers and thus double the width and height of each feature map. Additionally, the amount of feature maps is doubled or halved each contraction block or expansion block respectively. Corresponding contraction and expansion blocks are linked with skip connections and concatenations. The input of the network consists of one or multiple grayscale input image(s) of size 976x976 (detailed explanation in section II-C2) in addition to the metal mask, which is for now calculated from the ground truth. Following the contraction part of the network, the image size condenses to 7x7 pixels with 1024 feature maps in the network's bottleneck. The final output of the network is one grayscale image of size 976x976x1 again.

C. Training Procedure

1) Preprocessing: The measured RAW Data at the detector of the C-arm is converted into line integral data by using Lambert-Beer Law [12]. Thus, the measured intensities (I) are normalized by the initially emitted intensity (I_0) from the X-ray source. Using that normalization, the integral over the attenuation coefficient μ along the path x on the X-ray beam is calculated as follows:

$$\int \mu(x)dx = -\ln\left(\frac{I}{I_0}\right). \tag{1}$$

Furthermore, the 12 acquired 3D scans are splitted into eight scans for training, two scans for validation and two scans for testing. Thus, 3200 projections can be used for training and 800 projections each for validation and testing respectively. As an additional step the metal mask is (for now) calculated for each projection by subtracting the metal corrupted projection image from its corresponding metal free ground truth projection and a subsequent binarization.

2) Training: In order to assess the influence of providing additional information to the network in form of neighboring projections, the explained network was trained twice, as can be seen in Figure 1 a) and b). During the first training the network had two input channels, the metal corrupted projection that should be inpainted, as well as its corresponding metal mask. During the second training ten additional neighboring projections were provided such that the network got 12 input channels. The neighboring projections are selected using a sliding window operation with the current projection as center. The sliding window has a window size of 11 with an angular increment of 1° such that the neighboring projections are

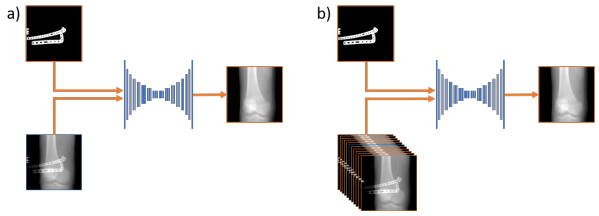


Fig. 1. a) shows the network's input w/o additional information, thus consisting of the current projection and its metal mask. b) shows the second training configuration concatenating the neighboring projections, the current projection as central channel and its metal mask.

selected at angular positions $\{\theta_i = \theta + i \mid i \in [-5°, 5°]\}$, with θ denoting the acquisition angle of the currently chosen projection. In both cases, the network was trained for 1000 epochs using Adam [13] as optimizer with an exponentially decaying learning rate applied every 100th epoch, starting with $1e^{-4}$, and mean squared error as loss function. Due to large image sizes and limited amount of memory, the batch size was chosen as 1.

D. Image Fusion

In order to preserve as much untouched data as possible, in addition to the network's outputted inpainted projection, a second so called "fused" result projection is calculated. The Image Fusion is calculated between the unprocessed input projection I_{in} and the network's inpainted output projection I_{out}:

$$I_{fusion} = I_{out} \circ M + I_{in} \circ (1 - M) \tag{2}$$

, where M denotes the metal mask, I_{fusion} the fused result image and $\circ$ the Hadamard product.

E. Evaluation

To exclusively evaluate the reduction of image artifacts present in the resulting volumes (no inpainting vs. inpainting w/o neighbors vs. inpainting w/ neighbors), it is necessary to quantify the volumes only in a metal-free region. Otherwise, solely the presence or absence of metal objects inside the metal-corrupted volume and inpainted volumes, would completely deteriorate validity of the comparison, thus distorting the evaluation. To solve this issue, the largest possible quadratic center crop (200x200 pixels) was chosen such that the cropped area does not include any metal objects for both test datasets, as is exemplary shown in Figure 2. Consequently, only the presence and severity of metal artifacts, which cover the whole image due to the underlying image formation process, thus penetrating the cropped region, are compared using the structured similarity index (SSIM) [14]. Besides the different inpainting methods, it is investigated whether the proposed Image Fusion helps the overall quality of the reconstructed volumes.

III. Results

In order to compare the influence of the projection domain based inpainting methods and the benefit of adding additional

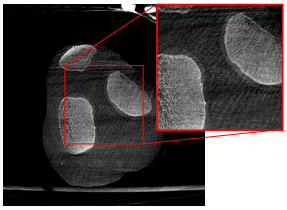


Fig. 2. Example of the quadratic center crop.

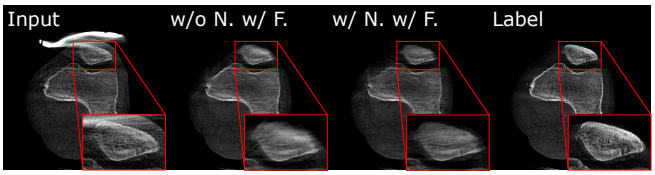


Fig. 3. Results for a slice containing dense metal objects. Including neighbor information (w/ N.) leads to a more distinct bone edge at the kneecap.

neighboring information to the network's input, we take a look at Table I. The quantitative evaluation of both test datasets (3D scans of 2 knees corrupted by metal objects), and thus a total of 800 projection images, shows that all average SSIM values of the different inpainting and fusion strategies are at least equal or above of those performing no inpainting at all. The highest SSIM of 0.826 is achieved using neighboring projections as additional information for the network and without performing the proposed Image Fusion. It is significantly higher than the SSIM of 0.763 reached when no inpainting is performed and also higher than the SSIM of 0.824 reached when not including additional information and not performing Image Fusion. Applying the proposed Image Fusion, the SSIM decreases to 0.763 without using additional neighboring projections and to 0.768 including the neighboring projections. Thus, adding additional information of the neighborhood to the network's input increases the SSIM in both cases whether Image Fusion is used or not. Taking a look at the results of the first and second test set separately, it is apparent that in both test cases using the additional neighboring information results in increased SSIMs.

TABLE I
THE AVERAGE SSIMS FOR THE DIFFERENT INPAINTING METHODS USING NEIGHBORING INFORMATION OR NOT AND IMAGE FUSION RESPECTIVELY

Test Set	Inpaint Config.	Fusion	SSIM
3D Scan (400 projs) of Knee 1 (1)	No Inpaint	-	0.759
	w/o Neigh.	No	0.824
	w/ Neigh.	**No**	**0.826**
	w/o Neigh.	Yes	0.759
	w/ Neigh.	**Yes**	**0.763**
3D Scan (400 projs) of Knee 2 (2)	No Inpaint	-	0.767
	w/o Neigh.	No	0.824
	w/ Neigh.	**No**	**0.825**
	w/o Neigh.	Yes	0.767
	w/ Neigh.	**Yes**	**0.772**
(1) + (2)	No Inpaint	-	0.763
	w/o Neigh.	No	0.824
	w/ Neigh.	**No**	**0.826**
	w/o Neigh.	Yes	0.763
	w/ Neigh.	**Yes**	**0.768**

Looking at the qualitative results presented in Figure 4 shows that all configurations of inpainting and fusion (cf. rows 2-5) reduce the amount and severity of metal artifacts present in the reconstructions. Using the fused projection images for reconstructing the volumes (cf. columns 2 + 3), it becomes visible that these volumes contain more high frequencies, thus creating an image impression, comparable to that of the metal- and artifact-free label volume (cf. col. 6). Furthermore, it can be seen that those slices that already initially do not include metal objects or artifacts, are not altered at all, when using the proposed Image Fusion (cf. row c)). Figure 3 presents a slice from the first test dataset that represents a case in which the kneecap is overlaid by a highly dense metal implant. Taking a closer look shows that including neighboring projections produces a reconstruction with a clearer and more distinct bone edge at the kneecap, while the omission of neighboring projections leads to a blurred one.

IV. Discussion and Conclusion

Summing up the quantitative and qualitative results, it becomes clear that performing a projection domain based inpainting method leads to a significant increase of image quality and consequently to a significant reduction of the amount and severity of metal artifacts in the evaluated cropped region. Furthermore, including additional information, in form of neighboring projections, seem to improve the performance of the network. Especially in cases in which parts of the projection images are superimposed by extremely dense metal objects and in which the previously proposed inpainting network [10] had difficulties, the incooperation of neighboring information creates more meaningful results, showing in e.g. more distinct bone edges as shown by the example of the kneecap (cf. Fig. 3). In order to use the proposed inpainting method in a complete MAR pipline, a deep learning based metal segmentation will be implemented. The metal masks are needed as 2D input for the inpainting network and as reconstructed 3D mask for the still needed Metal Injection, which inserts the deleted metal objects back into the artifact-reduced volume.

Although the quantitative results decrease when performing Image Fusion, the qualitative results (cf. Fig. 4) show that the overall image impression increases and is more similar to that of the label image. The decreased quantitative results are probably due to the influence of noise on the SSIM metric – we investigated that comparing two images with comparable amount of uncorrelated noise lead to a low SSIM value, whereas comparing a noisy image to a less noisy image lead to higher SSIM values. Thus, the decreased SSIMs, when using Image Fusion, is not directly translatable to less reduction of image artifacts but more likely to the presence of more uncorrelated noise. This assumption should be proven by suitable simulations, but is beyond the scope of this paper. Consequently, at least for our understanding, Image Fusion is preferable to use, due to its more natural image impression and more importantly in the interest of preserving as much unchanged original data as possible.

In order to solely assess the amount of metal artifact reduction, performing no or the proposed inpainting methods,

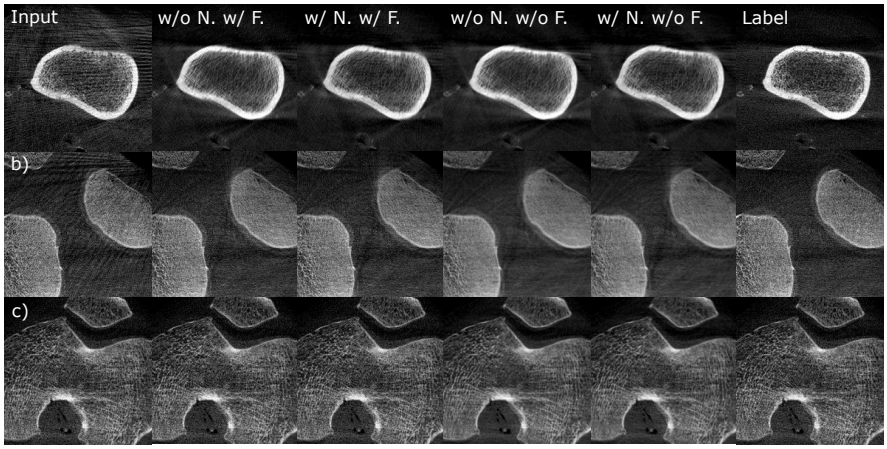


Fig. 4. Each row from a) to c) shows an identical center cropped slice (of the same two test datasets used for the quantitative evaluation shown in Table I), being reconstructed using different combinations of the proposed inpainting methods and Image Fusion. The first and last columns show the unprocessed metal corrupted volume (Input) and the metal- and artifact-free ground truth volume (Label). Columns 2 + 3 show the results achieved using Image Fusion without and with neighboring information. Columns 4 + 5 show the results achieved using no Image Fusion without and with neighboring information. Row c) shows results for a slice where already the initial slice does not contain any metal objects or artifacts.

it was necessary to center crop the volume slices as explained. Since the metal artifacts, which originate from the metal objects, extend over the whole image and thus also penetrate into the cropped region, the performed evaluation should give a meaningful intuition for this case study. However, the next step should be the implementation of the missing parts of a complete MAR pipeline, providing the possibility to evaluate the results on complete reconstructed volumes without cropping and also providing the possibility to compare the results with existing MAR methods.

Another important step will be to further improve the realism of the data by acquiring scans, where the metal is not only placed on the surface of the knee cadavers, but rather directly on and even inside the bones. In the course of acquiring such data, the variability of the data should be broaden by adding additional anatomical regions as well as additional metal objects, in order to further increase the robustness of the method. Furthermore, the use of augmentation strategies and simulation will be considered in order to increase the amount of data substantially. Nonetheless, we are certain that the current dataset, completely acquired from real knee cadavers, already provide a superior amount of realism over solely simulation based datasets, and thus providing a better generalization to real clinical cases.

Finally, as we were able to show that using neighboring projections as additional input to the network improves the image quality of the reconstructed volumes, it would be interesting to further investigate how different sampling patterns with different angular ranges of adjacent projections as well as different angular increments change the capabilities of the artifact reduction.

REFERENCES

[1] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Frequency split metal artifact reduction (FSMAR) in computed tomography," *Medical Physics*, vol. 39, no. 4, pp. 1904–1916, 2012. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3691902

[2] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess, "Normalized metal artifact reduction (NMAR) in computed tomography," *Medical Physics*, vol. 37, no. 10, pp. 5482–5493, 2010. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3484090

[3] D. Xinhui, Z. Li, X. Yongshun, C. Jianping, C. Zhiqiang, and X. Yuxiang, "Metal artifact reduction in CT images by sinogram TV inpainting," in *2008 IEEE Nuclear Science Symposium Conference Record*, Oct 2008, pp. 4175–4177.

[4] W. A. Kalender, R. Hebel, and J. Ebersberger, "Reduction of CT artifacts caused by metallic implants." *Radiology*, vol. 164, no. 2, pp. 576–577, 1987, pMID: 3602406. [Online]. Available: https://doi.org/10.1148/radiology.164.2.3602406

[5] L. Gjesteby, Q. Yang, Y. Xi, H. Shan, B. Claus, Y. Jin, B. D. Man, and G. Wang, "Deep learning methods for CT image-domain metal artifact reduction," in *Developments in X-Ray Tomography XI*, B. Müller and G. Wang, Eds., vol. 10391, International Society for Optics and Photonics. SPIE, 2017, pp. 147 – 152. [Online]. Available: https://doi.org/10.1117/12.2274427

[6] X. Huang, J. Wang, F. Tang, T. Zhong, and Y. Zhang, "Metal artifact reduction on cervical CT images by deep residual learning," *Biomedical engineering online*, vol. 17, no. 1, p. 175, 2018.

[7] M. U. Ghani and W. C. Karl, "Deep Learning Based Sinogram Correction for Metal Artifact Reduction," *Electronic Imaging*, vol. 2018, no. 15, pp. 472–1–4728, 2018. [Online]. Available: https://www.ingentaconnect.com/content/ist/ei/2018/00002018/00000015/art00005

[8] B. E. Claus, Y. Jin, L. A. Gjesteby, G. Wang, and B. De Man, "Metal-artifact reduction using deep-learning based Sinogram completion: Initial results," in *Proc. 14th Int. Meeting Fully Three-Dimensional Image Reconstruction Radiol. Nucl. Med.*, 2017, pp. 631–634.

[9] H. S. Park, S. M. Lee, H. P. Kim, J. K. Seo, and Y. E. Chung, "CT sinogram-consistency learning for metal-induced beam hardening correction," *Medical Physics*, vol. 45, no. 12, pp. 5376–5384, 2018. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13199

[10] T. M. Gottschalk, B. W. Kreher, H. Kunze, and A. Maier, "Deep Learning Based Metal Inpainting in the Projection Domain: Initial Results," in *Machine Learning for Medical Image Reconstruction*, F. Knoll, A. Maier, D. Rueckert, and J. C. Ye, Eds. Cham: Springer International Publishing, 2019, pp. 125–136.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[12] A. Maier, S. Steidl, V. Christlein, and J. Hornegger, *Medical Imaging Systems: An Introductory Guide*. Springer, 2018, vol. 11111.

[13] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2014.

[14] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

Decomposition of Multiple Contrast Agents in Photon-Counting Micro-CT: Iodine and Bismuth

Carlo Amato, Laura Klein, Joscha Maier, Stefan Sawall, Nicole Gehrke, Danielle Franke, Spyridon Gkoumas, Thomas Thüring, Andreas Briel, Christian Brönnimann, and Marc Kachelrieß

***Abstract*—One potential application of new photon counting detectors in computed tomography is the possibility of decomposing simultaneously more than one contrast agent. In previous studies, it was shown through simulations that iodine and bismuth are two potential contrast agents which individually provide the best performance in material decomposition in a preclinical scenario. In this study, we investigate the possibility to decompose bismuth and iodine at the same time and we determine the settings that optimize the process. We find that at a tube voltage of $60\,\mathrm{kV}$, decomposition of the two contrast agents can be obtained with no cross-talk between the two material maps.**

***Index Terms*—Photon-counting CT, material decomposition, micro CT, novel contrast agents.**

I. Introduction

PHOTON counting (PC) detectors are under investigation for several new applications in clinical and preclinical computed tomography (CT) [1], [2]. Thanks to the adjustable energy thresholds of PC detectors, the attenuation coefficient of matter can be sampled at different energies. This can be exploited to differentiate materials whose attenuation coefficients have different dependency on the photon energy [3], [4]. This process is called material decomposition and, in the case of two energy thresholds and two materials to decompose, it is based on the inversion of the following system

$$\mu(E_1, \boldsymbol{r}) = a_1(\boldsymbol{r})\mu_1(E_1) + a_2(\boldsymbol{r})\mu_2(E_1) \quad (1)$$

$$\mu(E_2, \boldsymbol{r}) = a_1(\boldsymbol{r})\mu_1(E_2) + a_2(\boldsymbol{r})\mu_2(E_2), \quad (2)$$

where $\mu(E_i, \boldsymbol{r})$ is the attenuation coefficient of matter sampled in the space $\boldsymbol{r}$ at two different energies E_1 and E_2, μ_1 and μ_2 are the basis functions in which the matter has to be decomposed in and $a_1(\boldsymbol{r})$ and $a_2(\boldsymbol{r})$ are the material maps and the weight factors which describe how the two basis functions are distributed in the space. In order to invert the system, the number of thresholds has to be equal to or higher than the number of materials to decompose and the two basis functions have to be linearly independent. The elements with k-edge are excellent candidates as basis function because of their discontinuities in the attenuation coefficient [5].

One possible and interesting application in CT enabled by PC detectors might be the simultaneous usage of multiple contrast agents [6], [7]. In a recent study [8], several elements with high atomic number and k-edge between $33.2\,\mathrm{keV}$ and $90.5\,\mathrm{keV}$ were investigated as potential contrast agents for micro PC-CT. The study showed that at the same concentration and at $60\,\mathrm{kV}$, iodine and bismuth provide the best performance in material decomposition w.r.t. the other elements. In this study, the possibility of decomposing both materials at the same time was investigated. Other combinations of contrast agents can be found in the literature (iodine-gadolinium [7], iodine-gold [9]), but according to the results obtained in Amato et al. [8], bismuth provides a higher contrast than gadolinium or gold. When decomposing two contrast agents simultaneously, it is essential to ensure that no material is decomposed as another one (cross-talk). As already mentioned, exploiting k-edges plays a crucial role in this. Unlike lighter elements with a k-edge at lower energies, the k-edge of bismuth ($90.5\,\mathrm{keV}$) cannot be exploited when using a tube voltage of $60\,\mathrm{kV}$. Therefore, a brute-force optimization was carried out to find the threshold settings which ensure that no cross-talk is present between iodine and bismuth. Simulations and experiments with a micro PC-CT setup were performed. A mouse-sized phantom with inserts of iodine and bismuth solutions was used. CT images were generated for all the combinations of two energy thresholds and decomposed into iodine and bismuth. Then, the material decomposition performance was evaluated to find optimal threshold settings for the simultaneous decomposition of water, iodine and bismuth.

II. Materials & Methods

A. Experimental setup

The imaging system is a table-top setup (figure 1) equipped with a X-ray source, a rotating stage and a PC detector. The X-ray tube is a Hamamatsu Microfocus operated with a tube voltage of $60\,\mathrm{kV}$, a tube current of $800\,\mu\mathrm{A}$ and a prefiltration of $2\,\mathrm{mm}$ aluminum. The PC detector is a Säntis ME, a prototype manufactured by Dectris AG (Baden-Dättwil, Switzerland). The detector has a $1\,\mathrm{mm}$ CdTe sensor layer, a pixel size of $150\,\mu\mathrm{m}$, a sensitive area of $77.25\,\mathrm{mm} \times 77.55\,\mathrm{mm}$ and 4 energy thresholds. Due to the low power of the tube, the acquisition time per projection was set to $0.8\,\mathrm{s}$. For each scan, a total of $384\,\mathrm{projections}$ with a sampling

C. Amato, L. Klein, Dr. J. Maier, Dr. S. Sawall, and Prof. Dr. M. Kachelrieß are with German Cancer Research Center (DKFZ), Division of X-Ray Imaging and CT, Heidelberg, Germany, and with the University of Heidelberg, Heidelberg, Germany.

Dr. N. Gehrke, Dr. D. Franke and Dr. A. Briel are with nanoPET Pharma GmbH, Berlin, Germany.

Dr. S. Gkoumas, Dr. T. Thüring,and Dr. C. Brönnimann are with DECTRIS Ltd., Baden-Dättwil, Switzerland.

Corresponding author: Carlo Amato (carlo.amato@dkfz.de)

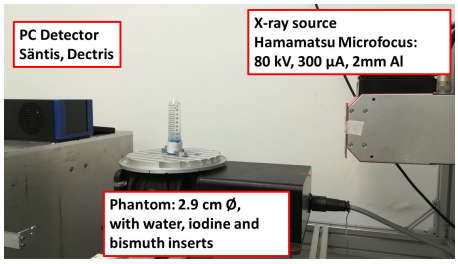


Fig. 1. Foto of the experimental table-top setup.

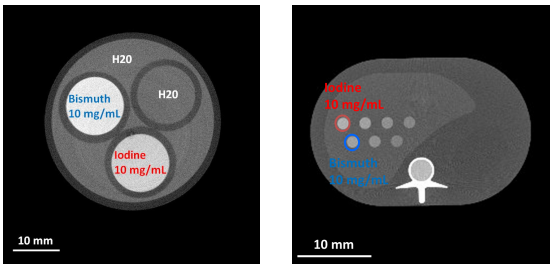


Fig. 2. Left: Slice of the phantom used in the experiments. Right: Slice of the phantom used in the simulations.

angle of $0.9375°$ were acquired to have a $360°$ coverage. The source-to-detector distance was set to $315\,\mathrm{mm}$ with a magnification of around 2.3

The used phantom (figure 2 left) consists of a water cylinder with a diameter of $29\,\mathrm{mm}$ containing three inserts filled with water, a $10\,\mathrm{mg/mL}$ bismuth solution and a $10\,\mathrm{mg/mL}$ iodine solution, respectively.

In order to find the combination of thresholds which provides the best performance in material decomposition, the phantom was scanned with all the thresholds between $16\,\mathrm{keV}$ and $38\,\mathrm{keV}$, in steps of $2\,\mathrm{keV}$. The threshold-projections were subtracted in order to obtain bin-projections for all the combination of the two threshold values. From 12 threshold projections, 66 combinations of 2 bin-projections were obtained.

The bin images were reconstructed with the FDK algorithm [10]. To prevent ring artifacts, the defect pixels and detector gaps were corrected through linear interpolation before the reconstruction.

B. Simulations

The simulations of the setup were performed according to the described experimental geometry. The 3D polychromatic forward projections of the phantom were generated using a Tucker spectrum [11] filtered with $2\,\mathrm{mm}$ of aluminum and realistic detector spectral-response obtained thought the model by Faby et al. [12]. Poisson noise was added in projection domain according to the number of transmitted photons. The used spectrum and the attenuation coefficients of iodine and bismuth are shown in figure 3. Images were reconstructed using the FDK algorithm [10].

The simulated phantom (figure 2 right) is a human liver phantom scaled to a size of $30\,\mathrm{mm}$ with contrast agent

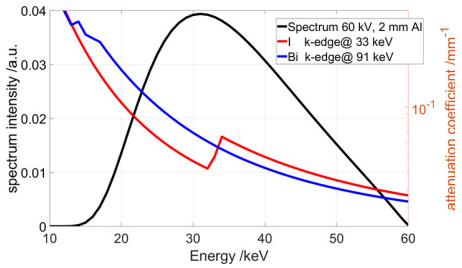


Fig. 3. The Tucker spectrum and the attenuation coefficients of iodine and bismuth used in the simulations. It can be seen how the k-edge of iodine ($33.2\,\mathrm{keV}$) is very close to the center of mass of the spectrum ($35.3\,\mathrm{keV}$), whereas the k-edge of bismuth ($90.5\,\mathrm{keV}$) is not visible.

inserts. For this investigation, only the inserts with iodine and bismuth $10\,\mathrm{mg/mL}$ solutions were taken into account. These are highlighted in figure 2 right in red and blue, respectively. The bin images were generated for all the combinations of two thresholds between $16\,\mathrm{keV}$ and $56\,\mathrm{keV}$ in steps of $2\,\mathrm{keV}$.

C. Material decomposition and image analysis

For a two material decomposition, the algorithm by Faby et al. [13] was used. Two bin images were used to decompose two materials: water and one of the contrast agents of interest. The material decomposition was performed for each investigated combination of two thresholds.

In order to quantify the performance of the two material decomposition, the CNR was chosen as figure of merit. The CNR is defined as:

$$\mathrm{CNR} = \frac{|\langle S_1\rangle - \langle S_2\rangle|}{\sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}}, \tag{3}$$

where $\langle S_i\rangle$ and σ_{S_i} are the mean and the standard deviation in two ROIs. The output of the two materials decomposition are the material maps of water and of the specific contrast agent. The CNR was evaluated in the contrast agent material map between ROIs in water and in the vial with the contrast agent of interest.

For each of the two contrast agents, a 2D map that shows the CNR as a function of the two energy thresholds was generated. Each map was normalized to its maximum value. The peaks on the CNR maps indicate which are the two thresholds which provide an optimal decomposition between the water and the specific contrast agent.

In order to chose the combination of two energy thresholds which optimize the decomposition of the two contrast agents at the same time, the overlap between the peaks of the two CNR maps was exploited. To do so, the normalized CNR maps of iodine and bismuth were averaged and the optimal threshold values were chosen in correspondence of the CNR peak in the mean CNR map.

Once the two optimized energy thresholds are found, a three material decomposition is performed using volume fraction

conservation [14]. The system of equations (1)-(2) can be rewritten as:

$$\mu(E_1, \boldsymbol{r}) = a_1(\boldsymbol{r})\mu_1(E_1) + a_2(\boldsymbol{r})\mu_2(E_1) + a_3(\boldsymbol{r})\mu_3(E_1) \quad (4)$$
$$\mu(E_2, \boldsymbol{r}) = a_1(\boldsymbol{r})\mu_1(E_2) + a_2(\boldsymbol{r})\mu_2(E_2) + a_3(\boldsymbol{r})\mu_3(E_2) \quad (5)$$
$$1 = a_1(\boldsymbol{r}) + a_2(\boldsymbol{r}) + a_3(\boldsymbol{r}) \quad (6)$$

allowing to decompose three materials with two energy thresholds. Equation (6) enforces the sum of the volume fractions $a_i(\boldsymbol{r})$ to add to unity.

III. Results

A. Simulations

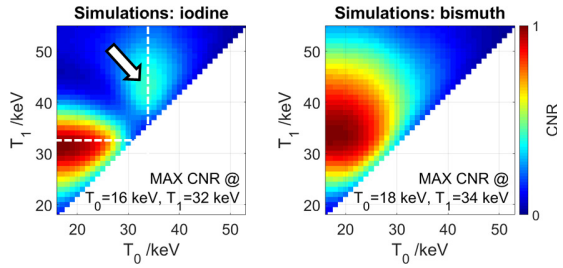


Fig. 4. CNR of the simulated material decomposition as function of the two energy thresholds for iodine (left) and bismuth (right). The k-edge of iodine was highlighted with a dashed white line at both thresholds. The white arrow points at the secondary peak of iodine.

The simulation results of the two energy threshold optimization for iodine and bismuth are shown in figure 4. Each of the 2D CNR maps shows at least one peak. The iodine CNR map (figure 4 left) exhibits two peaks: the main at T_0=16 keV, T_1=32 keV and the other at T_0=34 keV, T_1=44 keV. The main peak corresponds to a scenario where the highest threshold is at the iodine k-edge (33.2 keV), whereas the second peak has the lower threshold at the k-edge. Both peaks are broad and elongated in the direction of the threshold that is at the k-edge. The bismuth CNR map (figure 4 right) has a broad peak with no elongation, accordant to the fact that the k-edge of bismuth is at 90.5 keV and therefore outside the used spectrum. The maximum CNR is at T_0=18 keV, T_1=34 keV.

The material decompositions at the three CNR peaks are shown in figure 5. The material decomposition of iodine with thresholds at T_0=16 keV, T_1=32 keV (figure 5 top-left) does not show positive contribution by bismuth or bones in the iodine map. Thanks to the choice of thresholds, iodine is the only material that exhibits an increase of the attenuation coefficient between the two energy bins. This condition is not satisfied anymore if the vertical CNR peak of figure 4 left is exploited and iodine is decomposed using thresholds at T_0=34 keV, T_1=44 keV. Such material decomposition (figure 5 top-right) shows positive contributions of bismuth and bones in the iodine map. The bismuth material decomposition at its CNR maximum (T_0=18 keV, T_1=34 keV) is shown in figure 5 bottom. The bismuth map shows contributions from bones but not from iodine. This is expected because for that specific choice of thresholds, the k-edge of iodine is exploited and therefore cannot be decomposed as bismuth.

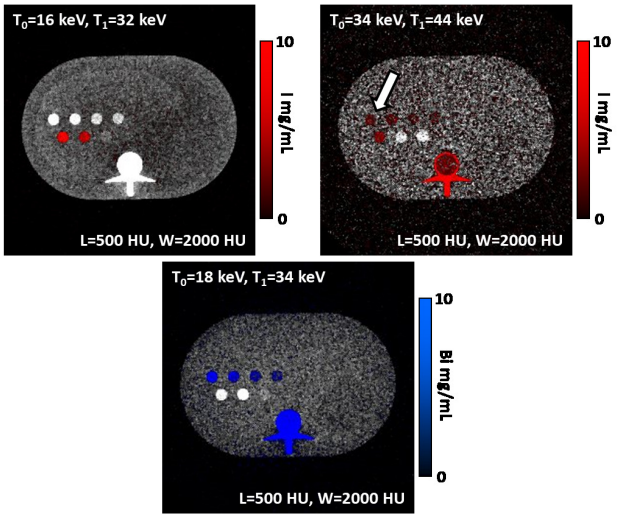


Fig. 5. Material decomposition of iodine with T_0=16 keV, T_1=32 keV (top-left), iodine with T_0=34 keV, T_1=44 keV (top-right) and bismuth with T_0=18 keV, T_1=34 keV (bottom). In the top-right image, it can be seen how bismuth is partially decomposed as iodine (white arrow) for the specific threshold choice.

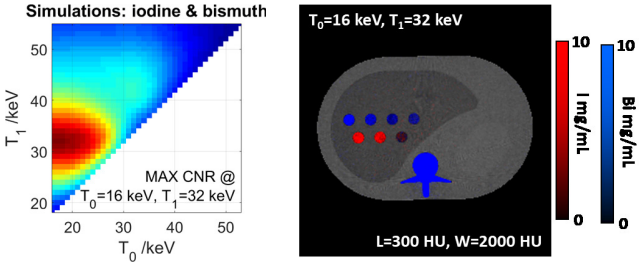


Fig. 6. Left: mean of the CNR maps of iodine and bismuth. Right: three material decomposition using T_0=16 keV and T_1=32 keV.

The mean of the two normalized CNR maps is shown in figure 6 left. The overlap between the horizontal peak of iodine and the broad peak of bismuth results in a peak with a maximum at T_0=16 keV and T_1=34 keV. The three material decomposition for these thresholds value is shown in figure 6 right. No cross-talk between the two contrast agents is present.

B. Measurements

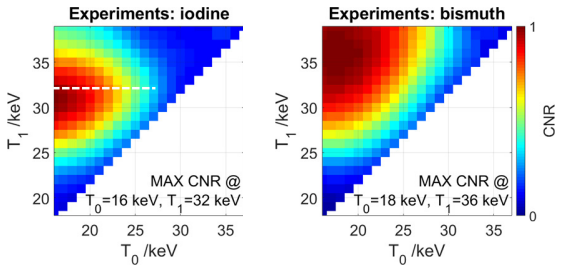


Fig. 7. CNR of the material decomposition as function of the two energy thresholds for iodine (left) and bismuth(right). The k-edge of iodine was highlighted with a dashed white line at T_1 =33.2 keV.

The measured results of the two energy thresholds optimization for iodine and bismuth are shown in figure 7. In

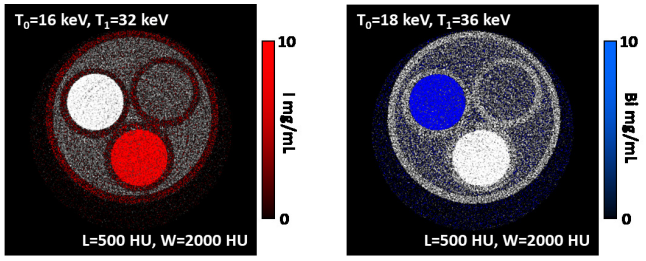


Fig. 8. Experimental material decomposition of: iodine with T_0=16 keV and T_1=32 keV (left) and bismuth at T_0=18 keV and T_1=36 keV (right).

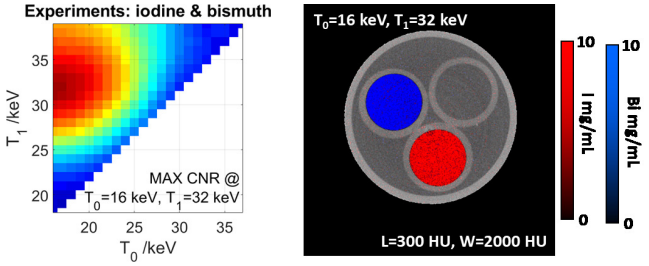


Fig. 9. Left: mean of the experimental CNR maps of iodine and bismuth. Right: 3 material decomposition using T_0=16 keV and T_1=32 keV.

the iodine CNR map (figure 7 left), the vertical CNR peak is not found because the interval of investigated energies is narrower than in simulations. The horizontal peak, with the higher threshold at the iodine k-edge, is visible and has a maximum at T_0=16 keV and T_1=32 keV. In the bismuth CNR map (figure 7 right), a broad peak is found, with a maximum at T_0=18 keV and T_1=36 keV.

The material decompositions at the two CNR peaks are shown in figure 8. The material decomposition of iodine with thresholds at T_0=16 keV, T_1=32 keV (figure 8 left) does not show positive contribution by bismuth or bones in the iodine map. The material decomposition of bismuth with thresholds at T_0=18 keV, T_1=36 keV (figure 8 right) shows no contribution from iodine.

The mean of the two normalized CNR maps is shown in figure 9 left. The overlap between the iodine and bismuth peaks results in a peak with a maximum at T_0=16 keV and T_1=32 keV. The material decomposition of the two contrast agents for these thresholds is shown in figure 9 right. No cross-talk between the two materials is present.

IV. Discussion & Conclusion

A threshold optimization was performed to optimize the CNR of the material decomposition of the individual contrast agents. For iodine, it was shown that there are two ways of exploiting the k-edge at 33.2 keV, depending if the lower or the higher energy threshold is set at the k-edge. Between the two possibilities, the scenario with the higher threshold at the k-edge exhibited no contribution of bismuth or bone in the iodine material maps. For bismuth, it was shown that even if the k-edge is not exploited, there is a combination of thresholds which optimize the CNR. Thanks to the overlap of CNR peaks of the two contrast agents, the combination of the CNR optimization for iodine and bismuth leads to a set of thresholds which optimize both material decompositions at the same time. Results have shown how it is possible to decompose both materials at the same time without cross-talk between the two. A very good agreement between simulations and measurements was found.

Acknowledgment

Parts of the reconstruction and simulation software were provided by RayConStruct® GmbH, Nürnberg, Germany. This work was supported in part by the Eurostars project E!11793 MultiContrastPhotonCT and by the Helmholtz International Graduate School for Cancer Research, Heidelberg, Germany.

References

[1] E. Fredenberg, "Spectral and dual-energy X-ray imaging for medical applications," *Nuclear Instruments and Methods in Physics Research, Section A*, vol. 878, pp. 74–87, 2018.

[2] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: Technical Principles and Clinical Prospects," *Radiology*, vol. 289, no. 2, pp. 293–312, 2018.

[3] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography," *Physics in Medicine and Biology*, vol. 21, no. 5, p. 002, 1976.

[4] E. Roessl and R. Proksa, "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," *Physics in Medicine and Biology*, vol. 52, no. 15, pp. 4679–4696, 2007.

[5] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler, A. Livne, and R. Proksa, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Physics in Medicine and Biology*, vol. 53, no. 15, pp. 4031–4047, 2008.

[6] R. Symons, B. Krauss, P. Sahbaee, T. E. Cork, M. N. Lakshmanan, D. A. Bluemke, and A. Pourmorteza, "Photon-counting CT for simultaneous imaging of multiple contrast agents in the abdomen: An ¡i¿in vivo¡/i¿ study," *Medical Physics*, vol. 44, no. 10, pp. 5120–5127, 2017.

[7] C. T. Badea, D. P. Clark, M. Holbrook, M. Srivastava, Y. Mowery, and K. B. Ghaghada, "Functional imaging of tumor vasculature using iodine and gadolinium-based nanoparticle contrast agents: a comparison of spectral micro-CT using energy integrating and photon counting detectors," *Physics in Medicine and Biology*, 2019.

[8] C. Amato, L. Klein, J. Maier, S. Sawall, N. Gehrke, D. Franke, S. Gkoumas, T. Thüring, A. Briel, C. Brönnimann, and M. Kachelrieß, "Potential of High-Z Elements in Photon-Counting Micro-CT for Optimized Material Decomposition," *accepted EMIM*, 2020.

[9] D. P. Clark, K. Ghaghada, E. J. Moding, D. G. Kirsch, and C. T. Badea, "In vivo characterization of tumor vasculature using iodine and gold nanoparticles and dual energy micro-CT." *Physics in medicine and biology*, vol. 58, no. 6, pp. 1683–704, 2013.

[10] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, no. 6, p. 612, 1984.

[11] D. M. Tucker, G. T. Barnes, and D. P. Chakraborty, "Semiempirical model for generating tungsten target x-ray spectra," *Medical Physics*, vol. 18, no. 2, pp. 211–218, 1991.

[12] S. Faby, J. Maier, S. Sawall, D. Simons, H.-P. Schlemmer, M. Lell, and M. Kachelrieß, "An efficient computational approach to model statistical correlations in photon counting x-ray detectors," *Medical Physics*, vol. 43, no. 7, pp. 3945–3960, 2016.

[13] S. Faby, S. Kuchenbecker, D. Simons, H.-P. Schlemmer, M. Lell, and M. Kachelrieß, "CT calibration and dose minimization in image-based material decomposition with energy-selective detectors," B. R. Whiting and C. Hoeschen, Eds., vol. 9033. International Society for Optics and Photonics, 3 2014, p. 903318.

[14] P. V. Granton, S. I. Pollmann, N. L. Ford, M. Drangova, and D. W. Holdsworth, "Implementation of dual- and triple-energy cone-beam micro-CT for postreconstruction material decomposition," *Medical Physics*, vol. 35, no. 11, pp. 5030–5042, 2008.

Convolutional Regularization for Accelerated Spectrotemporal CT Reconstruction

Darin P. Clark and Cristian T. Badea

***Abstract*—Machine learning approaches to x-ray computed tomography (CT) denoising and regularization are poised to advance the state of the art. Specifically, computationally efficient convolutional neural networks (CNNs) have shown enormous potential for learning signal restoration models directly from paired training data. With this power, we hypothesize that CNNs will fundamentally enable routine reconstruction of combined time and energy (spectrotemporal, ST) CT data. We expect the demand for ST-CT will rise dramatically in the coming years thanks to the advancement of photon-counting detector (PCD) technology, which allows the acquisition of multi-energy CT data with a single scan. In previous work, we demonstrated an effective iterative reconstruction scheme for preclinical ST-CT and showed that the temporal portion of the iterative reconstruction problem may be replaced with a single pass of a 4D (3D+time) CNN. With the ultimate goal of producing a fast approximation for iterative ST-CT reconstruction, here, we advance the spectral regularization subproblem. Specifically, we propose a novel CNN architecture based on the split Bregman method, which robustly denoises spectral CT data. Using *in vivo* cardiac micro-CT data, we then demonstrate the performance of the proposed network for material decomposition and regularization of ST data. Ultimately, CNN-based regularization delivers a 12-fold computational speed-up relative to our iterative ST reconstruction algorithm.**

***Index Terms*—machine learning, photon counting, preclinical, x-ray computed tomography**

I. Introduction

Recent advancements in machine learning have the potential to impact nearly every facet of x-ray computed tomography (CT) data acquisition, reconstruction, post-processing, and analysis. Here, we are interested in a subset of this work which demonstrates state-of-the-art image domain denoising performance (e.g. [1, 2]) with the potential to bridge the gap between analytical reconstruction methods, which are sensitive to sampling conditions and noise, and iterative reconstruction methods, which often reduce noise at the expense of increased complexity. Specifically, we hypothesize that the representational power and efficiency of convolutional neural networks (CNNs) will fundamentally enable routine reconstruction of combined time and energy (spectrotemporal, ST) CT data.

Currently, ST-CT is in its infancy: a preclinical and clinical research topic with several promising clinical applications demonstrated for dual-energy cardiac imaging [3, 4]. However, we expect the demand for ST-CT protocols will increase sharply within the next ten years, following the maturation of photon-counting detector (PCD) technology, which allows multi-energy CT with a single scan. Widespread adoption of this technology is anticipated due to the myriad of benefits associated with photon-counting CT [5] and with spectral information in cardiac imaging applications [6]. However, significant advancements are needed to overcome fundamental PCD limitations: distortion of recorded spectral information and increased noise associated with photon binning at fixed dose [7].

Preclinical micro-CT provides an excellent testbench for PCD-CT and ST-CT. Working with a prototype PCD micro-CT system, we have proposed an iterative reconstruction scheme which enables high-fidelity reconstruction of *in vivo*, ST-CT data acquired in a mouse model of atherosclerosis. Based on the split Bregman method, our iterative reconstruction scheme exploits the low-rank structure of spectral contrast and the sparse structure of temporal contrast [8]. We have further shown that sparse temporal contrast can be highly amenable to regularization with CNNs, possibly replacing the need for iterative reconstruction of temporal data with a single pass of a 4D (3D+time) CNN [9].

Toward a fast approximation for iterative ST-CT reconstruction (analytical reconstruction + one CNN pass), here we investigate multi-channel regularization of the spectral dimension. Specifically, we propose a new CNN architecture based on the split Bregman method [10]. Key design features include Laplacian pooling, to control the flow of low and high spatial frequencies, and several tunable parameters. After being trained on time-averaged spectral data, adjusting these parameters allows our network to denoise time-resolved spectral data with higher noise, a scenario where CNNs may generalize better than classic denoising methods [11]. We also demonstrate the fidelity of three-material decomposition following regularization.

II. Materials and Methods

A. In Vivo Cardiac Micro-CT

A detailed description of our preclinical, PCD-based micro-CT system and imaging procedures can be found in our previous work [8]. In brief, our PCD micro-CT system used a SANTIS 0804 ME prototype PCD (DECTRIS Ltd., Baden-Daettwil, Switzerland; dectris.com): 1 mm thick CdTe sensor; 150-µm isotropic pixels (515x257); four independent energy thresholds (set to 25, 34, 40, 55 keV). Projection data sets were acquired with a G297 x-ray tube (Varian Medical Systems, Palo Alto, CA; 0.3 mm focal spot size; tungsten

All work was performed at the Duke Center for In Vivo Microscopy. This work was supported by the NIH National Cancer Institute (R01 CA196667, U24 CA220245).

D. P. Clark and C. T. Badea are with the Center for In Vivo Microscopy, Duke University Medical Center, Dept. of Radiology, Durham, NC 27710 USA (e-mail: cristian.badea@duke.edu).

rotating anode; filtration: 0.1 mm Cu; 80 kVp, 5mA, 10 ms exposure/projection). Scanning used a helical trajectory with 3 rotations and 1.25 cm of total translation (9000 projections/ threshold; ~190 mGy dose).

Animal scanning was approved by the Duke University Institutional Animal Care and Use Committee. Six mice were scanned: female ApoE-/- mice obtained from The Jackson Laboratory (B6.129P2-Apoe^{tm1Unc}/J). ApoE-/- mice have germline deletion of the apolipoprotein E gene [12] and are prone to develop calcified atherosclerotic lesions [13]. At 8–12 weeks of age, the ApoE-/- mice were exposed to 25 fractions of 2 Gy partial-heart irradiation, mimicking breast cancer treatment. After irradiation, the mice were kept on a regular diet for a year prior to imaging. Three days before imaging, the mice were intravenously injected with gold nanoparticles (15 nm AuroVist, nanoprobes.com) at a dose of 0.004 mL/g mouse (to label myocardial injury [14]). Three days later, immediately prior to imaging, the same mice were injected with a liposomal iodinated contrast agent (described in [15]; 0.012 mL/g mouse). The recorded ECG signal was used for retrospective gating and projection weighting during reconstruction. Respiratory gating was not performed.

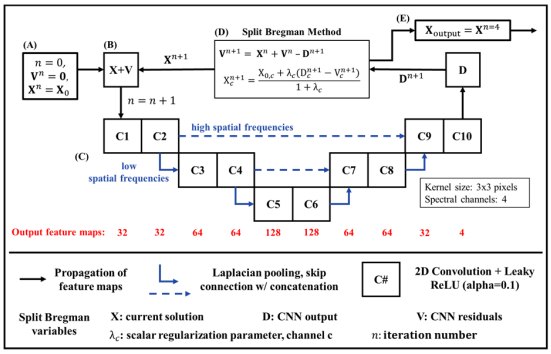


Fig. 1. 2D CNN for spectral denoising. Given an input image, (A) $\mathbf{X}_0$, the network iteratively solves for a denoised output image, (E) $\mathbf{X}_{output}$, by employing the split Bregman method for denoising (B-D). Key features of the proposed network are the use of Laplacian pooling layers to partition high and low spatial frequencies and the tracking of the regularization residuals, $\mathbf{V}$. Removing uncorrelated residuals from the network input may be a key to training networks which generalize across input noise levels.

B. *Split Bregman Network Architecture*

Fig. 1 summarizes the proposed network architecture. Broadly, the network is a fusion of the popular U-net [16] and ResNet [17] architectures and the split Bregman optimization method [10]. The input data, $\mathbf{X}_0$, is denoised over several iterations, indexed by n, through a symmetric, multi-resolution CNN (U-net; bold indicates a tensor with multiple independent channels, indexed by c). Similar to the ResNet, the output of CNN regularization, $\mathbf{D}$, is added back to $\mathbf{X}_0$ (Fig. 1D). Following application of the split Bregman method for denoising, $\mathbf{D}$ is added back subject to regularization weighting parameters, $\boldsymbol{\lambda}$. Values of $\boldsymbol{\lambda}$ less than one favor the input data, while values of $\boldsymbol{\lambda}$ greater than one favor CNN regularization, potentially allowing the optimal regularization parameters to be learned during network training, or even tuned after network training has completed.

For iterative reconstruction, the split Bregman method can be applied to solve problems of the following form:

$$\mathbf{X}=\underset{\mathbf{X}}{\operatorname{argmin}}\,\tfrac{1}{2}\|\mathbf{AX}-\mathbf{G}\|_F^2+\mu R(\mathbf{X}), \qquad (1)$$

where $\mathbf{X}$ (reconstructions) is the minimum norm solution to the inverse problem, $\mathbf{AX}=\mathbf{G}$ (log-transformed projection data), evaluated channel-wise and subject to the multi-channel penalty function(s) $R(\mathbf{X})$ and regularization parameter μ. The difficulty of regularizing $\mathbf{X}$ through the system matrix $\mathbf{A}$ is alleviated through the introduction of auxiliary variables $\mathbf{D}$ and $\mathbf{V}$, which split the data fidelity update and the regularization into independent sub-problems which can be solved sequentially [18].

Practically, we have had great success in applying this same Bregman framework to the problem of image-domain spectral denoising (where $\mathbf{A}$ in Eq. 1 is the identity matrix, $\mathbf{G}$ is the noisy input data). Following the expectation of low-rank structure in spectral contrast, iterative, joint regularization of spectral channels enables precise control over the trade-off between intensity bias and noise variance, which can dramatically improve the fidelity of spectral regularization [19]. In this work, we extend our spectral regularization framework to incorporate a CNN-based regularizer and attempt to replace the entire iterative reconstruction process with 2D, image-domain denoising (Fig. 1; $\mathbf{G}=\mathbf{X_0}$).

Consistent with findings on tight frames within network architectures [2], our network employs "Laplacian" pooling layers. These layers construct a Laplacian pyramid [20] within the context of a multi-resolution U-net. They operate by performing normalized Gaussian convolution (FWHM = 2 pixels) and 2x2 average pooling to downsample (per spectral channel, feature map), and 2x2 nearest-neighbor upsampling and Gaussian convolution to upsample. The residual associated with these conjugate operations is passed through skip connections, while the downsampled component is passed to the next level of the U-net (Fig. 1C).

C. *Training and Computation*

The proposed 2D CNN was trained to match analytical reconstruction results (helical, weighted FBP [21]) to multi-channel, iterative reconstructions produced using the split Bregman framework (outlined above, detailed in [8, 19]). During training the temporal dimension of the six ST-CT data sets was averaged out to maximize image quality. Five of the six data sets were used for training, while the sixth data set was used for validation. Tests conducted post-training were also applied to the validation data; however, during testing, the data was fully resolved both in time and energy (10 cardiac phases x 4 energy bins = 40 volumes, each with 360x360x374, 123-μm, isotropic voxels). Training minimized the cost function shown in Fig. 2A, which consists of three terms: (1) image domain root-mean-squared error (RMSE) ($\mathbf{Y}$, iterative reconstruction label), (2) material domain RMSE (see below), and (3) a debiasing term. To combat under representation of valid spectral contrast in the training data, debiasing involved choosing uniformly distributed, random material combinations of iodine, gold, and water (ranges determined from training set), and used them to construct constant intensity test images, $\overline{\mathbf{X}}$, which were penalized to remain constant at the output of the CNN, $\mathrm{CNN}(\overline{\mathbf{X}})$. RMSE values were averaged over all image pixels and spectral or material channels (units: HU/1000, normalized material concentration). Data augmentation included random image orientation, horizontal and vertical

image flipping, and up to 360° rotation (Fig. 2B-E).

Training executed with a batch size of four 2D CT images and four debiasing images, each with four spectral channels (450 mini-batches per training epoch), for a total of 130 epochs. Following training, the model with the lowest validation cost (epoch 60, $\boldsymbol{\lambda}$=0.55 for all channels) was used for further testing. Training was performed using Tensorflow v2.0.0 (tensorflow.org) and Python 3.6.8 (python.org). Training was performed with a NVIDIA Quadro RTX 8000 GPU on an Ubuntu 18.04 Linux workstation with 64 GB of RAM and an Intel i7 5960X CPU (~12 hours).

Analytical (WFBP) and iterative CT reconstructions were performed with our GPU-based, multi-channel CT reconstruction toolkit [22] on an Ubuntu Linux workstation (16.04) with 4x NVIDIA Titan Xp GPUs, 256 GB of system RAM, and 2x Intel Xeon E5-2650 v4 CPUs. Analytical reconstruction took 7 minutes per data set (40x, 360x360x374 voxel volumes). Iterative reconstruction took ~10 hours per data set (unregularized algebraic reconstruction + 3 Bregman iterations with regularization [8]). Application of the trained CNN to 40 volumes required 14 minutes (5 internal iterations, see *Results*). Including 30 minutes of preprocessing time (projection normalization, ring artifact correction, etc.), this brings the total time for analytical reconstruction and CNN application to 51 minutes vs. more than 10 hours for iterative reconstruction of each ST-CT data set (12x speed-up).

D. Characterization and Material Decomposition

CNN-based regularization is judged by comparing analytical reconstruction and CNN evaluation results to iterative reconstruction results using RMSE and structural similarity (SSIM, dynamic range: 1.5, [23]). In light of distortions inherent in PCD data and the low-rank and sparse signal model assumed by our iterative reconstruction [8], our iterative reconstruction results cannot be considered a gold standard; however, alternative strategies for obtaining high-fidelity training labels (e.g. higher dose acquisitions) are impractical for *in vivo*, preclinical, ST micro-CT, which requires short exposures and fast acquisition times to image the mouse heart (300-600 bpm). To improve the quality of the trained network, we train with spectral data which is averaged in time and then apply the network to time-resolved spectral data, using network iteration to address increased noise.

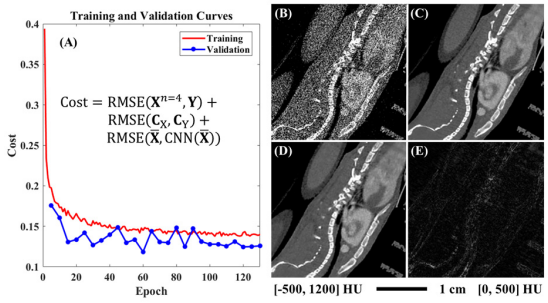


Fig. 2. CNN training and validation. (A) CNN training and validation curves computed for the cost function shown (units: HU/1000, normalized material concentration; averaged over all spectral channels). (B) One of four network input channels (55-80 keV) shown following random data augmentation. (C) Training label. (D) Network output. (E) Absolute residuals between (C) and (D) with increased contrast.

We further judge the quality of our CNN-based results through material decomposition. Specifically, we decompose our results into iodine, gold, and water material maps:

$$\mathrm{C} = \underset{\mathrm{C}}{\operatorname{argmin}} \|\mathrm{CM} - \mathrm{X}\|_{\mathrm{F}}^2 \text{ subject to } \mathrm{C} \geq 0. \qquad (2)$$

Unlike in Eq. 1, here the channels of X and C (material maps) are vectorized as columns of a 2D matrix. The material sensitivity matrix, M, is calibrated using a physical phantom [24] and relates material concentrations (iodine, gold; mg/mL) and fractions (pure water=1.0) to attenuation measures for each spectral channel. Each material sensitivity vector within M is unit normalized for cost function computations (Fig. 2A). The non-negativity constraint on C is only applied post-training. Gold is used as a material but represents both gold and calcium because the 80 kVp source spectrum does not span the K-edge of gold (80.7 keV).

III. Results

A. Training and Validation

Fig. 2 summarizes training and validation results for the proposed network architecture. Specifically, 2A shows the cost function values per epoch for the five training data sets and the single validation data set (evaluated every 5 epochs). Based on prior network training with a simpler cost function (image domain RMSE only; not shown), turbulence in the validation cost appears tied to the inclusion of the material decomposition in the cost function. Attempting to minimize material decomposition error amplifies noise variance but better conditions the results for material decomposition. Fig. 2B-E illustrate augmented network inputs and outputs (time-averaged; 55-80 keV shown). The network reduces the noise standard deviation in water from 309 HU to 26HU without significant bias (4.6 HU relative to input, Fig. 2E).

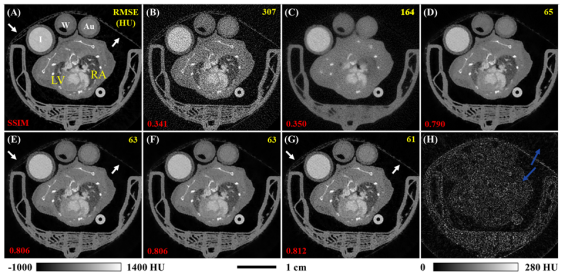


Fig. 3. Application of the spectral CNN to *in vivo* ST-CT data. (A) Reference iterative reconstruction results (ventricular systole, 40-80 keV). (B) Input analytical reconstruction results (WFBP). (C-F) Denoised network output, **X**, after 1, 3, 5, and 7 internal Bregman iterations (λ_c = 2.0). (G) Denoised output with 5 internal iterations and noise weighting by channel. (H) Absolute residuals between (A) and (G) with increased contrast. Arrows and measurements are referenced in the text.

B. Testing

All post-training testing was conducted using the time and energy resolved ST data. Fig. 3 summarizes these post-training results produced with the testing data set and illustrated for the third energy threshold (40-80 keV, systole). Comparing the iterative reconstruction results (A) with the analytical reconstruction results (B), the improvement in image quality and the apparent reduction in the average noise level (307 HU) without loss of detail demonstrates the value of multi-channel iterative reconstruction, despite its cost.

Fig. 3C illustrates the consequences of training the proposed network on temporally averaged data and then applying it to temporally resolved data with higher noise

(~2.6x): the network heavily over-smooths the result after a single internal iteration ($\mathbf{X}^{n=2}$, Fig. 1D). Interestingly, however, when the network is applied for three internal iterations (as during training; $\mathbf{X}^{n=4}$), the solution improves significantly (RMSE: 65 HU, SSIM: 0.79; averaged over all cardiac phases, energies). Furthermore, the network exhibits convergence behavior, with the output unchanged between 5 and 7 iterations (optimal $\lambda_c = 2.0$, all channels). Finally, when the regularization parameter is scaled by the relative noise standard deviation in each energy channel (optimal $\lambda_c = 1.0 * \sigma_c / \min_c(\boldsymbol{\sigma})$) as we commonly do for spectral regularization [19], the solution further improves (5 internal iterations; RMSE: 61 HU, SSIM: 0.81). White arrows indicate low-contrast, high-frequency tape better resolved with iteration and noise scaling. Blue arrows highlight a significant change in noise variance relative to the iterative reconstruction for regions of air (-HU) vs. higher attenuation (+HU), potentially caused by the over representation of air during training or by the use of Leaky ReLU activations [25].

Finally, Fig. 4 compares material decomposition results between WFBP, WFBP and the CNN (5 iterations, noise scaling), and the output of iterative reconstruction (non-negativity constraint applied, Eq. 2). Quantitatively, the percent difference in the material measurements between the CNN output and iterative reconstruction results was 7.4% for 12 mg/mL of iodine in water, 22% for 4 mg/mL of gold in water, and 4.8% for pure water. Qualitatively, two calcified plaques are clearly visible in both sets of decomposition results (yellow arrows); however, a "false positive" calcification appears in the CNN output results, and the liver, which is known to contain iodine (red) and gold (green), appears patchy in the CNN output (blue arrows).

IV. Discussion and Conclusions

In this work, we have advanced machine-learning based approaches to spectral regularization by proposing a novel CNN architecture which synergizes the popular U-net and ResNet architectures within the framework of the split Bregman optimization method. Through preliminary experiments, we show that the proposed architecture is capable of robust spectral regularization, even generalizing to data with a significantly higher level of noise. We believe that advancements which improve the generalization capabilities of machine learning algorithms will ultimately lead to routine and trusted application of such methods in medical imaging.

A number of questions, quantitative characterizations, and possible extensions of our method remain to be explored. For instance, we employ a linear material decomposition model to photon-counting data which we know to be inherently distorted. Given the ideal low-rank structure of spectral CT data, correcting such distortions prior to CNN training and regularization is likely to significantly improve the network's performance, since training data will be easier to generalize. Also, we have trained our network to reproduce the results of a specific iterative reconstruction method. While this provides higher fidelity labels for training than would otherwise be available for *in vivo* cardiac micro-CT, it also reinforces biases and assumptions made by our reconstruction method (e.g. low-rank and sparse signal model, log-normalization of spectral CT data). Future revisions to our method could benefit from unsupervised training approaches.

In future work, we intend to combine our spectral regularization method with our previous temporal regularization method [9] to yield a fast approximation for ST-CT reconstruction which synergizes the space, time, and energy dimensions in performing denoising. We expect that advancements in PCD technology, spectral distortion correction, and such fast approximations will make routine ST-CT imaging a reality.

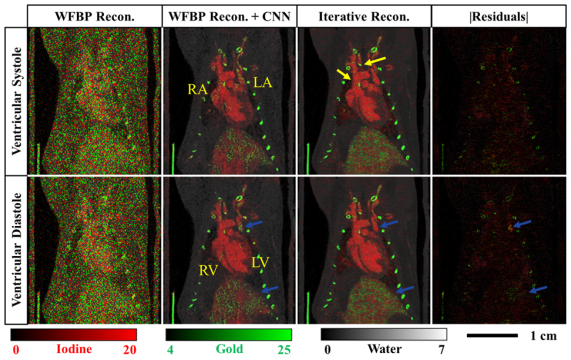


Fig. 4. Material decomposition results (2D, coronal slices). Material decompositions are shown following WFBP reconstruction, CNN evaluation, and iterative reconstruction (columns) at ventricular systole and diastole (rows). Absolute residual images compare CNN results with iterative reconstruction results on the same intensity scale. Yellow arrows indicate the locations of calcified plaque (green), while blue arrows highlight differences between the CNN and iterative results. Iodine, gold units: mg/mL. Water units: pure water = 1.0.

Acknowledgment

This work was made possible by the loan of a prototype SANTIS 0804 ME photon-counting x-ray detector from DECTRIS, Ltd. (Baden-Daettwil, Switzerland; dectris.com).

References

[1] H. Chen *et al.*, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE transactions on medical imaging,* vol. 36, no. 12, pp. 2524-2535, 2017.
[2] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE transactions on medical imaging,* vol. 37, no. 6, pp. 1418-1429, 2018.
[3] D. De Santis *et al.*, "Heavily calcified coronary arteries: advanced calcium subtraction improves luminal visualization and diagnostic confidence in dual-energy coronary computed tomography angiography," *Investigative radiology,* vol. 53, no. 2, pp. 103-109, 2018.
[4] I. Danad, B. Ó Hartaigh, and J. K. Min, "Dual-energy computed tomography for detection of coronary artery disease," *Expert review of cardiovascular therapy,* vol. 13, no. 12, pp. 1345-1356, 2015.
[5] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: technical principles and clinical prospects," *Radiology,* vol. 289, no. 2, pp. 293-312, 2018.
[6] I. Danad, Z. A. Fayad, M. J. Willemink, and J. K. Min, "New applications of cardiac computed tomography: dual-energy, spectral, and molecular CT imaging," *JACC: Cardiovascular Imaging,* vol. 8, no. 6, pp. 710-723, 2015.
[7] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical Physics,* vol. 40, no. 10, p. 100901, 2013.
[8] D. Clark, M. Holbrook, C. Lee, and C. Badea, "Photon-counting cine-cardiac CT in the mouse," *PLoS ONE,* vol. 14, no. 9, p. e0218417, 2019.
[9] D. Clark and C. Badea, "Convolutional regularization methods for 4D, x-ray CT reconstruction," presented at the Medical Imaging 2019: Physics of Medical Imaging, 2019.
[10] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM Journal on Imaging Sciences,* vol. 2, no. 2, pp. 323-343, 2009.
[11] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing,* vol. 27, no. 9, pp. 4608-4622, 2018.
[12] J. A. Piedrahita, S. H. Zhang, J. R. Hagaman, P. M. Oliver, and N. Maeda, "Generation of mice carrying a mutant apolipoprotein E gene inactivated by gene targeting in embryonic stem cells," *Proceedings of the National Academy of Sciences,* vol. 89, no. 10, pp. 4471-4475, 1992.
[13] S. H. Zhang, R. L. Reddick, J. A. Piedrahita, and N. Maeda, "Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E," *Science,* vol. 258, no. 5081, pp. 468-471, 1992.
[14] C.-L. Lee *et al.*, "Assessing cardiac injury in mice with dual energy-microCT, 4D-microCT, and microSPECT imaging after partial heart irradiation," *International Journal of Radiation Oncology* Biology* Physics,* vol. 88, no. 3, pp. 686-693, 2014.
[15] S. Mukundan Jr *et al.*, "A liposomal nanoscale contrast agent for preclinical CT in mice," *American Journal of Roentgenology,* vol. 186, no. 2, pp. 300-307, 2006.
[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234-241: Springer.
[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
[18] H. Gao, H. Yu, S. Osher, and G. Wang, "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)," *Inverse Problems,* vol. 27, no. 11, p. 115012, 2011.
[19] D. P. Clark and C. T. Badea, "Hybrid spectral CT reconstruction," *PLOS ONE,* vol. 12, no. 7, p. e0180324, 2017.
[20] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on communications,* vol. 31, no. 4, pp. 532-540, 1983.
[21] K. Stierstorfer, A. Rauscher, J. Boese, H. Bruder, S. Schaller, and T. Flohr, "Weighted FBP—a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch," *Physics in Medicine & Biology,* vol. 49, no. 11, p. 2209, 2004.
[22] D. P. Clark and C. T. Badea, "GPU-Based Tools for Multi-Channel X-ray CT Reconstruction," in *The Fifth International Conference on Image Formation in X-Ray Computed Tomography*, Salt Lake City, Utah, 2018.
[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing,* vol. 13, no. 4, pp. 600-612, 2004.
[24] C. T. Badea, D. P. Clark, M. Holbrook, M. Srivastava, Y. Mowery, and K. B. Ghaghada, "Functional imaging of tumor vasculature using iodine and gadolinium-based nanoparticle contrast agents: a comparison of spectral micro-CT using energy integrating and photon counting detectors," *Physics in medicine and biology,* 2019.
[25] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, 2013, vol. 30, no. 1, p. 3.

Low-dose CT Denoising with Convolutional Neural Network for Unknown Noise Levels

Zhitao Guo[1,2], Hengyong Yu[2,*]

[1] College of Electronic Information Engineering, Hebei University of Technology, Tianjin, 300401, China
[2] Department of Electrical and Computer Engineering, University of Massachusetts, Lowell, MA, 01854, USA
*Corresponding author, Email: hengyong-yu@ieee.org

***Abstract*— Low-dose computed tomography (LDCT) can reduce the risks of radiation, but it will lead to significantly degraded image quality. How to improve the LDCT image quality is a hot topic in the medical imaging field. In this paper, we propose a denoising method for LDCT images via deep convolutional neural network (CNN). Different from the existing LDCT image post-processing models which usually train a specific model at a certain low-dose level, our model is able to handle the LDCT image restoration with unknown dose levels and distributions. Experimental results demonstrate a competing performance of the proposed method.**

***Index Terms*—Low-dose CT, noise reduction, deep learning, convolutional neural network**

I. INTRODUCTION

X-ray computed tomography (CT) is one of the most valuable imaging techniques in clinical diagnosis and interventional tasks. The increasing number of CT scans attracts the public concerns due to the potential cancer risks of x-ray radiation [1]. To reduce the x-ray radiation as low as reasonably achievable, the low-dose CT (LDCT) techniques have been extensively studied, and the reduction of x-ray photon number using tube current modulation is considered as one of the solutions. However, a drawback of this approach is the significantly degraded image quality due to the excessive quantum noise caused by an insufficient number of photons.

Till now, many methods have been developed to improve LDCT image quality. These methods generally can be categorized into three classes: sinogram filtration before reconstruction, iterative reconstruction, and image post-processing after reconstruction. Sinogram filtering directly smoothens raw projection data before image reconstruction (e.g. filtered backprojection method). The noise characteristic is well modeled for the sinogram-domain filtration. However, this kind of method often is limited because sinogram data of commercial CT scanners usually are not readily available to users and it may suffer from low resolution and edge blurring.

Iterative algorithms (IR) have a great potential in noise reduction and detail preservation for LDCT image reconstruction, aided by some kinds of prior information on the imaging object to be reconstructed. In the last decades, it attracted lots of research interests to search appropriate prior information. In the CT field, popular image priors include total variation (TV) minimization and its variants [2],[3], as well as dictionary learning based sparse representation. These iterative reconstruction method greatly improved the image quality. However, they suffer from a high computational cost with repeated projection/backprojection operations, which is a bottleneck in practical applications.

Differently from the sinogram denoising and iterative reconstruction, image post-processing methods directly operate on reconstructed images rather than relying on the raw projection data. It is underlined that the noise in LDCT images is different from a common noise distribution. The CT-specific noise is more complicated with streaking artifacts due to photon starvation and beam hardening. Extensive methods were developed to suppress artifacts and improve LDCT image quality via post-processing after the image reconstruction. Non-local means (NLM) methods take the advantages of the feature similarity and achieve high level performance for LDCT image denosing. As NLM extensions, the block-matching 3D (BM3D) algorithm [4] and weighted nuclear norm minimization (WNNM) algorithm [5] have been proved efficient in image restoration for different noise types. One drawback of this kind of approach is that the prior knowledge of the noise level is assumed to be known. However, the noise level and distribution of an actual LDCT image is often unkown.

In the image processing field, denoising and restoration algorithms using deep learning have been intensively investigated and have shown impressive performance. In the denoising framework based on convolutional neural network (CNN), the parameters of network are trained by supervised learning using large training datasets. The trained CNN can keep abundant priors with a complex hierarchy framework and remove noises from the test datasets. The recent development of net units (e.g. rectified linear unit (RelU), maxpooling, dropout and batch normalization) mitigates the risks of gradient vanishing or exploding [6]. Some flexible and complex networks achieve the challenged effects with less time compared to the classical algorithms. Zhang *et al* purposed a DnCNNs to deal with the blind Gaussian noise via training and achieved exciting results [7].

Inspired by the success of the CNN, we propose a LDCT denoising framework to detect and remove CT-specific noise patterns [8]. Specifically, we generate training sets with various CT-specific noise levels through numerical simulations, leaving the enhancement of ability to transform

LDCT images towards the corresponding normal-dose CT images. The rest of this paper is organized as follows. The proposed method is described in Section II. The experiments and results are presented in Section III. Finally, the conclusion is drawn.

Ⅱ. METHODS

A. Noise reduction model

Let $\mathbf{X} \in \mathbb{R}^{M\times N}$ be a low-dose CT image and $\mathbf{Y} \in \mathbb{R}^{M\times N}$ be the corresponding normal-dose CT image (NDCT). Then the relationship can be formulated as:

$$\mathbf{X} = \boldsymbol{G}(\mathbf{Y}), \tag{1}$$

where $\mathbf{G}: \mathbb{R}^{M\times N} \to \mathbb{R}^{M\times N}$ represents the corrupting process due to the quantum noise that contaminates the normal-dose CT image. Thus, the denoising issue can be converted to find a functional f:

$$\underset{f}{\operatorname{argmin}} \left|\left| f(\mathbf{X}) - \mathbf{Y} \right|\right|_2^2, \tag{2}$$

where f is treated as the best approximation of $\boldsymbol{G}^{-1}$. The CT-specific noise is different from natural image due to the combination of Poisson quantum noise and Gaussian electronic noise in x-ray photon measurements. The process of reconstruction makes the noise model more complicated and non-uniformly distributed across the whole image. What's more, streaking artifacts usually appear in low-dose CT because the lower energy X-ray photons are seldom received by detectors in lose-dose CT. All these problems are ignored in CNN-based methods. This is because this kind of learning based methods are strongly dependent on the training samples instead of noise type.

B. Convolutional neural network

The convolutional neural network, inspired by the neural network of the visual cortex in animals, is a special case of an artificial neural network. Similar to typical neural networks, it consists of successive linear and nonlinear functions, but the linear parts are specifically designed.

1) CNN model

The general form of CNN output y is expressed as

$$y = F(\Theta, x)$$
$$= f_L(W_L f_{L-1}(\cdots(f_2(W_2 f_1(W_1 x + b_1) + b_2)\cdots) + b_L), \tag{3}$$

where x is the input, y is the output, W_l is the convolution matrix of the l^{th} $(l = 1,2,\dots,L)$ layer, b_l is the bias of the l^{th} convolution layer, f_l is a nonlinear function, and Θ is the set of all tunable parameters including W_l and b_l. The goal of the CNN framework is to find an optimal parameter set Θ with K inputs to minimize the empirical loss:

$$\sum_{k=1}^{K} L_s(y_k, F(\Theta, x_k)). \tag{4}$$

In Eq.(4), x_k and y_k respectively denote the k^{th} input and output, and L_s typically denotes the cross-entropy loss in classification problems or the Euclidean distance in regression problems such as image denoising. Assuming that all nonlinear functions and the loss functions are differentiable, the minimization problem or network training in Eq. (4) can be addressed by an error back-propagation method. In general, the performance of the deep network is determined by the network architecture and methods that overcome the overfitting problem.

2) Network architecture

The input of our network is a series of LDCT images, which can be viewed as a set of noisy observations x_k. Residual learning formulation is adopted to train a residual mapping $F(\Theta, x_k) = y_k$, where y_k is the difference between the LDCT and NDCT images. Formally, the loss function is defined as the averaged mean squared error between the desired residual images and estimated ones from noisy input.

$$\ell(\Theta) = \frac{1}{2K}\sum_{k=1}^{K} \left|\left| F(x_k) - y_k \right|\right|. \tag{5}$$

Figure 1 illustrates the architecture of the proposed network to learn $F(\Theta, x_k)$. In the following, we explain the details of architecture and parameters.

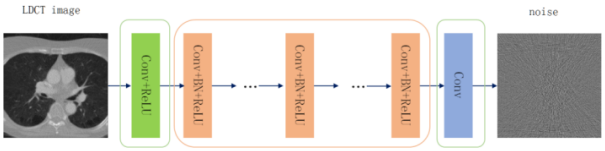


Figure1.The architecture of the purposed network.

There are three types of layers with a network depth D, shown in Figure. 1. (i) Sparse representation layer: Sparse representation of LDCT is implemented in the first layer via convolution operations with a series of filters. 64 filters of size 3×3 are used to generate 64 feature maps, and ReLUs are then utilized for nonlinearity. (ii) Non-linear filtering layers: for layers 2~(*D*−1), 64 filters of size 3×3×64 are used, and batch normalization is added between convolution and ReLU. (iii) Reconstruction layer: for the last layer, a filter of size 3×3×64 are used to reconstruct the output.

Meanwhile, our network model has two main features: the residual learning formulation is adopted to learn $F(\Theta, x)$, and batch normalization is incorporated to speed up training as well as boost the denoising performance. By incorporating the convolution with ReLU, the proposed network can gradually separate image structure from the noisy observation through the hidden layers.

We construct two kinds of CNNs. One is for a specific noise level (CNN-S), and the other is for unknown noise level (CNN-U). The depth of CNN-U is set as 20 which is greater than 15 layers of CNN-S, considering the depth of CNN improving the expression ability of the network.

3) Network training

Network training was performed by minimizing the loss function (5). Minimization of the cost function was performed by means of conventional error back-propagation with the Adaptive Moment Estimation (Adam) and the gradient clipping method. The convolution kernel weights were initialized using random Gaussian distributions. The gradient

clipping method in the range of $[10^{-3},10^{-4}]$ was used to facilitate the use of a high learning rate in the initial training steps. This allows rapid convergence and avoids the gradient explosion problem. If the learning rate is high, the speed of convergence is fast. However, this introduces the gradient explosion problem. When we attempted to find an optimal environment setting for a stable learning process, we found that the gradient should be held within the range of $[-10^{-3},10^{-3}]$ to prevent gradients from exploding.

Ⅲ. EXPERIMENTS

To evaluate the performance of the proposed approach, the state-of-the-art methods were selected for comparison. They include the TV minimization based iterative reconstruction and image post-processing approach WNNM. The parameters for the completing methods were set according to the recommendations in the original references. The peak signal to noise ratio (PSNR) and structure similarity index (SSIM) were used as the quantitative metrics. All the experiments were conducted using MATLAB 2019b on a PC (Intel i5 9600K CPU, 32 GB RAM and RTX 2080 graphics card).

A. Dataset preparation

In 2019, Leuschner *et al*. built a standard database for low-dose CT, named LoDoPaB-CT, which can be used to serve as a benchmark to train networks of low-dose CT reconstruction methods [9]. Over 40,000 paired samples are created with size of 362x362 from around 800 patients by simulating low-dose CT measurements. The database constitutes the most complete training data and could be used for all kinds of learning. In order to obtain CT images with different noise levels reflecting different levels of dose, we preprocess the projection data by simulating the number of photons and adding Poisson noise. Specifically, assuming that N_0 photons are emitted from the x-ray source towards each detector aperture, we can simulate the Poisson noise inherent with the normalized projection data (p). The expected photon number arriving at the detector can be calculated as $\overline{N} = N_0 \times \exp(-p)$, obeying Lambert-Beer law. is the X-ray linear attenuation coefficient for water. The noisy projection data is $\widetilde{p}=(\ln(N_0/\widehat{N})$, here $\widehat{N}$ is Poisson random variable whose mean and variance equal to $\overline{N}$ [10]. Finally, FBP reconstruction is performed on the noise projection data to obtain the corresponding LDCT images. Table I gives the PNSR and SSIM for different low-dose levels. The representative reconstruction images via FBP are shown in figure 2.

Table I: Dose ratios and SNRs for different low-dose levels.

N_0	10^7	10^6	10^5	10^4	10^3
PSNR	38.361	38.2944	37.682	34.912	26.114
SSIM	0.975672	0.97371	0.95529	0.81474	0.40139

We found that using a larger training dataset can only bring little improvement. To train a network for specific dose level (i.e. $N = 10^4$), we set the patch size as 32x32 and crop 1,454,592 patches to train the model with image operations of scale, rotation, mirror and so on. To feed different CNNs, two kinds of training datasets are generated, specific datasets and mixed datasets. The former is based on single noise level images, while the latter is composed by randomly extracted data of different noise levels.

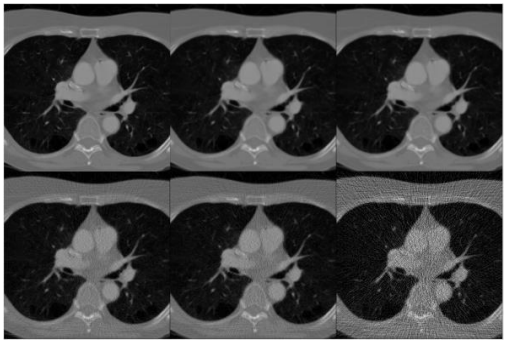


Figure 2. Illustration of degraded images with different low-dose levels. From top to bottom and from left to right, the images are the ground truth and with different noise levels for N=10^7, 10^6, 10^5, 10^4, and 10^3.

In the test data set, different levels of simulated low-dose data are established according to the LoDoPaB database for testing. Meanwhile, real quarter-dose fan-beam CT data from 2016 AAPM [11] are used to evaluate the real application effect. For fairness, 128 low-dose samples were randomly selected from the original dataset to form the testing set.

B. Visualization-based evaluation

Some representative LDCT images are select from LoDoPaB and AAPM to evaluate the performance in data simulation and real data. The iterative TV minimization algorithm and NLM based WNNM algorithm are used as comparison. Figure 3 shows the representative results from different methods with an example selected in datasets of LoDoPaB, and CNN-S is applied in this comparison. The noisy image was reconstructed by FBP from noisy projection with Poisson noise parameter $N_0 = 10^4$, because the noise is not obvious when the parameter N_0 is greater than 10^5 and the image degradation with $N_0=10^3$ is too serious. In Figure 4, the corresponding noisy example is a real quarter-dose image. Both of the two original images have serious noise and artifacts which degraded the reconstructed image quality. Some details and structures cannot be discriminated. Both the TV and WNNM can suppress the noise and artifacts to different degrees. However, the denoised image appears blocky with the edge damaging for TV method. The WNNM could not eliminate the noise and artifacts clearly, and it also erase the image details. From the zoom parts, the detail restoration can be compared. By combining Figs. 3 and 4, one can see that the purposed CNN-S method can achieve the best image quality.

C. Quantitative evaluation

To quantitatively evaluate the proposed CNN algorithm, PSNR and SSIM were measured for all the images in the the testing set. The test set included 128 samples in LoDoPaB and 128 samples in AAPM databases, respectively. We compared the results of CNN-S(N_0=10^4), CNN-U and two benchmark algorithms using different noise level data sets in the LodoPaB and AAPM. We record the mean values of PSNR and SSIM in Table II. For the LoDoPaB test set, we can find that both CNN-S and CNN-U are better than TV and WNNM in denoising performance. CNN-S gets the best effect at the same-noise

level as expected, but it is inferior to CNN-U in other noise levels. CNN-U shows stable restoration effect and strong robustness at all noise levels. In the actual test of AAPM dataset, the performance of CNN-U is close to that of CNN-S, which are obviously better than the other two algorithms.

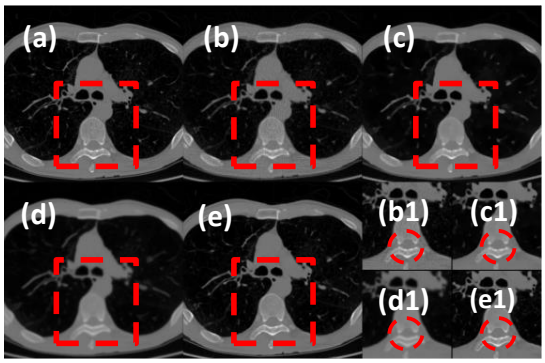


Figure 3. Representative results of a LDCT image of LoDoPaB datasets. (a) Original normal-dose image; (b) the low-dose images; (c) the TV reconstructed image; (d) the WNNM processed low-dose image; (e) the purposed CNN-S processed low-dose image; (b1)-(e1) the zoomed regions specified with the red boxes in (b)-(e).

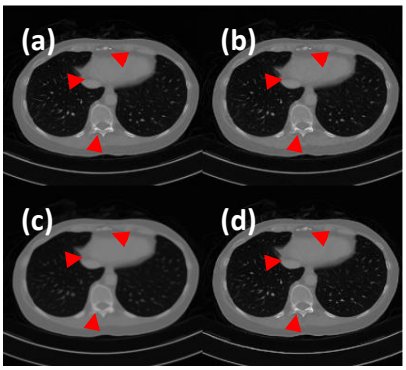


Figure 4. Representative results of a LDCT image of AAPM datasets. (a) Original normal-dose image; (b) the low-dose images (c) the WNNM processed low-dose image; (d) the purposed CNN-S processed low-dose image.

Table II. Quantitative evaluation associated with different algorithms.

		TVIR	WNNM	CNN-S ($N=10^4$)	CNN-U
LoDoPaB 10^6	PSNR	43.164	35.696	43.466	**43.572**
	SSIM	0.9902	0.9774	0.9901	**0.9950**
LoDoPaB 10^5	PSNR	41.704	35.635	42.764	**42.965**
	SSIM	0.9774	0.9557	0.9740	**0.9779**
LoDoPaB 10^4	PSNR	38.256	35.562	**40.265**	40.167
	SSIM	0.9750	0.8750	**0.9789**	0.9770
AAPM	PSNR	38.488	33.690	40.568	**41.678**
	SSIM	0.9530	0.9026	0.9691	**0.9693**

IV. CONCLUSION

In this paper, we propose and evaluate a deep convolutional neural network for noise reduction in low-dose CT. The framework convergence is guaranteed from the mathematical model design. The results demonstrate the capability of CNN based method for medical imaging denoising even the noise level is unknown. This is because the complex expression ability of neural network can form a unified processing of image denoising with different noise levels. A mixed image set with different noise levels is generated to train such a CNN-U. In the future, we will combine the proposed network structure with the iterative algorithm to further improve the image quality, and we will also apply the proposed CNN network in CT reconstruction to solve the problems of metal artifact reduction and material decomposition in spectral CT.

REFERENCES

[1] D. J. Brenner and E. J. Hall, “Computed tomography–An increasing source of radiation exposure,” New Eng. J. Med., vol. 357, pp. 2277–2284, 2007.

[2] Sidky EY, Pan X. “Image reconstruction in circular cone-beam computed tomography by constrained, total- variation minimization , ” Physics in Medicine & Biology. , vol.53, no.17,pp.4777,2008.

[3] Liu Y, Ma J, Fan Y, Liang Z. "Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction," Physics in Medicine & Biology., vol.57,no. 23, pp.7923,2012

[4] Feruglio PF, Vinegoni C, Gros J, Sbarbati A, "Block matching 3D random noise filtering for absorption optical projection tomography," Physics in Medicine & Biology, vol.33,no.18, pp.5401, 2010.

[5] S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2862–2869, 2014.

[6] Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. Proceeding of the 27th International Conference on Machine Learning (ICML-10),. vol. 1.pp.807–814, 2010.

[7] Zhang, K., Zuo, W., Chen, Y., Meng, D. and Zhang, L, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising". IEEE Transactions on Image Processing, vol.26, no.7, pp.3142-3155, 2017.

[8] Yang, Q., Yan, P., Zhang,.and Wang, G, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," IEEE transactions on medical imaging, vol.37,no.6, pp.1348-1357,2018.

[9] Leuschner, Johannes and Schmidt, "The LoDoPaB- CT Dataset: A Benchmark Dataset for Low-Dose CT Methods".. arXiv:1910.01113, 2019.

[10] H.Yu, Y.Ye, S.Zhao, and G.Wang, "Local ROI Reconstruction via Generalized FBP and BPF Algorithms along More Flexible Curves," International Journal of Biomedical Imaging, pp.1-7,2006.

[11]Kang, E., Min, J. and Ye, J.C.," A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," Medical physics, vol.44,no.10, pp.e360-e375, 2017.

Investigating the Utility of Virtual Monoenergetic Images in Non-Contrast Lung CT Scans

T. W. Holmes, C. N. De Cecco, and A. Pourmorteza

***Abstract*—Background: Spectrally-derived virtual monoenergetic images have traditionally been used to maximize the CNR of iodine and minimize beam hardening artifacts in contrast-enhanced CT scans. In this study we explored the utility of VMIs in improving the CNR of ground glass nodules in non-contrast lung cancer screening protocols.**

Methods: We scanned the COPDGENE lung phantom at 48 different settings of x-ray tube voltage and current settings which include single- and dual-energy acquisition modes with and without tin pre-filtration. We measured the CNR of different inserts in the phantom which mimic different healthy and pathological tissue types found in the lungs.

Results: Single-energy tube voltage set at 100 kVp with tin pre-filtration (100Sn) showed highest CNR for ultra-low-dose CT scans ($CTDI_{vol}$ = 0.1 to 1 mGy). For all dose-matched scans above 1.0 mGy, the VMI at 190 keV provided the highest CNR for detection of GGN.

Conclusion: Our preliminary results show that dual-energy derived VMI can improve the detection of GGN in low- and high-dose lung CT scans.

I. Introduction

VIRTUAL monoenergetic images (VMIs) are spectrally derived images that provide a map of mass attenuation coefficients at specific energies of x-ray photons. Spectrally-derived images have been extensively investigated in contrast enhanced CT acquisition protocols ([1]–[5]). However, their utility in non-contrast scans has not been thoroughly explored. In contrast-enhanced scans, low kiloelectron volt-potential (kVp) acquisitions have been suggested in order to increase the CNR of iodine. They were shown to provide better or comparable contrast-to-noise ratios (CNRs) to 120 kVp acquisitions [6]. However, their utility in non-contrast scan tasks remains to be explored. Additionally, spectral beam-shaping with tin (Sn) pre-filters have been proposed to reduce radiation dose to patients in low-dose non-contrast screening protocols [7]–[9]. The aim of this study was to compare the performance of various x-ray tube voltages, Sn beam pre-filtration, and dual-energy VMI protocols in a dedicated lung phantom.

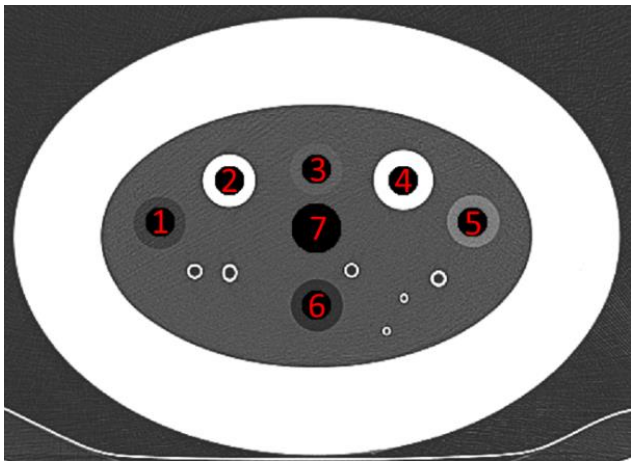


Figure 1. CT scan of the COPDGENE 2 phantom used in this study. The scan was performed at 120 kVp and 300 mAs resulting in $CTDI_{vol}$ = 20.06 mGy. Visualization window width and level were 500/-700 HU. The black circles represent the regions of interest (ROIs) used in this study: 1, Foam NIST 4lb; 2, Acrylic; 3, Foam NIST 12lb; 4, Water; 5, Foam NIST 20lb; 6, Air; 7, Simulated Lung Density Material (SLDM).

II. Materials and Methods

We used the COPDGENE Phantom II (The Phantom Laboratory, Salem, NY, USA) in this study. The phantom includes inserts with calibrated materials which produce HU values that mimic the lung parenchyma (simulated lung density material: SLDM), two NIST calibrated foams representing ground-glass nodule (GGN) densities of 12lb. and 20lb., as well as water, air, and acrylic.

We performed a series of CT scans at various tube voltage and current settings as described in Table 1. All scans were performed on a third generation dual-energy scanner

T. W. Holmes is with the Radiology and Imaging Sciences Department at Emory University, 1701 Uppergate Dr, Suite 5018A, Atlanta, GA, 30322 (e-mail: twholmes@emory.edu).

C. N. De Cecco is with the Radiology and Imaging Sciences Department and the Department of Biomedical Informatics at Emory University (e-mail: carlo.dececco@emory.edu)

A. Pourmorteza is with the Radiology and Imaging Sciences Department, the Winship Cancer Institute, and the Department of Biomedical Engineering at Emory University, 1701 Uppergate Dr, Suite 5018A, Atlanta, GA, 30322 (e-mail: apourmo@emory.edu).

TABLE 1
CT acquisition settings and corresponding radiation doses for the single- and dual-energy scans.
The x-ray tube power could not produce mAs of higher than 770 in the 100Sn mode.
*Reported mAs belongs to the low kVp setting

Voltage [kVp]	***Current [mAs]***										
$CTDI_{vol}$ [mGy]	0.10	0.15	0.20	0.33	0.68	1.02	1.34	3.34	6.69	10.03	20.06
100 Sn	24	35	47	78	158	-	316	770	-	-	-
Dual-Energy: 100/150Sn*	-	-	-	-	-	18	24	59	118	178	335
140	-	-	-	-	7	-	13	34	68	103	206
120	-	-	-	5	10	-	20	50	100	150	300
100	-	-	-	8	16	-	33	83	167	250	500
80	-	-	10	17	35	-	70	174	348	522	1044
70	-	-	17	28	58	-	114	285	572	857	-

(SOMATOM FORCE, Siemens Healthcare, Forchheim, Germany). The scans consisted of one Sn pre-filtered protocol at 100 kVp (100Sn), a dual-energy protocol (100/150Sn), and five single-energy protocols.

We performed the scans at multiple tube current-time settings and used CT dose index by volume ($CTDI_{vol}$) values to match the radiation dose among all scan protocols when possible. These scans were acquired in helical mode with a pitch of 1.2 and z-collimation: 96 x 0.6 mm. The images were reconstructed with weighted FBP, quantitative Qr40 kernel, matrix size = 512x512, FOV=370 mm, and slice thickness / increment of 2 mm / 1 mm. No automatic exposure control was used to ensure a fair comparison of all the different tube voltage settings.

We placed circular regions of interest (ROIs) with diameters of at least 17 mm inside each tissue mimicking material of the center slice as shown in figure 1. Mean and standard deviation of the ROIs were measured and used to CNR between selected ROIs and the background lung (SLDM).

III. Results

A total of 48 scans were performed as shown in Table 1. All data was reconstructed on the scanner workstation. The VMIs were created using the Monoenergetic+ algorithm of Syngo.Via™ post-processing software (Siemens Healthcare, Forchheim, Germany) [5], [10], [11]. We used software written in Matlab to place co-registered ROIs across all images and calculate their mean, standard deviation, and CNR. Table 2 shows the mean Hounsfield units (HU) of select ROIs and table 3 summarizes the CNR measurements for various ROIs. The maximum radiation dose output of the 100Sn tube voltage setting was 3.34 mGy compared to > 20.06 mGy for most of the other tube settings. However, the 100Sn setting could provide radiation dose outputs as low as 0.1 mGy. In addition, figure 2 illustrates CNR – $CTDI_{vol}$ curves for all single-energy acquisition protocols. The 100Sn setting consistently showed higher CNR for all single-energy scans with $CTDI_{vol}$ < 3.34 mGy followed by the 80 kVp setting. At $CTDI_{vol}$ settings of > 5 mGy, the 150Sn setting resulted in the highest CNR values.

Figure 3 illustrates the mean Hounsfield values and CNR curves of the VMIs in the energy range of 40 to 190 keV. We observed a steady increase in the CNR up to 80 keV followed by a slow increase in CNR to 190 keV. In almost all energies and dose settings, the 190 keV image provided the best CNR

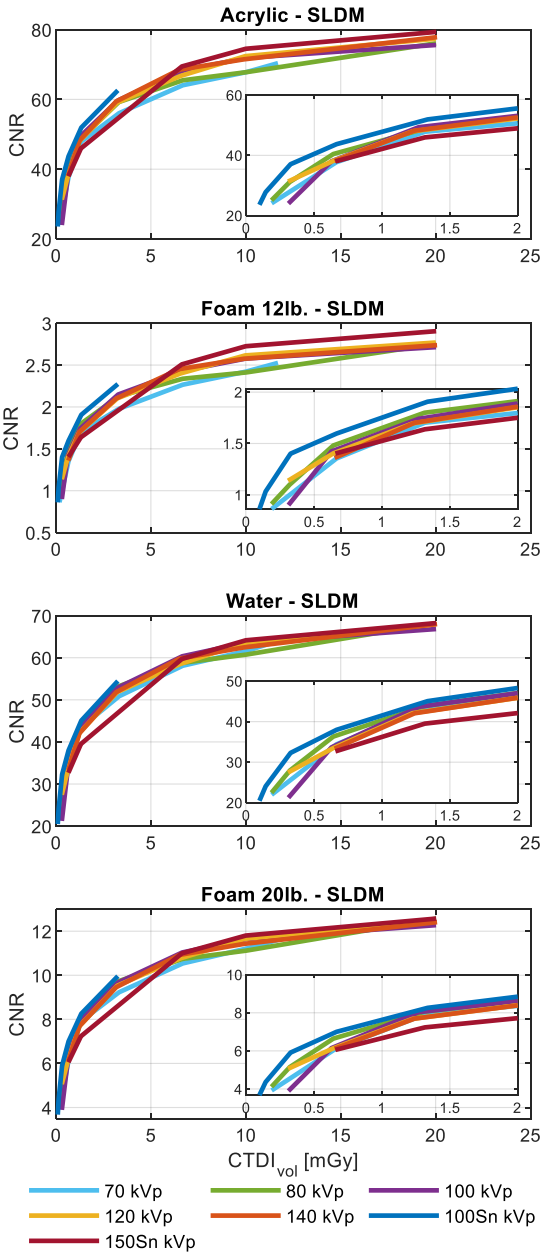


Figure 2. The CNR between acrylic, foam NIST 12lb., water, and foam NIST 20lb. vs. simulated lung density material (SLDM). The CNR values at each voltage potential were calculated over a range of doses from 0.1 mGy to 20.06 mGy.

for GGN, water, and acrylic materials.

In all dose-matched settings, VMI at 190 keV consistently had the highest CT numbers and CNR values, with little difference in performance for VMIs with energies > 80 keV. Figure 4 shows CT number and CNR values for all scans performed at $CTDI_{vol}$ of 1.34 mGy.

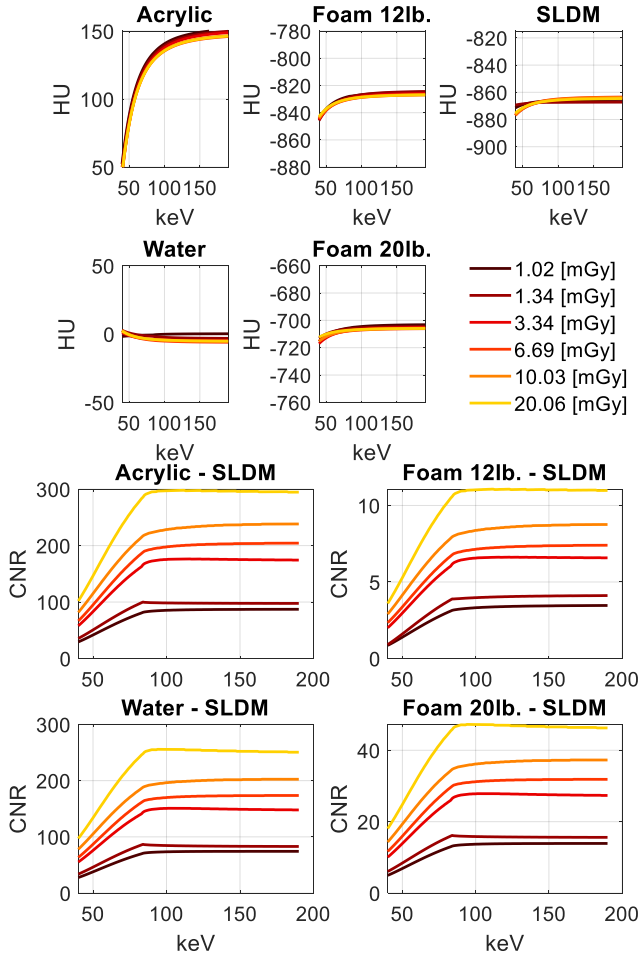


Figure 3. The mean CT numbers and CNR values measured in the virtual monoenergetic images across the energies ranging from 40 to 190 keV for radiation doses ranging between 1.02 to 20.06 mGy.

IV. Discussion

Low-dose lung cancer screening CT scans are usually performed at 120 kVp with $CTDI_{vol}$ of 1.5-3 mGy. In this study, we performed a sweep of the tube voltage and current parameters to compare performance of single-energy acquisition modes to that of the dual-energy CT for non-contrast CT acquisitions in the range of $CTDI_{vol}$ values from 0.1 to 20.06 mGy. Our results indicate that the 100 kVp with Sn pre-filtration provided superior CNR for detection of ground glass nodules at very low dose settings (0.10-1.0 mGy). The lowest possible dose setting for dual-energy acquisition mode was 1.03 mGy. The CNR for GGN detection of the dual-energy derived VMIs was superior to all single-energy modes, including the 100Sn setting, for all $CTDI_{vol} \geq 1.03$ mGy with 190 keV demonstrating the highest CNR.

Given that VMIs have been traditionally used to maximize the CNR of iodinated contrast agents, the results of this study indicate the utility of VMIs in non-contrast CT applications.

While the results of this study were noteworthy, this was a preliminary study with many limitations. This investigation only used simple image quality metrics and large ROIs within each of the materials. In future studies we plan to investigate modulation transfer function (MTF) and noise power spectrum (NPS) of the different scan protocols. We used filtered back-projection (FBP) reconstruction algorithm to reduce the shift-variant effects of nonlinear and statistical reconstruction algorithms. While such algorithms may reduce the standard deviation of image noise, their effect on CT numbers of lung tissue remains to be investigated. Lastly, we scanned a geometrically, simple phantom at one patient size and only single measurements were taken at each acquisition configuration. Similar analysis should be conducted on anthropomorphic phantoms of various sizes, and each scan should be performed at least twice to ensure a bias-free measurement of noise.

The dual-energy images could be combined to create images with minimum noise, or maximum contrast as described in [2]. We plan to perform similar analysis; however, our results suggest that VMIs between 80-190 keV image seems to provide the best CNR for GGN detection.

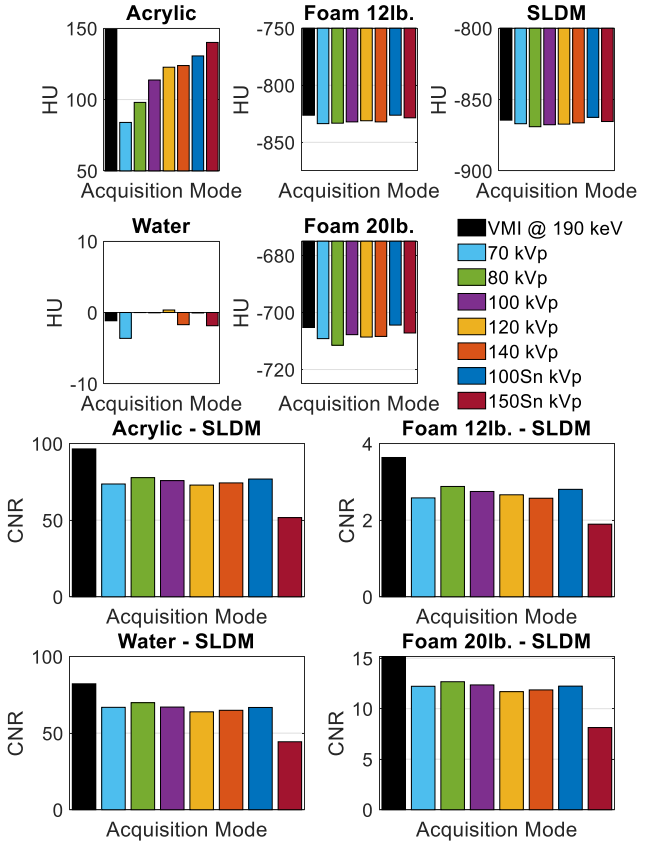


Figure 4. Mean CT numbers and CNR values for dose-matched scans all with $CTDI_{vol}$ of 1.34 mGy. Virtual monoenergetic images consistently had the densest CT numbers, and highest CNRs for all ROIs.

TABLE 2
The mean Hounsfield unit (HU) in acrylic, foam 12lb., water, foam 20lb., and background lung (SLDM) at three different radiation doses.

ROI	***CTDI$_{vol}$***	*VMI @ 190 [keV]*	*70[kVp]*	*80[kVp]*	*100[kVp]*	*120[kVp]*	*140[kVp]*	*100Sn[kVp]*	*150Sn[kVp]*
Acrylic	1.34	149.16	83.96	98.00	113.71	122.69	123.84	130.60	140.07
	3.34	148.82	85.95	98.25	111.89	119.28	125.57	127.16	137.74
	10.03	146.62	82.99	97.53	112.33	118.53	123.40	-	137.01
Foam 12lb.	1.34	-826.26	-833.61	-833.22	-832.07	-831.09	-832.08	-826.23	-828.52
	3.34	-825.84	-833.91	-833.92	-832.28	-831.55	-830.78	-826.09	-828.28
	10.03	-826.65	-833.97	-835.08	-832.75	-831.45	-830.87	-	-827.33
Water	1.34	-1.16	-3.63	-0.01	-0.04	0.35	-1.71	-0.06	-1.86
	3.34	-4.83	-3.51	-2.19	-0.06	-2.22	-1.90	-2.16	-4.20
	10.03	-5.42	-5.98	-2.87	-0.58	-2.59	-2.52	-	-3.26
Foam 20lb.	1.34	-705.26	-709.16	-711.49	-707.77	-708.61	-708.38	-704.41	-707.18
	3.34	-705.16	-711.00	-709.63	-708.75	-707.61	-708.29	-705.50	-707.07
	10.03	-705.72	-710.91	-711.11	-709.03	-708.29	-708.17	-	-706.21
SLDM	1.34	-864.45	-866.95	-869.05	-867.67	-867.24	-866.36	-862.47	-865.41
	3.34	-864.05	-867.03	-869.25	-866.94	-867.18	-866.07	-862.39	-864.94
	10.03	-863.77	-868.15	-869.53	-867.41	-867.58	-866.87	-	-864.16

TABLE 3
The CNR between acrylic, foam 12lb., water, and foam 20lb. compared with background lung (SLDM) at three different radiation doses.

ROI	***CTDI$_{vol}$***	*VMI @ 190[keV]*	*70[kVp]*	*80[kVp]*	*100[kVp]*	*120[kVp]*	*140[kVp]*	*100Sn[kVp]*	*150Sn[kVp]*
Acrylic - *SLDM*	1.34	96.53	73.63	77.76	75.82	72.92	74.34	76.87	51.66
	3.34	174.33	111.80	122.98	114.69	112.11	116.71	125.65	77.86
	10.03	238.30	180.95	183.90	188.30	192.81	201.74	-	129.29
Foam 12lb. - *SLDM*	1.34	3.64	2.58	2.88	2.75	2.66	2.57	2.81	1.90
	3.34	6.58	3.89	4.49	4.06	4.05	4.15	4.61	2.85
	10.03	8.75	6.50	6.55	6.66	7.06	7.33	-	4.76
Water - *SLDM*	1.34	82.22	66.84	69.88	67.03	63.91	64.91	66.75	44.37
	3.34	147.89	101.30	110.22	101.57	98.31	101.71	109.23	66.83
	10.03	202.44	164.02	164.81	166.60	169.13	176.08	-	111.18
Foam 20lb. - *SLDM*	1.34	15.16	12.22	12.67	12.35	11.69	11.86	12.23	8.13
	3.34	27.35	18.30	20.29	18.54	18.14	18.57	19.92	12.26
	10.03	37.28	29.91	30.13	30.44	31.14	32.33	-	20.40

REFERENCES

[1] F. F. Behrendt *et al.*, "Image fusion in dual energy computed tomography: effect on contrast enhancement, signal-to-noise ratio and image quality in computed tomography angiography," *Invest. Radiol.*, vol. 44, no. 1, pp. 1–6, 2009.

[2] L. Yu, A. N. Primak, X. Liu, and C. H. McCollough, "Image quality optimization and evaluation of linearly mixed images in dual-source, dual-energy CT," *Med. Phys.*, vol. 36, no. 3, pp. 1019–1024, 2009.

[3] J. C. Schenzle *et al.*, "Dual energy CT of the chest: how about the dose?," *Invest. Radiol.*, vol. 45, no. 6, pp. 347–353, 2010.

[4] S. Mangold *et al.*, "Virtual nonenhanced dual-energy CT urography with tin-filter technology: determinants of detection of urinary calculi in the renal collecting system," *Radiology*, vol. 264, no. 1, pp. 119–125, 2012.

[5] V. Sandfort *et al.*, "Optimized energy of spectral CT for infarct imaging: Experimental validation with human validation," *J. Cardiovasc. Comput. Tomogr.*, vol. 11, no. 3, pp. 171–178, 2017.

[6] L. Yu, J. A. Christner, S. Leng, J. Wang, J. G. Fletcher, and C. H. McCollough, "Virtual monochromatic imaging in dual-source dual-energy CT: radiation dose and image quality," *Med. Phys.*, vol. 38, no. 12, pp. 6371–6379, 2011.

[7] C. Thomas *et al.*, "Differentiation of urinary calculi with dual energy CT: effect of spectral shaping by high energy tin filtration," *Invest. Radiol.*, vol. 45, no. 7, pp. 393–398, 2010.

[8] C. Li, B. Liu, H. Meng, W. Lv, and H. Jia, "Efficacy and Radiation Exposure of Ultra-Low-Dose Chest CT at 100 kVp with Tin Filtration in CT-Guided Percutaneous Core Needle Biopsy for Small Pulmonary Lesions Using a Third-Generation Dual-Source CT Scanner," *J. Vasc. Interv. Radiol.*, vol. 30, no. 1, pp. 95–102, 2019.

[9] K. Narikiyo *et al.*, "The evaluation of ultralow dose CT urography using Tin filter: Comparison with standard-dose CT urography.," 2019.

[10] D.-K. Jeong *et al.*, "Effects of energy level, reconstruction kernel, and tube rotation time on Hounsfield units of hydroxyapatite in virtual monochromatic images obtained with dual-energy CT," *Imaging Sci. Dent.*, vol. 49, no. 4, pp. 273–279, 2019.

[11] K. L. Grant, T. G. Flohr, B. Krauss, M. Sedlmair, C. Thomas, and B. Schmidt, "Assessment of an advanced image-based technique to calculate virtual monoenergetic computed tomographic images from a dual-energy examination to improve contrast-to-noise ratio in examinations using iodinated contrast media," *Invest. Radiol.*, vol. 49, no. 9, pp. 586–592, 2014.

An Investigation on Bayesian Image Reconstruction without Freely Adjustable Parameter

Zhengrong Liang, Yongfeng Gao, Yongyi Shi, and Wei Hou

***Abstract*—In the past decades, most penalized maximum likelihood or penalized weighted least-squares tomographic image reconstruction from projections, based on the Bayes' law, include a freely-adjustable parameter to balance the two basic terms of data fidelity and prior penalty for an acceptable result. The freely-adjustable parameter is determined empirically by a trial-error fashion in different applications. With recent machine learning advancement, the free parameter could be determined by multiple reconstruction trials, but demanding high computing power. A theory-based strategy is desired to determine the free parameter, and/or to guide machine learning for reduction of the computing complexity. This study investigates a theory-based strategy for parameter estimation in Bayesian image reconstruction without freely-adjustable parameters. A main idea is the introduction of parameters to ensure the condition of each probability density function (PDF) in the Bayes' Law is satisfied. Thus maximum *a posteriori* probability solution, given acquired data, shall satisfy the PDF condition for both data likelihood and *a priori* probability of the Bayes' Law. Computer simulation experiments were carried out to evaluate the performance of the parameter estimation under various conditions, such as initialization of the iterative process, the PDF specification, and renormalization in an iterative manner. The experimental outcomes reveal that the parameter estimation is relatively robust under these conditions. If the *a priori* probability cannot satisfy the PDF definition, this study provides some insights on weighting the *a priori* penalty for the above penalized image reconstructions.**

***Index Terms*—Tomographic image reconstruction, Bayesian inference, PDF, EM algorithm.**

I. Introduction

TOMOGRAPHIC image reconstruction from projections is mathematically expressed by the Radon transform [1]. The inversion of the Radon transform is well-established by the analytical filtered back-projection (FBP) method [2], which has been adopted by all commercial computed tomography (CT) vendors. In the photon-limited applications, such as positron emission tomography (PET), an alternative method for reconstructing the photon-limited or statistical projection data has shown some good properties as compared to the FBP method. This alternative method applies the well-known Expectation-Maximization (EM) algorithm [3] to iteratively estimate the maximum likelihood (ML) solution of the statistical projection data and is called MLEM method [4]. The MLEM method and its variate by adding ordered-subsets (OS) in the iterative update stage, called OSEM method, have also been adopted by all commercial PET vendors. In the statistical domain, given the statistical projection data, the desired solution shall be governed by the Bayes' Law or Bayesian theory, which includes the above likelihood (or data fidelity) term and an additional term, called prior penalty term. Both terms are expressed in terms of probability density function (PDF). This study applies a similar concept as the EM algorithm to specify PDF for both the data fidelity term and the prior penalty term to construct the posteriori probability as objective function and uses the local update (LU) strategy [5] to maximize the cost function for parameter estimation. The study is presented with implication for low-dose CT (LdCT) application.

This work was partially supported by the NIH/NCI Grant CA206171.

Z. Liang (corresponding author) is with the Departments of Radiology, Electrical & Computer Engineering, Computer Science, and Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794, USA (e-mail: Jerome.Liang@sunysb.edu).

Y. Gao and Y. Shi are with the Department of Radiology, Stony Brook University, Stony Brook, NY 11794, USA.

W. Hou is with the Department of Preventive Medicine, Stony Brook University, Stony Brook, NY 11794, USA.

II. Methods

This section starts by brief descriptions of the Bayesian theory and the PDF definition of the two basic terms of data likelihood and prior penalty, followed by presentation of the LU strategy to determine the parameters, called maximum *a posteriori* probability-LU (MAP-LU) method.

A. Bayesian Theory for LdCT Reconstruction

Bayesian theory is expressed, given acquired data { $\boldsymbol{y}$ } for the corresponding solution { $\boldsymbol{\mu}$ }, as:

$$p(\boldsymbol{\mu}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\mu})p(\boldsymbol{\mu})/p(\boldsymbol{y}) \quad (1)$$

where $p(.)$ represents the PDF of the concerned variable, vector $\boldsymbol{y} \in R^{I\times 1}$ represents the LdCT sinogram or projection data acquired from the body (with I data elements) and vector $\boldsymbol{\mu} \in R^{J\times 1}$ represents the body tissue attenuation coefficient map or image of the body (with J image elements).

The acquired data can be written as:

$$\boldsymbol{y} = \boldsymbol{A\mu} + \boldsymbol{\zeta} \quad (2)$$

where $\boldsymbol{A}$ is the projection operator, called projection or system matrix, and $\boldsymbol{\zeta}$ indicates the noise vector corresponding to the data vector. The PDF for the acquired data or data fidelity term can be written as:

$$p(\boldsymbol{y}|\boldsymbol{\mu}) = \frac{\exp\left(-\frac{1}{2\alpha}(\boldsymbol{y}-\mathbf{A}\boldsymbol{\mu})\Sigma_{\boldsymbol{y}}^{-1}(\boldsymbol{y}-\mathbf{A}\boldsymbol{\mu})^T\right)}{\sqrt{(2\pi\alpha)^I\,|\Sigma_y|}} \quad (3)$$

where Σ_y represents the covariance matrix for the acquired data vector, which is usually estimated based on some established knowledge and/or principles [6] up to a scale or parameter of α, and π is a constant, thus $\{\boldsymbol{y}, \boldsymbol{A}, \Sigma_{\boldsymbol{y}}\}$ are observable but incomplete (because only one set of data $\boldsymbol{y}$ is given, not sufficient to estimate the unknown $\boldsymbol{\mu}$), and the parameter α is introduced as a unobservable parameter [3] to ensure that the PDF is completely specified according to the Bayesian theory of Eq.(1). Under the condition that all measurements $\{y_i\}$ are independent each other, then the covariance matrix is diagonal. We can denote the diagonal element of *i*th ray of matrix $\Sigma_{\boldsymbol{y}}^{-1}$ as ϱ_i.

While the tissue attenuation coefficients are not random variables, their distribution patterns across the body or field of view (FOV) can be assumed as random variable and the patterns are frequently modeled as the Markov Random Field (MRF) [7]. Thus the prior penalty PDF for the pattern distribution can be expressed as a Gibbs distribution across the MRF [7]. Following the variance analysis for a pair-wise neighbor system below:

$$Var(\mu_j - \mu_k) = Var(\mu_j) + Var(\mu_k) + 2CoVar(\mu_j, \mu_k) = \sigma + \sigma + 0 = 2\sigma \quad (4)$$

the prior PDF can be written as:

$$p(\boldsymbol{\mu}) = \frac{1}{\sqrt{(2\pi\sigma)^J|\boldsymbol{\omega}_j{}^{-1}|}}\, exp\left(-\frac{1}{2\sigma}\sum_{j=1}^{J}\sum_{k\in\Omega_j}\omega_{jk}\left(\mu_j-\mu_k\right)^2\right) \quad (5)$$

where ω_{jk} represents the normalized pair-wise MRF coefficients among the neighbors Ω_j around the center *j*, $\sum_{k\in\Omega_j}\omega_{jk} = 1$, and can be estimated based on some established knowledge and/or principles [8] up to a scale or parameter of σ, thus $\{\omega_{jk}, \Omega_j\}$ are observable but incomplete, and the parameter σ is introduced as a unobservable parameter [3] to ensure that the prior PDF is completely specified for the pattern distribution across the FOV.

B. Parameter Estimation by the LU Strategy

Above we have obtained the explicit form of the conditional data probability of Eq.(3) and the prior probability of Eq.(5) with each symbol well defined. Bringing Eqs.(3) & (5) into the Bayesian theory of Eq.(1), the completely specified log *posteriori* PDF becomes [9]:

$$\Psi(\boldsymbol{\mu},\alpha,\sigma) = \log\{p(\boldsymbol{\mu}|\boldsymbol{y})\} = \frac{1}{2\alpha}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\mu}\|^2_{\Sigma_{\boldsymbol{y}}^{-1}} + \frac{I}{2}log\,\alpha + \frac{1}{2\sigma}\sum_{j=1}^{J}\sum_{k\in\Omega_j}\omega_{jk}\left(\mu_j-\mu_k\right)^2 + \frac{J}{2}\,log\,\sigma + C(.) \quad (6)$$

or in an alternative form under the independent condition:

$$\Psi(\boldsymbol{\mu},\alpha,\sigma) = \frac{1}{2\alpha}\sum_{i=1}^{I}\varrho_i\left(y_i-\sum_{j=1}^{J}A_{ij}\mu_j\right)^2 + \frac{I}{2}log\,\alpha + \frac{1}{2\sigma}\sum_{j=1}^{J}\sum_{k\in\Omega_j}\omega_{jk}\left(\mu_j-\mu_k\right)^2 + \frac{J}{2}\,log\,\sigma + C(.) \quad (7)$$

where $C(.)$ includes terms, which are not dependent on$\{\boldsymbol{\mu},\alpha,\sigma\}$.

According the LU strategy [5] or the EM terminology [3], when the *j*-th pixel is under consideration for update, there are j-1 pixels having been updated to the (*n*+1)th iteration cycle and J-j+1 pixels remaining at the *n*-th iterated cycle. Given these J pixels of $\boldsymbol{\mu}^{(n*)}$ at the *n*-th and (*n*+1)th stages, the conditional expectation of the two parameters $\{\boldsymbol{\alpha},\boldsymbol{\sigma}\}$ by Eq.(7) would be:

$$\Psi\left(\alpha,\sigma|\boldsymbol{\mu}^{(n*)}\right) = \frac{1}{2\alpha}\sum_{i=1}^{I}\varrho_i\left(y_i-\sum_{j=1}^{J}A_{ij}\mu_j^{(n*)}\right)^2 + \frac{I}{2}log\,\alpha + \frac{1}{2\sigma}\sum_{j=1}^{J}\sum_{k\in\Omega_j}\omega_{jk}\left(\mu_j^{(n*)}-\mu_k^{(n*)}\right)^2 + \frac{J}{2}\,log\,\sigma + C(.) \quad (8)$$

and the maximization of Eq.(8) for the two PDF parameters $\{\boldsymbol{\alpha},\boldsymbol{\sigma}\}$ at the *j*-th pixel is respectively given by:

$$\frac{\partial\Psi(\alpha,\sigma|\boldsymbol{\mu}^{(n*)})}{\partial\alpha} = \partial\left\{\frac{1}{2\alpha}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\mu}^{(n*)}\|^2_{\Sigma_{\boldsymbol{y}}^{-1}} + \frac{I}{2}\,log\,\alpha\right\}\partial\alpha|_{\alpha=\alpha_j^{(n*)}} = 0 \quad (9)$$

$$\frac{\partial\Psi(\alpha,\sigma|\boldsymbol{\mu}^{(n*)})}{\partial\sigma} = \partial\left\{\begin{array}{c}\frac{1}{2\sigma}\sum_{j=1}^{J}\sum_{k\in\Omega_j}\omega_{jk}\left(\mu_j^{(n*)}-\mu_k^{(n*)}\right)^2 \\ +\frac{J}{2}\,log\,\sigma\end{array}\right\}\partial\sigma|_{\sigma=\alpha_j^{(n*)}} = 0 \quad (10)$$

The update for the *j*-th pixel under consideration is given by:

$$\frac{\partial\Psi\left(\boldsymbol{\mu}|\alpha_j^{(n*)},\sigma_j^{(n*)},\boldsymbol{\mu}^{(n*)}\right)}{\partial\mu_j} = \partial\left\{\frac{1}{2\alpha_j^{(n*)}}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\mu}\|^2_{\Sigma_{\boldsymbol{y}}^{-1}} + \frac{1}{2\sigma_j^{(n*)}}\sum_{j=1}^{J}\sum_{k\in\Omega_j}\omega_{jk}\left(\mu_j-\mu_k\right)^2\right\}\partial\mu_j|_{\mu=\mu_j^{(n+1)}} = 0 \quad (11)$$

The solutions of Eqs.(9) and (10) are respectively:

$$\alpha_j^{(n*)} = \frac{1}{I}\sum_{i=1}^{I}\varrho_i\left(y_i-\sum_{j=1}^{J}A_{ij}\mu_j^{(n*)}\right)^2 \quad (12)$$

$$\sigma_j^{(n*)} = \frac{1}{J}\sum_{j=1}^{J}\sum_{k\in\Omega_j}\omega_{jk}\left(\mu_j^{(n*)}-\mu_k^{(n*)}\right)^2 \quad (13)$$

The solution of Eq.(11) for updating μ_j from $\mu_j^{(n)}$ to $\mu_j^{(n+1)}$ is given by the following formula. We use the symbol *s* to indicate pixel index for summation to avoid index confusion. Thus, Eq.(11) can be rewritten as:

$$\frac{\partial\Psi\left(\boldsymbol{\mu}|\alpha^{(n*)},\sigma^{(n*)},|\boldsymbol{\mu}^{(n*)}\right)}{\partial\mu_j} = \partial\left\{\frac{1}{2\alpha_j^{(n*)}}\sum_{i=1}^{I}\varrho_i\left(y_i - \sum_{s=1,s\neq j}^{J}A_{is}\mu_s^{(n*)} - A_{ij}\mu_j\right)^2 + \frac{1}{2\sigma_j^{(n*)}}\left(\sum_{s=1,s\neq j}^{J}\sum_{k\in\Omega_s}\omega_{sk}\left(\mu_s^{(n*)}-\mu_k^{(n*)}\right)^2 + \sum_{k\in\Omega_j}\omega_{jk}\left(\mu_j-\mu_k^{(n*)}\right)^2\right)\right\}\partial\mu_j|_{\mu=\mu^{(n+1)}}$$

$$= \left(\frac{1}{\alpha_j^{(n*)}}\sum_{i=1}^{I}\varrho_i\left(A_{ij}\right)^2 + \frac{1}{\sigma_j^{(n*)}}\right)\mu_j + \frac{1}{\alpha_j^{(n*)}}\sum_{i=1}^{I}\varrho_i\left(-A_{ij}\right)\left(y_i - \sum_{s=1,s\neq j}^{J}A_{is}\mu_s^{(n)}\right) - \frac{1}{\sigma_j^{(n*)}}\sum_{k\in\Omega_j}\omega_{jk}\mu_k^{(n)} = 0 \quad (14)$$

Then, we can obtain:

$$\mu_j{}^{(n+1)} = \frac{\sum_{i=1}^{I}\varrho_i(A_{ij})\left(y_i-\sum_{j=1}^{J}A_{ij}\mu_j^{(n*)}\right)+\sum_{i=1}^{I}\varrho_i(A_{ij})^2\mu_j^{(n*)}+\frac{\alpha_j^{(n*)}}{\sigma_j^{(n*)}}\sum_{k\in\Omega_j}\omega_{jk}\mu_k^{(n*)}}{\sum_{i=1}^{I}\varrho_i(A_{ij})^2+\frac{\alpha_j^{(n*)}}{\sigma_j^{(n*)}}} \quad (15)$$

After Eqs.(12), (13) & (15), i.e. the MAP-LU method, have been applied to all the *J* pixels from *j=1* to *J*, one iteration of

image reconstruction is completed. According to Eq. (12) and Eq. (13), the parameters for data fidelity and prior terms are of the symmetry form and reflect the normalized strength of the data fidelity and the prior.

III. Experiments and Results

Numerical studies were carried out to evaluate the proposed strategy for free parameter estimation. With fully controlled simulation, we can explore the effectiveness and stability of this MAP-LU method.

A. Experiments Setup

The Shepp-Logan phantom of size 512×512 was used to generate the line-integral of size $672(bins) \times 1160(views)$ by forward projections. Then the Poisson noise were added directly in the line-integral domain. By simulating so, the variance $\Sigma_{\boldsymbol{y}}$ or ${\varrho_i}^{-1}$ will be equal to the mean value of line-integral without any approximation. The ω_{jk} used is the spatial distance dependent Huber MRF weights with widow size 3×3.

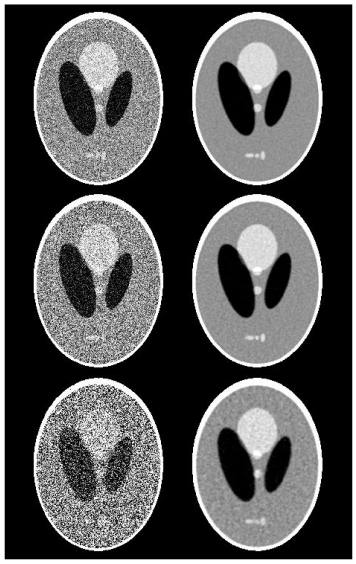


Fig. 1. Performance of the MAP-LU across three noise levels (from top to bottom). For each row, left image is reconstructed by FBP without filter and right one is reconstructed by the proposed EM Bayesian method.

We scaled the numerical phantom by 1, 5 and 10 to mimic different noise levels. The reconstructed results of this EM Bayesian approach were shown in Fig. 1 with comparison of FBP. The reconstruction performance is robust across the three noise levels with good eye appealing images. Fig. 2 shows the estimated free parameters $\alpha^{(n)}$, $\sigma^{(n)}$ and $\alpha^{(n)}/\sigma^{(n)}$ along with iterations. Although the iteration times are different for different noise levels to be converged, consistent transforming trends are observed in Fig. 2.

As we discussed in the method section, the parameter will be updated by the LU strategy. Fig. 3 is an example of our updated parameters along with the pixel ID and global iteration times n. According to the estimation, scale factors for the first and last pixels are observably different initially. With the iteration increases, the difference becomes smaller.

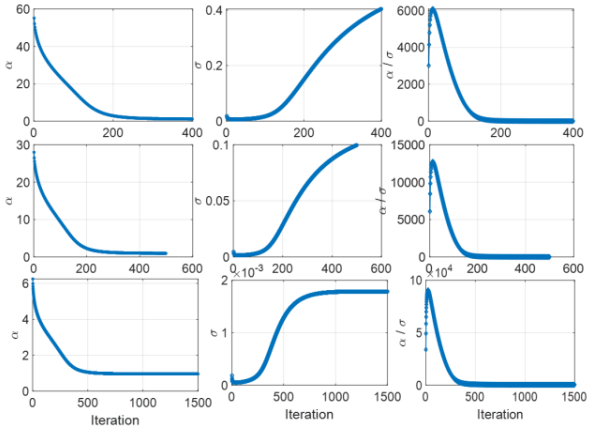


Fig. 2. Parameter estimation by the MAP-LU along with iterations across three noise levels (from top to bottom).

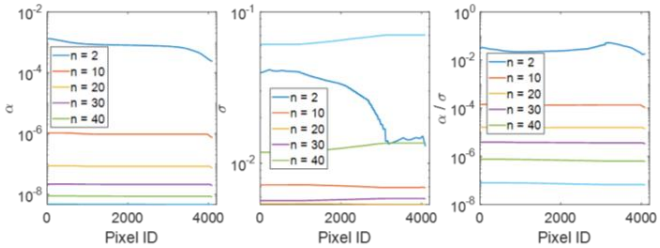


Fig. 3. Estimated free parameters by the MAP-LU along with pixel ID and global iteration times n.

B. Effects of Initialization

We further compared the effect initialization on the proposed method. Three initials are tested: uniform initial by averaging the FBP results, FBP without filter and FBP with hann50 filter. The reconstructed images and their initials are presented in Fig. 4. The estimated parameters along iterations are presented in Fig. 5. According to Fig. 5, the pattern of $\alpha^{(n)}$ does not change too much for different initials. However, the pattern of $\sigma^{(n)}$ varies. Their ratio $\alpha^{(n)}/\sigma^{(n)}$ kind of follows similar pattern with different converges rate. According to Fig. 4, the reconstructed results are stable. It implies that the MAP-LU can effectively estimate the parameter values starting with different initials and provide stable solutions.

C. Effect of Variance

Effect of variance on MAP-LU was also explored. Fig. 6 below presents the reconstructed images by using $\Sigma_{\boldsymbol{y}}$ given by $\boldsymbol{y}$ (left) and $\boldsymbol{A\mu}^{(n)}$(right). Comparable reconstructed images are obtained. The estimated ratio of $\alpha^{(n)}/\sigma^{(n)}$ starts from different values but finally converged to very similar values.

Figure 7 presents the reconstructed images for $\Sigma_{\boldsymbol{y}}$ without (upper) and with normalization (lower) by $sum(\Sigma_{\boldsymbol{y}}^{-1}) = 1$. Comparable reconstructed images are obtained. The estimated parameters follow the same pattern with linear scaling strength. This observation agrees with our expectation that even though the magnitude of $\alpha^{(n)}/\sigma^{(n)}$ is scaled due to normalization, the fidelity strength of $(\boldsymbol{y} - \mathbf{A}\boldsymbol{\mu})\Sigma_{\boldsymbol{y}}^{-1}(\boldsymbol{y} - \mathbf{A}\boldsymbol{\mu})^T$ will also be scaled by the same degree. Therefore, the results would not be affected.

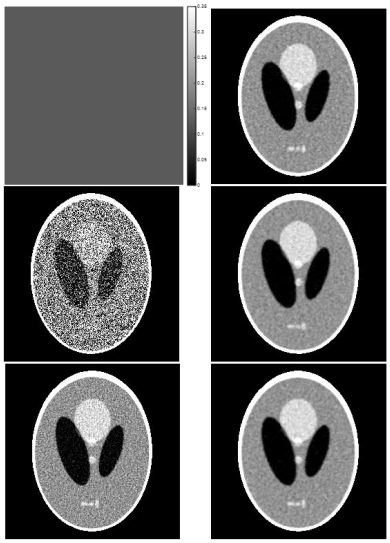


Fig. 4. Performance of MAP-LU with different initials. From top to bottom, the initials (left) are uniform averaged from FBP, FBP without filter and FBP with hann50 filter.

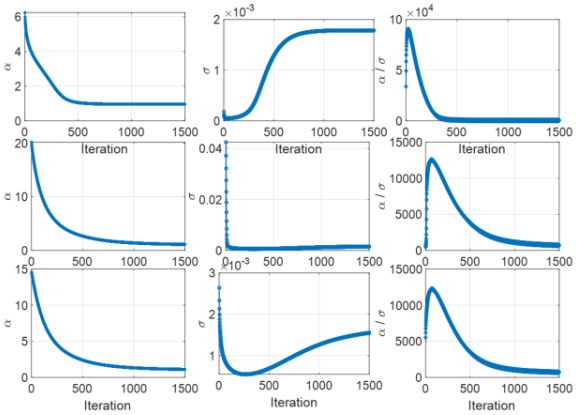


Fig. 5. Estimated free parameters by MAP-LU along with iterations with three different initials in accordance with Fig. 3.

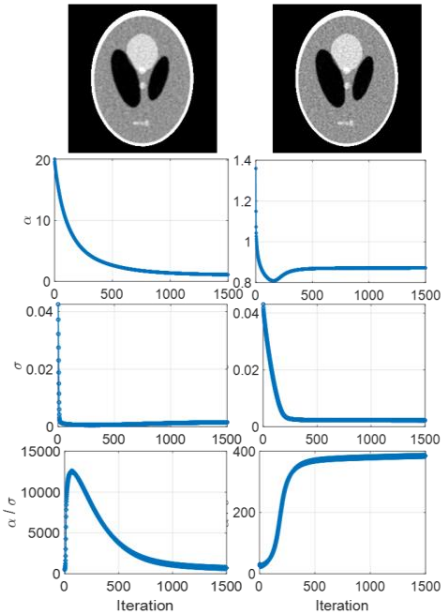


Fig. 6. Performance of the MAP-LU and its corresponding parameter transforming by using $\Sigma_{\boldsymbol{y}}$ given by $\boldsymbol{y}$ (left) and $\boldsymbol{A\mu}^{(n)}$(right).

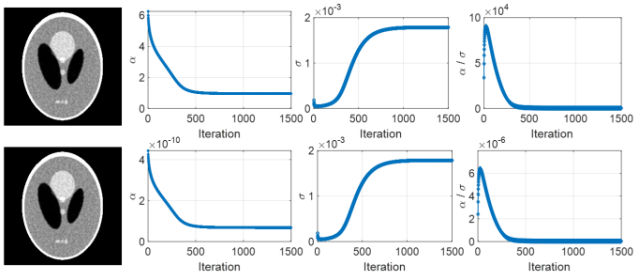


Fig. 7. Performance of MAP-LU images for $\Sigma_{\boldsymbol{y}}$ without (upper) and with normalization (lower) by $sum(\Sigma_{\boldsymbol{y}}^{-1}) = 1$.

IV. Conclusion

A theory-based strategy is presented to determine the freely-adjustable parameter for penalized maximum likelihood or penalized weighted least-squares tomographic image reconstruction from projections. The strategy is based on the PDF requirements of the Bayes' Law and LU numerical calculation. The freely-adjustable parameter is related to the ratio of the data fidelity variance (confidence) and the prior variance (confidence). This implies that given the count-limited imaging, such as LdCT, we desire high confidence or low variance on the prior penalty. In other words, we desire the medical truth to derive prior [10] and train machine learning [11].

References

[1] Radon Transform: https://en.wikipedia.org/wiki/Radon_transform .

[2] L.A. Shepp and B.F. Logan, "The Fourier Reconstruction of a Head Section," *IEEE Transactions on Nuclear Science*, vol. 21, pp. 21-43, 1974.

[3] A. Dempster, N. Laird and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B* (Methodological), vol. 39, no. 1, pp.1-38, 1977.

[4] L.A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Transactions on Medical Imaging*, vol. 1, pp.113-122, 1982.

[5] K. Sauer and C. Bouman, "A Local Update Strategy for Iterative Reconstruction from Projections," *IEEE Transactions on Signal Processing*, vol. 41, no.2, pp. 534-548, 1993.

[6] J. Ma, Z. Liang, J. Huang, W. Chen, and H. Lu, "Variance Analysis of X-ray CT Sinograms in the Presence of Electronic Noise Background," *Medical Physics*, vol. 39, no. 7, pp. 4051-4065, 2012.

[7] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no.6, pp. 721-741, 1984.

[8] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized Weighted Least-squares Approach to Sinogram Noise Reduction and Image Reconstruction for Low-dose X-ray CT," *IEEE Transactions on Medical Imaging*, vol. 25, no. 10, pp. 1272-1283, 2006.

[9] G. Zhang, F. Liu, J. Liu, J. Luo, Y. Xie, J. Bai, and L. Xing, "Cone Beam X-ray Luminescence Computed Tomography Based on Bayesian Method," *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp .225-235, 2017.

[10] H. Zhang, H. Han, Y. Liu, W. Moore and Z. Liang, "Deriving adaptive MRF coefficients from previous normal-dose CT scan for low-dose image reconstruction via penalized weighted least-squares minimization", *Medical Physics*, vol. 41, no.4, pp. 041916-1 - 041916-15, 2014.

[11] Jiaxiang Tan, Yongfeng Gao , Zhengrong Liang, Fellow, IEEE , Weiguo Cao , Marc J. Pomeroy, Yumei Huo , Lihong Li, Matthew A. Barish, Almas F. Abbasi, and Perry J. Pickhardt, "3D-GLCM CNN: A 3-dimensional gray-level co-occurrence matrix-based CNN model for polyp classification via CT colonography." *IEEE Transactions on Medical Imaging*, in press, 2019.

Ultra-low Dose CT for PET Attenuation Correction: Preliminary Study

Kai Mei*, Nadav Shapira, Joshua Scheuerman, Janet Reddin, Amy E. Perkins,
Joel S. Karp and Peter B. Noël

***Abstract*—The purpose of this study is to investigate ultra-low dose Computed Tomography (CT) acquisition schemes for Positron Emission Tomography (PET) attenuation correction. This study includes patient data acquired on a research PET/CT scanner. Ultra-low dose scans (3.2, 1.1 and 0.36 mAs) were simulated by combining simulated tube current reductions (39%, 13%, 4% or 1%) and sparse sampling (1/2, 1/6, 1/18 or 1/54). A modal-based statistical iterative algorithm with regularization was used to reconstruct noise- and artifact-suppressed images. The structural similarity index measurement (SSIM) was used as a metric to compare the ultra-low dose CT and the corrected PET images with their corresponding data at clinical low dose. Compared to the current standard (50 mAs), ultra-low dose CT images with 0.36 mAs (0.7%) provide accurately attenuation corrected PET images (SSIM 0.86). In the case that CT data are not utilized for diagnostics, our results illustrate with clinical data that ultra-low dose CT is feasible for attenuation correction.**

***Index Terms*— Computed tomography, positron emission tomography, attenuation correction, sparse sampling.**

I. Introduction

POSITRON emission tomography (PET) and computed tomography (CT) are standard imaging tools for oncology diagnosis and tumor staging. In PET/CT, CT provides anatomical information for localization and necessary attenuation profiles for PET, mitigating the limited resolution of PET and the unavoidable attenuation of the positrons by the patient's body during the scan. Over the last decades, growing community awareness about exposure-associated health risks has initialized significant ongoing efforts to reduce CT radiation exposure. The purpose of this study is to investigate CT dose reduction strategies and the associate level of CT dose required for accurate PET attenuation correction.

In CT-based attenuation correction (CTAC) for PET images, CT images are used to generate transmission maps, by piecewise-linearly re-scaling the attenuation coefficients to the PET energy of 511 keV photons according to tissue type. Unlike diagnostic CT, where the image quality has to be sufficiently high enough for radiologists to distinguish, for example, small soft tissue differences, CT images acquired for the correlation with PET and the transmission maps for attenuation correction have lower requirements, where spatial resolution and soft-tissue contrast are less demanding. Recent research has shown simulated CT scans of a human abdominal phantom [1] at less than 0.1 mAs (per axial slice) using low tube current and pulsing (sparse sampling) techniques [2] may be already sufficient for CTAC PET. This correction with such ultra-low dose CT can provide significant radiation dose savings for patients undergoing frequent PET examinations, as well as for certain research applications.

In this paper, we investigated the possibility of ultra-low dose CT images for PET attenuation correction with clinical PET/CT scans. We combined sparse sampling and tube current reduction simulation to mimic ultra-low dose CT scan at 0.36, 1.1 and 3.2 mAs. The ultra-low dose raw CT data are reconstructed using model-based statistical iterative reconstruction with regularization to suppress noise and artifacts. We compared the resulting CT images and the corresponding corrected PET images with the original images at a relatively low dose (50 mAs). Our result shows that ultra-low dose CT under 1 mAs does not significantly alter the attenuation correction effect on PET images and provides sufficient image quality for correction purposes.

II. Material and Methods

A. Patient and Scanning Protocols

In this preliminary study, we identified a patient undergoing a standard PET/CT examination in the past year. The patient was a 56-year-old female (BMI = 21.6). The patient was scanned with an Ingenuity TF PET/CT scanner (Philips Healthcare, Cleveland, OH, USA). The CT scan was acquired with a clinical low dose (LD) protocol. The scanned length was 95 cm, tube voltage was 120 kVp, tube current was 70 mA and rotation time was 0.5 seconds. The scan was spiral with a 0.704 pitch factor, capturing 2400 projections per rotation. The reconstructed field of view was 700 mm by 700 mm, and each axial slice thickness was 4 mm. The axial slices were not overlapping. For each slice, the average exposure was 50 mAs. For the PET scan, data acquisition was performed using a standard of care (SOC) protocol for 18-Fluorodeoxyglucose (FDG), 90 sec/bed.

B. Ultra-low Dose CT Simulation

Raw projection data were exported from the CT scanner.

K. Mei, N. Shapira, J. Sheuerman, J. Raddin, J. Karp and P. Noël are with the Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA USA.

A. Perkins is with Philips Healthcare, Cleveland, OH USA

*Email: kai.mei@pennmedicine.upenn.edu

TABLE I
LIST OF ULTRA-LOW DOSE PROTOCOLS APPLIED

Exposure	Tube current	Views	Sampling rate	Exposure Time	Percentage	Notation
50 mAs	70 mA	2400	1	714 ms	100%	Low Dose (LD)
3.2 mAs	9 mA	1200	1/2	357 ms	6.4%	1A
3.2 mAs	27 mA	400	1/6	119 ms	6.4%	**1B**
1.1 mAs	3 mA	1200	1/2	357 ms	2.1%	2A
1.1 mAs	9 mA	400	1/6	119 ms	2.1%	2B
1.1 mAs	27 mA	133	1/18	40 ms	2.1%	**2C**
0.36 mAs	1 mA	1200	1/2	357 ms	0.7%	3A
0.36 mAs	3 mA	400	1/6	119 ms	0.7%	3B
0.36 mAs	9 mA	133	1/18	40 ms	0.7%	**3C**
0.36 mAs	27 mA	44	1/54	13 ms	0.7%	3D

Tube Current is constant and there is no modulation. **Views** stands for number of views per gantry rotation. **Sampling rate** is the ratio of processed views per rotation. **Exposure time** is exposure time per axial slice, which is rotation time divided by spiral pitch factor. **Exposure** is the average radiation dose per axial slice, considering the spiral pitch factor.

Ultra-low dose CT scans were generated at three levels: 0.36 mAs, 1.1 mAs and 3.2 mAs by combining tube current reduction and sparse sampling. Tube current reduction was achieved using a simulation algorithm [3][4] that adds scaled electronic noise obtained from blank scans (scans without any object in the field of view). Original projections at low dose with 70 mA x-ray intensity were used to generate projection data simulating x-ray intensities of 1 mA, 3 mA, 9 mA or 27 mA. Sparse sampling was achieved by disregarding projections evenly in each gantry rotation. Hence, 1/2, 1/6, 1/18 or 1/54 sampling of 2400 projections per rotation, resulting in average 1200, 400, 133 or 44 projections per rotation. Mixing both methods, ultra-low dose simulations were achieved at 6.4%, 2.1% and 0.7% of the original clinical low dose level, Table I.

Reconstruction was performed off-line with a model-based statistical iterative algorithm [5] with a momentum approach [6] on a GPU server (Nvidia Geforce TITAN X). The iterative reconstruction utilized ordered subsets, where each subset consists of 1/24 of the undersamped projections randomly selected but evenly distributed across different gantry rotations. In a 1/2 sampling case, iteration was manually stopped after each projection was processed 30 times, i.e., after 30 iterations. The reconstruction grid was 1 mm voxels with isotropic voxel size. The images were resampled to match the resolution of the original reconstruction mentioned in Section 2.1 and the attenuation coefficients were converted to Hounsfield Unit using the water attenuation at room temperature (0.0192 cm^{-1}).

C. PET Attenuation Correction

Reconstructed CT images were converted into DICOM format and imported to the PET/CT scanner as input for the attenuation correction of the original PET scan. New PET image series were generated, simulating ultra-low dose CT attenuation corrections at 3.2, 1.1 and 0.36 mAs.

D. Image Analysis

Images were analyzed using structural similarity index measurement (SSIM) [7]. SSIM assesses the similarities between two images considering luminance, contrast and structure. Values approaching 1 signifies that two images are identical while a value approaching 0 represents that they are completely different.

III. RESULT

We investigated the three ultra-low dose levels CT images at 0.36 mAs, 1.1 mAs and 3.2 mAs (per slice), with comparison to the 50 mAs original clinical LD scan (see Table I). For each level of simulated radiation dose, we studied two to four combinations of tube current reductions and sparse

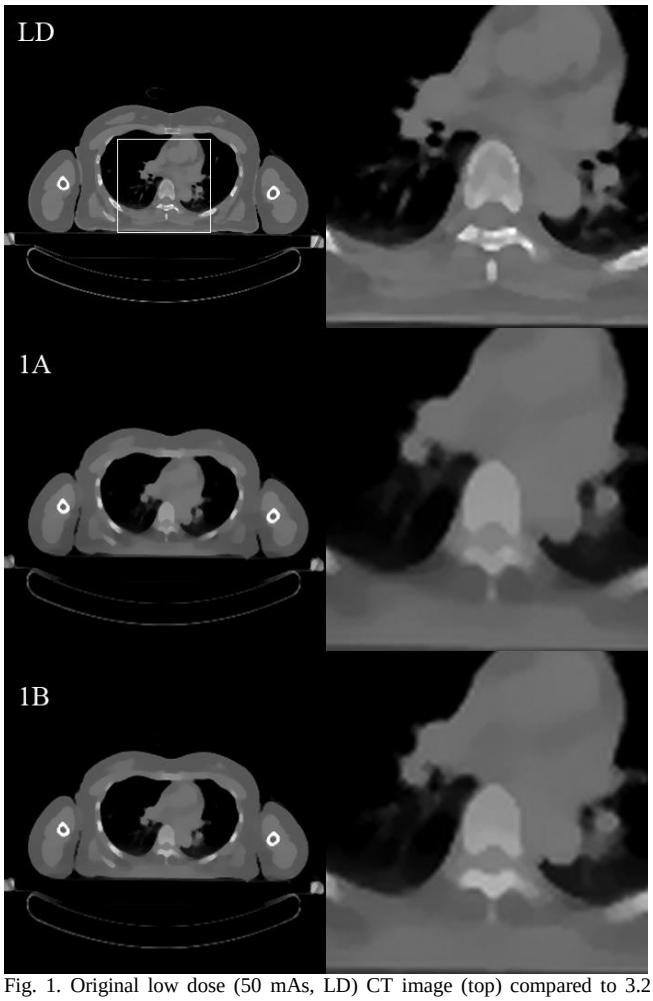


Fig. 1. Original low dose (50 mAs, LD) CT image (top) compared to 3.2 mAs (6.4%) CT images (middle and bottom). Window W/L: 1400/0 HU.

TABLE II
SSIM OF ULTRA-LOW DOSE CT IMAGES

		A 1/2	B 1/6	C 1/18	D 1/54
	Heart				
1	3.2 mAs	0.45	**0.48**		
2	1.1 mAs	0.40	0.43	**0.44**	
3	0.36 mAs	0.36	0.39	**0.41**	0.39
	Kidney				
1	3.2 mAs	0.44	**0.49**		
2	1.1 mAs	0.38	0.41	**0.45**	
3	0.36 mAs	0.35	0.38	**0.40**	0.39
	Liver				
1	3.2 mAs	0.41	**0.44**		
2	1.1 mAs	0.37	0.39	**0.41**	
3	0.36 mAs	0.35	0.37	**0.38**	0.37
	Average				
1	3.2 mAs	0.43	**0.47**		
2	1.1 mAs	0.38	0.40	**0.43**	
3	0.36 mAs	0.35	0.38	**0.40**	0.39

sampling reconstructions, depending on the level. We found that a moderate level of sparse sampling (133 to 400 views per gantry rotation) performed well with a moderate level of tube current reduction (9 to 27 mA), Fig. 1, 2 and 4. However, either highly-sparsely sampled data or over-reduced tube current resulted in reduced image quality.

Soft tissue boundaries between muscle and fat, for example, can be reconstructed and distinguished at dose level 3.2 mAs in both combinations. There are only slight visual differences for images with lower tube current (1A) or the image with a sparser sampling rate (1B), Fig. 1.

The separation between soft tissues starts to blur out at dose level 1.1 mAs. Areas appear blurry behind the patient's spine for the lowest tube current image (2A), probably due to the cost of heavy regularization mitigating the effect of photon starvation and noise, Fig. 2.

At 0.36 mAs, artefacts become prominent (3A). The edges between muscle and fat tissue started to deteriorate in the lowest sampling rate reconstruction (3D). At this simulated dose level, the image with the best visual quality is 133 projections (3C, 1/18 sparse sampling), Fig. 3.

SSIM showed similar values at different axial locations: heart (lung), liver and kidney, Table II. The sparse sampling and tube current combinations with the highest SSIM values are high-lighted with bold numbers, which are approximately at a sampling rate of 1/18 of the full sampling (133 projections per gantry rotation).

We selected the best images in each ultra-low dose level (1B, 2C and 3C) to generate new CTAC PET images. The resulting images showed no visual difference compared to the original CTAC PET image, Fig. 4. The SSIM is 0.86. Quantitative measurement showed no difference among images at ultra-low dose levels 3.2, 1.1 and 0.36 mAs (SSIM = 1.0).

IV. DISCUSSION

This study investigated CT images with simulated ultra-low exposure levels and the use of the resulting images as inputs to compute CTAC PET. We have found that for an optimal combination of tube current and sampling rate, the produced ultra-low dose CT image with 0.36 mAs resulted in accurate CTAC PET images, as compared to the original clinical low dose CTAC PET images (50 mAs).

This preliminary study has limitations. We first acknowledge that it includes only one patient so far. We are in the process of collecting more cases and are expecting more data for an extensive follow-up study. In addition, while our processing relies only on a high regularization term in the reconstruction aiming to mitigate the quantum and electronic noise and the sparse sampling rate, other filters, such as denoising techniques or learning-based algorithms may perform better and benefit the final result.

V. CONCLUSION

We studied the possibility of performing attenuation correction for PET with ultra-low dose CT images by utilizing clinical patient PET/CT scan data. Our result showed that ultra-low dose CT under 1 mAs could provide sufficiently accurate information for PET attenuation correction purposes.

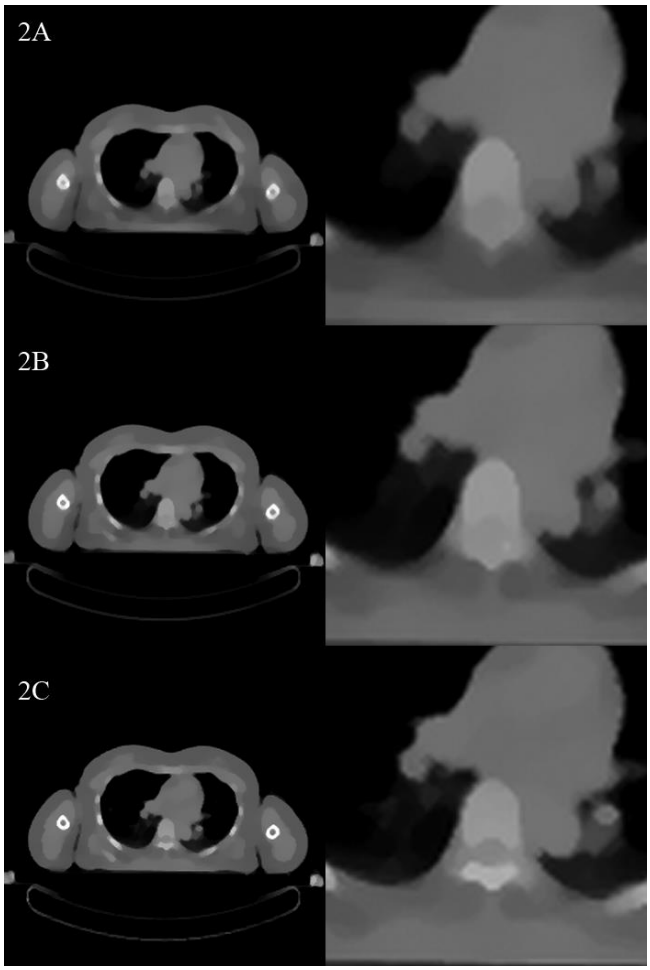


Fig. 2. 1.1 mAs (2.1%) CT images with mixed combinations of sparse sampling and reduced tube current. Window W/L: 1400/0 HU.

DISCLOSURES

Institutional Review Board approval was obtained. Patient provided informed consent for the use of his/her data for research purposes.

ACKNOWLEDGMENT

We acknowledge support through grant ROI-CA113941 and the University of Pennsylvania Research Foundation (URF).

REFERENCES

[1] X. Rui, L. Cheng, Y. Long, L. Fu, A. M. Alessio, E. Asma, P. E. Kinahan, and B. De Man, "Ultra-low dose CTP attenuation correction for PET/CT: analysis of sparse view data acquisition and reconstruction algorithms," *Physics in Medicine & Biology*, vol. 60, no. 19, p. 7437, 2015.

[2] U. Wiedmann, V. B. Neculaes, D. Harrison, E. Asma, P. E. Kinahan, and B. De Man, "X-ray pulsing methods for reduced-dose computed tomography in PET/CT attenuation correction," in *Medical Imaging 2014: Physics of Medical Imaging*, vol. 9033. International Society for Optics and Photonics, 2014, p. 90332Z.

[3] S. Zabic, Q. Wang, T. Morton, and K. M. Brown, "A low dose simulation tool for CT systems with energy integrating detectors," *Medical physics*, vol. 40, no. 3, p. 031102, 2013.

[4] D. Muenzel, T. Koehler, K. Brown, S. Zabic, A. A. Fingerle, S. Waldt, E. Bendik, T. Zahel, A. Schneider, M. Dobritz et al., "Validation of a low dose simulation technique for computed tomography images," *PloS one*, vol. 9, no. 9, p. e107843, 2014.

[5] J. A. Fessler, M. Sonka, and J. M. Fitzpatrick, "Statistical image reconstruction methods for transmission tomography," *Handbook of medical imaging*, vol. 2, pp. 1–70, 2000.

[6] D. Kim, S. Ramani, and J. A. Fessler, "Combining ordered subsets and momentum for accelerated x-ray CT image reconstruction," *IEEE transactions on medical imaging*, vol. 34, no. 1, pp. 167–178, 2014.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli et al., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

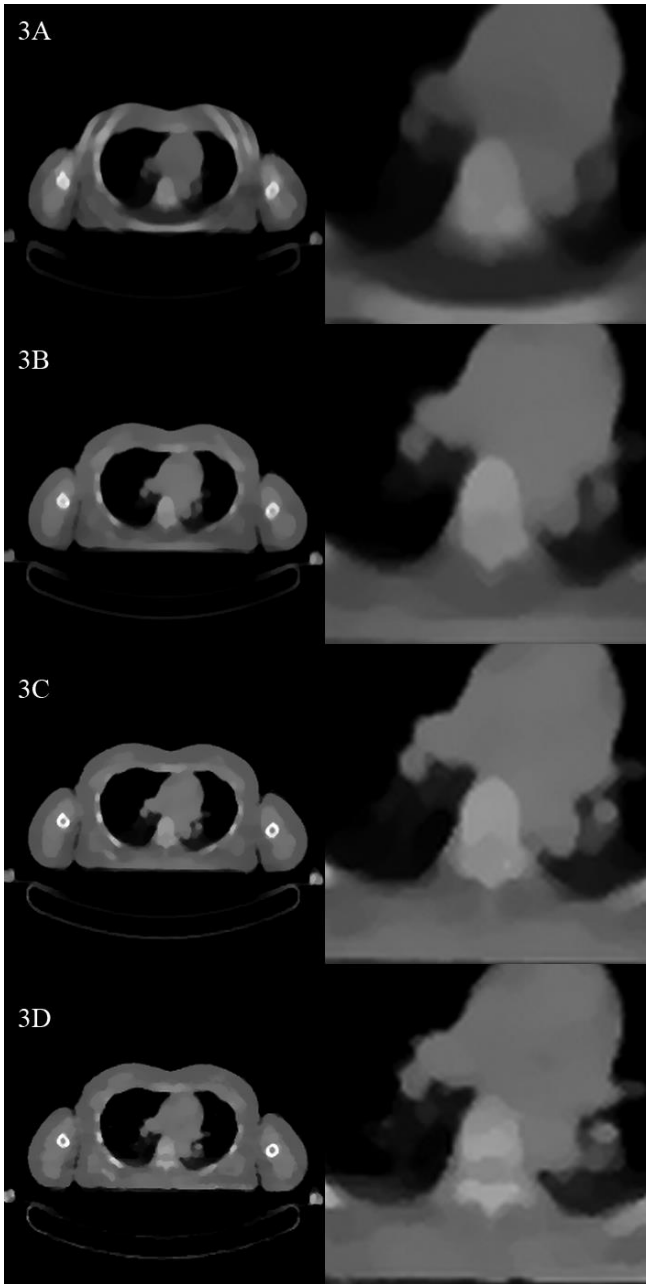


Fig. 3. 0.36 mAs (0.7%) CT images with mixed combinations of sparse

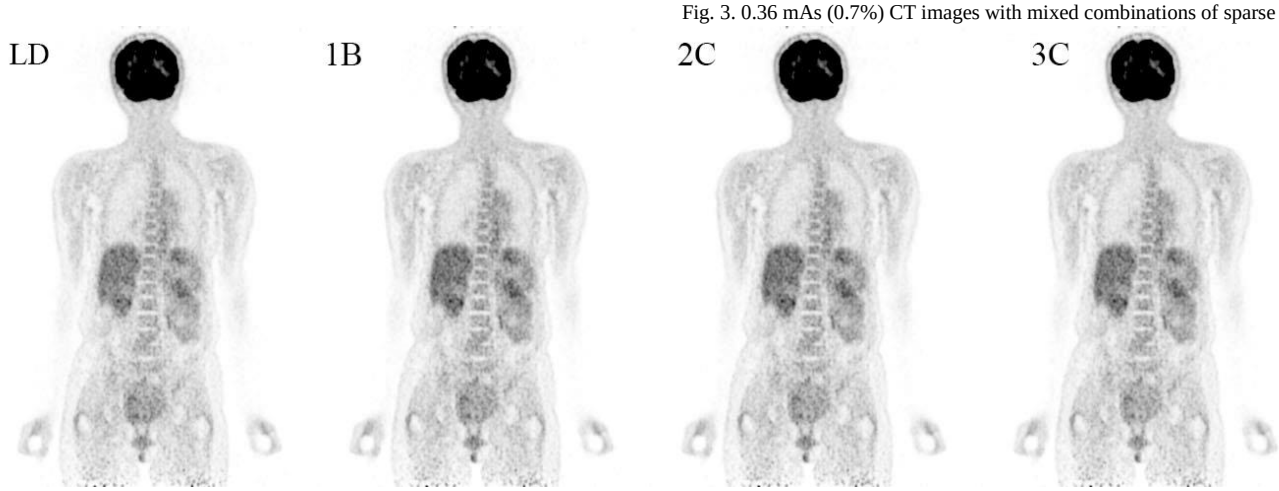


Fig. 4 Original low dose (50 mAs) CTAC PET image (leftmost, LD) and ultra-low dose (3.2, 1.1 and 0.36 mAs) CTAC PET images (1B, 2C, 3C).

Method of deep distribution learning from noisy samples and an application for low dose CT

Kaichao Liang, Li Zhang, Le Shen, Zhiqiang Chen, Yuxiang Xing

***Abstract*—Low-dose CT (LDCT) gained much attention for clinical CT in these decades. With the power of deep learning, the LDCT data processing is significantly improved in recent years. However, the data-driven deep learning depends highly on large normal-dose CT (NDCT) dataset to guide the training process. In most cases, paired LDCT and NDCT images are required thus supervised training can be developed to match LDCT images and NDCT images. While manifold learning searches for NDCT manifold and relaxes the requirement for paired data, it learns the manifold of NDCT from additional NDCT dataset and adopt the manifold to enhance LDCT images. However, the NDCT image dataset are often expensive to acquire or sometimes cannot be reached, which limits the applications of deep learning. In this work, to relax the requirement of NDCT dataset, we proposed an unsupervised deep learning denoising method by distribution learning. We proved that the clean signal distribution can be learned from noisy samples in case of known noise model. The method is referred as deep distribution learning from noisy samples (DDLN). We applied DDLN to unsupervised LDCT projection denoising combined with data fidelity loss. The network was in the form of a generative adversarial network but trained with LDCT dataset only. In our simulation study, the proposed method outperforms Hanning-window filtering method and achieved results comparable with supervised learning after reconstruction and even better in terms of small structures. The results verified the success of the proposed DDLN.**

***Index Terms*— Low-dose CT, distribution learning, generative adversarial network**

I. Introduction

REDUCING X-ray CT scanning dose to the lowest level on the condition of acceptable image quality is of great significance protecting patients from radiation over exposure. Low-dose CT (LDCT) problem has been a major challenge in X-ray CT field during the decades and steady progresses have been made. Generally, the LDCT processing methods can be distinguished into three categories based on process sequence. The first category is LDCT data pre-processing that reduces LDCT noises in the original raw data or projection data before CT image reconstruction. Various types of fixed or structure adaptive filters can suppress noise effectively as the projection domain is the source domain of noise [1-3]. In contrast with pre-processing methods, post-processing methods restore structures and suspense artifacts in the image domain after common CT reconstructions [4-7]. Adaptive noise filters remove striking artifacts and restore clinical structures. Block-matching method is one of the widely used post-processing methods which is effective and efficient [5]. The third category is LDCT reconstruction algorithms which enhance image quality in the reconstruction process. Generally, model-based iterative reconstructions (MBIR) are to acquire high-quality images by maximum-a-posterior (MAP) [7-9]. MBIR combines both data-fidelity and image prior constraints. With noise-model and various manually designed image priors, MBIR achieves great LDCT reconstruction results. However, the manually designed smoothing priors may not be accurate for NDCT image distribution and introduce additional errors. Dictionary learning prior extracts NDCT image manifold based on NDCT dataset which is more flexible but the dictionary transformation is limited to be linear [10].

In recent years, deep learning has shown its power for LDCT image enhancement. The most common deep learning method is to train a post-processing network to suspense noise in image domain [11-14]. While deep learning can also be integrated into iterative reconstruction process and improve the MBIR reconstruction [15-16]. However, these methods generally need paired LDCT and NDCT images and the networks are trained supervised. With generative adversarial learning [17] or manifold learning [18], the distribution or manifold of NDCT images can be learned unsupervised, but these methods are still highly depended on NDCT images. We developed an unsupervised LDCT reconstruction network based on MAP loss [19], but the results were limited by manually designed prior.

In this work, we proposed a new method of deep distribution learning from noisy samples (DDLN) thus the NDCT projection distribution can be inferred from LDCT projection dataset with known noise model. Further, we developed an unsupervised network for LDCT denoising with only LDCT projection data. To illustrate the effectiveness of the theory and validate the method, a denoising network was implemented in projection domain for dental LDCT denoising.

II. Methods

A. LDCT noise normalization

In an X-ray CT scan, there are mainly two causes of noises in acquired data: the statistical fluctuations of photons and the electric noise. Gaussian-Poisson mixed model is generally adopted to approximate the noise distribution for low-dose mono-energy CT [20]. The noisy raw data I_N can be formulated

Kaichao Liang, Li Zhang and Yuxiang Xing are members of Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education. Yuxiang Xing is the corresponding author (E-mail: xingyx@mail.tsinghua.edu.cn).

This work is supported by National Natural Science Foundation of China (No. 61771279 and No. 61527807) and National Science Fundation for Distinguished Young Scholars of China (No. 11525521).

as:

$$I_{\mathrm{N}} = \gamma \mathrm{Poisson}(C) + \mathrm{Gaussian}\left(0, \sigma_e^{\,2}\right) \tag{1}$$

with C denoting the mean photon number to a detector bin, γ the mean contribution of each photon towards electric signal, $\sigma_e^{\,2}$ the variance of electric noise. Here, $\mathrm{Poisson}(\bullet)$ means the Poisson distribution, and $\mathrm{Gaussian}(\bullet)$ the Gaussian distribution. According to [21, 22], I_{N} can be transformed to a new random variable with additive standard Gaussian noise independent of signal by generalized Anscombe transformation (GAT):

$$a_{\mathrm{N}} \equiv \mathcal{A}\left(I_{\mathrm{N}}\right) = \frac{2}{\gamma} \times \sqrt{\gamma I_{\mathrm{N}} + 0.375\gamma^2 + \sigma_e^{\,2}} \tag{2}$$

Here, $\mathcal{A}(\bullet)$ denotes GAT. After GAT, a_{N} is of standard Gaussian distribution:

$$a_{\mathrm{N}} \cong \bar{a}_{\mathrm{N}} + n, \quad n \sim \mathrm{Gaussian}(0,1) \tag{3}$$

The GAT normalizes the noise, thus it eases the denoising problem to a large extent.

B. Learning NDCT distribution from LDCT dataset

Generally, raw data are transformed into projection domain with negative logarithmic operation to obtain projection data for image reconstruction.

$$p_{\mathrm{N}} = -\ln\left(I_{\mathrm{N}} / I_0\right) \tag{4}$$

Here, the I_0 denotes none-attenuated signal, p_{N} the noisy projection. Projection data denoising is of common interest in CT imaging. In this work, we adopt a maximum a posteriori (MAP) objective function with a prior constraint after GAT as for low dose CT data denoising:

$$\begin{aligned} &\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmin}}\, \ell_{\mathrm{f}}\left(\mathbf{p}_{\mathrm{N}}; \mathbf{p}\right) + \lambda \ell_{\mathrm{P}}\left(\mathbf{a}\right) \\ &\text{s.t.} \quad \mathbf{a} = \mathcal{A}\left(I_0 \exp(-\mathbf{p})\right) \end{aligned} \tag{5}$$

Here, $\mathbf{p}_{\mathrm{N}}$ is a vector of acquired noisy projections along all X-ray paths of a scan, $\hat{\mathbf{p}}$ the projection estimation, $\mathbf{a}$ the GAT of $\mathbf{p}$, ℓ_{f} the data fidelity loss between estimated projection and LDCT projection. Here, $\ell_{\mathrm{P}}\left(\mathbf{a}\right) = -\ln\left(f_{\mathbf{A}}(\mathbf{a})\right)$ is a prior constraint on $\mathbf{a}$. We assume $\mathbf{a}$ is a sample value of a random vector $\mathbf{A}$ and $f_{\mathbf{A}}(\mathbf{a})$ denotes the probability density function (pdf) of $\mathbf{A}$. When a large number of samples are estimated, the problem can be transformed to a new whole distribution constraint as Eq. 7.

$$\begin{aligned} &\hat{\mathbf{p}}_{1,\cdots,\mathrm{M}} = \underset{\mathbf{p}_{1,\cdots,M}}{\operatorname{argmin}} \sum_m \ell_{\mathrm{f}}\left(\mathbf{p}_{\mathrm{N},m}; \mathbf{p}_m\right) / \mathrm{M} + \lambda \ell_{\mathrm{D}}\left(f_{\tilde{\mathbf{A}}}, f_{\mathbf{A}}\right) \\ &\text{s.t.}\ \tilde{\mathbf{A}} \equiv \mathcal{X}\{\mathbf{a}_m, m = 1, \cdots, \mathrm{M}\} \\ &\&\mathbf{a}_m = \mathcal{A}\left(I_0 \exp(-\mathbf{p}_m)\right), m = 1, \cdots, \mathrm{M} \end{aligned} \tag{6}$$

Here, M denotes the total number of projection data samples, m the m^{th} sample, $\tilde{\mathbf{A}} \equiv \mathcal{X}\{\mathbf{a}_m, m = 1:\mathrm{M}\}$ denotes a random vector defined by the set of $\mathbf{a}_m$, $\ell_{\mathrm{D}}\left(f_{\tilde{\mathbf{A}}}, f_{\mathbf{A}}\right)$ is a measurement of the distribution distance in terms of pdfs.

However, when only the LDCT dataset is available, the distribution $f_{\mathbf{A}}$ is unknown. Hence, we consider an alternative optimization problem:

$$\begin{aligned} &\hat{\mathbf{p}}_{1,\cdots,M} = \underset{\mathbf{p}_{1,\cdots,M}}{\operatorname{argmin}} \sum_m \ell_{\mathrm{f}}\left(\mathbf{p}_{\mathrm{N},m}; \mathbf{p}_m\right) / \mathrm{M} + \lambda \ell_{\mathrm{D}}\left(f_{\tilde{\mathbf{A}}+\mathbf{N}}, f_{\mathbf{A}_{\mathrm{N}}}\right) \\ &\text{s.t.}\ \tilde{\mathbf{A}} \equiv \mathcal{X}\{\mathbf{a}_m, m = 1, \cdots, \mathrm{M}\} \\ &\&\mathbf{a}_m = \mathcal{A}\left(I_0 \exp(-\mathbf{p}_m)\right), m = 1, \cdots, \mathrm{M} \end{aligned} \tag{7}$$

Since $f_{\mathbf{A}_{\mathrm{N}}}$ can be accessed. Here $\mathbf{A}_{\mathrm{N}} = \mathbf{A} + \mathbf{N}$ denotes the random vector of noisy projection after GAT, $\mathbf{N}$ the random vector of noise. If the noise in $\mathbf{A}_{\mathrm{N}}$ is of non-zero characteristic function, Eq. (7) asymptotic approximates to Eq. (6) because:

$$f_{\tilde{\mathbf{A}}} = f_{\mathbf{A}} \quad \text{iff} \quad f_{\tilde{\mathbf{A}}+\mathbf{N}} = f_{\mathbf{A}_{\mathrm{N}}} \tag{8}$$

Proof: Considering a general denoising problem for independent additive noise: $\mathbf{Z} = \mathbf{S} + \mathbf{N}$ with $\mathbf{Z}$ denoting the noisy data, $\mathbf{S}$ the clean signal and $\mathbf{N}$the noise which are all random vectors. If $\mathbf{S}$ and $\mathbf{N}$ are independent in a common measurable space, the characteristic function of $\mathbf{Z}$ can be obtained by the multiplication of the characteristic functions of $\mathbf{S}$ and $\mathbf{N}$:

$$\mathcal{F}_{\mathbf{Z}}(\boldsymbol{\mu}) = \mathcal{F}_{\mathbf{S}}(\boldsymbol{\mu}) \mathcal{F}_{\mathbf{N}}(\boldsymbol{\mu})$$

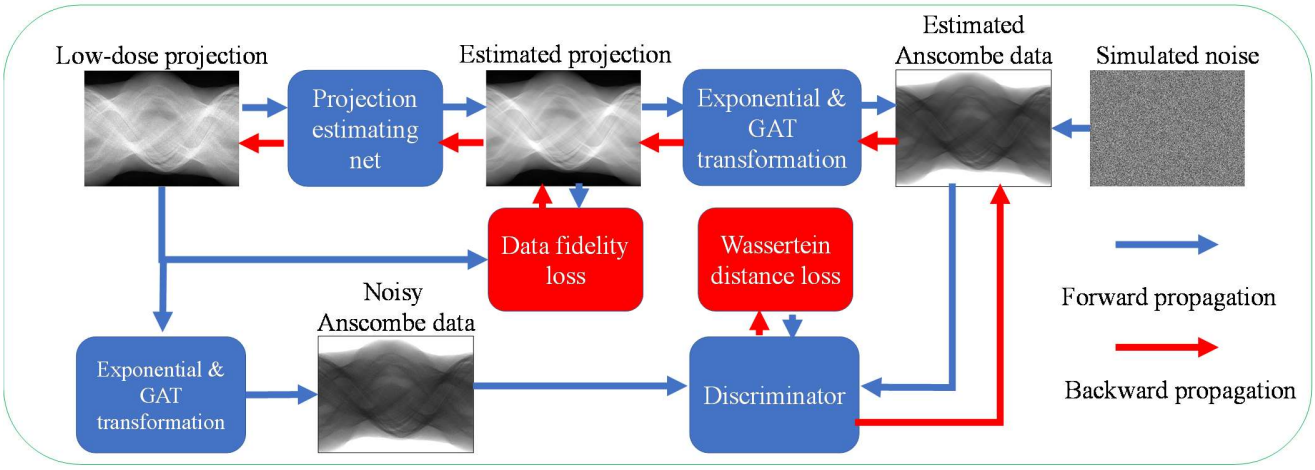


Fig. 1: The working flow of semi-blind denoising GAN.

Here $\mathcal{F}(\bullet)$ denotes a characteristic function of probability distribution. According to this, assuming the characteristic functions are none-zero, either distribution characteristic function can be inferred with the other two known. In our case, we are to estimate NDCT projections by incorporating the distribution of $\mathbf{A}$ that is about clean data. Thus,

$$\mathcal{F}_{\tilde{\mathbf{A}}+\mathbf{N}} = \mathcal{F}_{\tilde{\mathbf{A}}}\mathcal{F}_{\mathbf{N}}$$

$$\mathcal{F}_{\mathbf{A}_\mathbf{N}} = \mathcal{F}_{\mathbf{A}}\mathcal{F}_{\mathbf{N}}$$

$$\therefore \mathcal{F}_{\tilde{\mathbf{A}}+\mathbf{N}} = \mathcal{F}_{\mathbf{A}_\mathbf{N}} \Leftrightarrow \mathcal{F}_{\tilde{\mathbf{A}}} = \mathcal{F}_{\mathbf{A}} \quad \text{if} \quad \mathcal{F}_{\mathbf{N}} \neq 0 .$$

Now, with Eq. (7), we can work on $f_{\mathbf{A}_\mathbf{N}}$ that can be obtained from noisy data. This relaxes the needs of clean or high dose projection data that is hard or expensive to acquire.

C. *Network implementation and training*

In implementation, we construct a network to estimate clean projection data from corresponding noisy projection:

$$\hat{\mathbf{p}} = \varphi_{\mathrm{Es}}\left(\mathbf{p}_{\mathrm{N}};\boldsymbol{\theta}_{\mathrm{Es}}\right) \tag{9}$$

Here, $\mathbf{p}_{\mathrm{N}}$ denotes the noisy projection data, φ_{Es} the network operator with parameter set $\boldsymbol{\theta}_{\mathrm{Es}}$. The network is trained in an unsupervised manner with the loss function defined in Eq. (7). We adopt Wasserstein GAN [23] to minimize the distribution distance with samples. The total training is implemented as Eq. (10).

$$\min_{\boldsymbol{\theta}_{\mathrm{Es}}} \min_{\boldsymbol{\theta}_{\mathrm{D}}} \sum_{m=1:\mathrm{M}} \left[\ell_{\mathrm{f}}\left(\mathbf{p}_{\mathrm{N},m}, \mathbf{p}_m\right) + \lambda\left(\varphi_{\mathrm{D}}\left(\mathbf{a}_{\mathrm{N},m}\right) - \varphi_{\mathrm{D}}\left(\mathbf{a}_m + \mathbf{n}_{\mathrm{s}}\right)\right)\right] / M \tag{10}$$

Here, $\mathbf{a}_{\mathrm{N},m} = \mathcal{A}\left(I_0 \exp(-\mathbf{p}_{\mathrm{N},m})\right)$ is the GAT of LDCT projection $\mathbf{p}_{\mathrm{N},m}$, $\mathbf{n}_{\mathrm{S}}$ is a randomly generated noise sample from known noise model (thus, the independency is guaranteed), φ_{D} denotes a discriminator network with parameter set $\boldsymbol{\theta}_{\mathrm{D}}$.

The total network flowchart is displayed in Fig. 1. Here, we use a standard four-stage UNet for the projection estimation. The discriminator is a traditional classification network with four convolution layers each followed by a max-pooling down-sampling and two fully connected layers. It takes randomly extracted image patches of size 64 × 64 as inputs.

III. Experiments and Results

We conducted our experiments on low dose dental CT images. We collected real 3D dental CT images of 93 patients. Each patient's data is 640×640 in cross-section and 240 slices in total. The voxel size is 0.25^3 mm^3. We simulated clean projections of fan-beam CT scanning with ray-tracing method slice-by-slice. In simulation, both the source-to-center distance and detector-to-center distance were set 400 mm. There are 640 detectors with bin size 0.512 mm. We set 720 views uniformly distributed over 2π. The clean projection is used as ground-truth for performance evaluation but **not** used in network training. We further simulated low-dose projections with Poisson noise of 5000 incident photons. The reconstruction images were of the same scale of original patient data. Among the 93 patients, data of 85 patients were used to train our unsupervised denoising network, the other eight were used to test the performance of network.

We compared the proposed DDLN with projection-domain supervised deep learning and projection-domain Hanning-window filtering. All the methods were conducted based on sinogram data.

Fig. 2 displays one case of projection denoising results from a specific view of detector plane. The difference can be clearly seen in the projection residual images. Hanning-window filtering reduced noise to a certain extent. The residual error is still obvious with clear coherence. The supervised-learning denoising method achieves best results in projection-domain denoising task with little structure loss. This is expected because it is trained with full knowledge of labeled data. The DDLN also gives obviously denoised projection. However, the DDLN result has bigger residual error compared with supervised denoising in projection domain.

We further reconstruct theresults of different projection denoising method with FBP reconstruction. The reconstruction results of a slice are displayed in Fig. 3. Although supervised learning denoising achieved best results in projection domain, it seems slightly over-smoothed after reconstructions. The proposed DDLN method gives the most valuable reconstruction result with detail structures well restored.

IV. Conclusion

In this work, we proposed a distribution prior based MAP method for data denoising. We first proved that clean signal distribution can be inferred from onoisy samples. We further developed an unsupervised denoising network for LDCT projection data. The proposed method is validated with

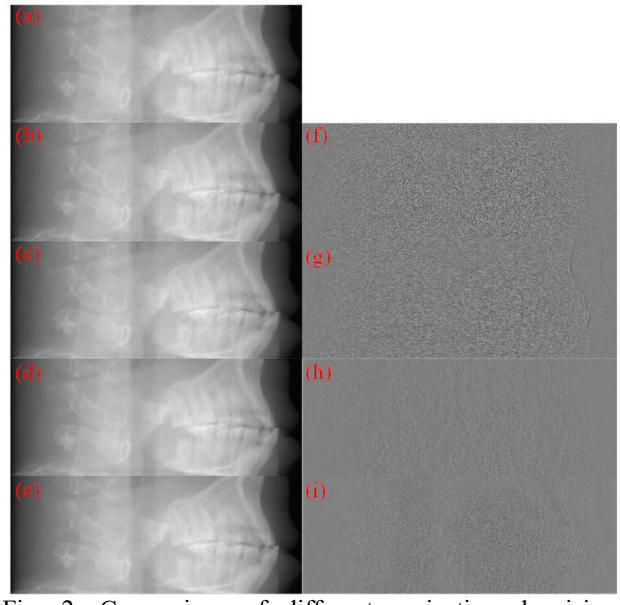


Fig. 2: Comparison of different projection denoising methods. (a) Ground-truth projection. (b) Low-dose projection. (c) Hanning-filter denoised results. (d) Supervised denoising network results. (e) DDLN results. (f), (g), (h), (i) are corresponding residual errors between (b), (c), (d), (e) and (a).

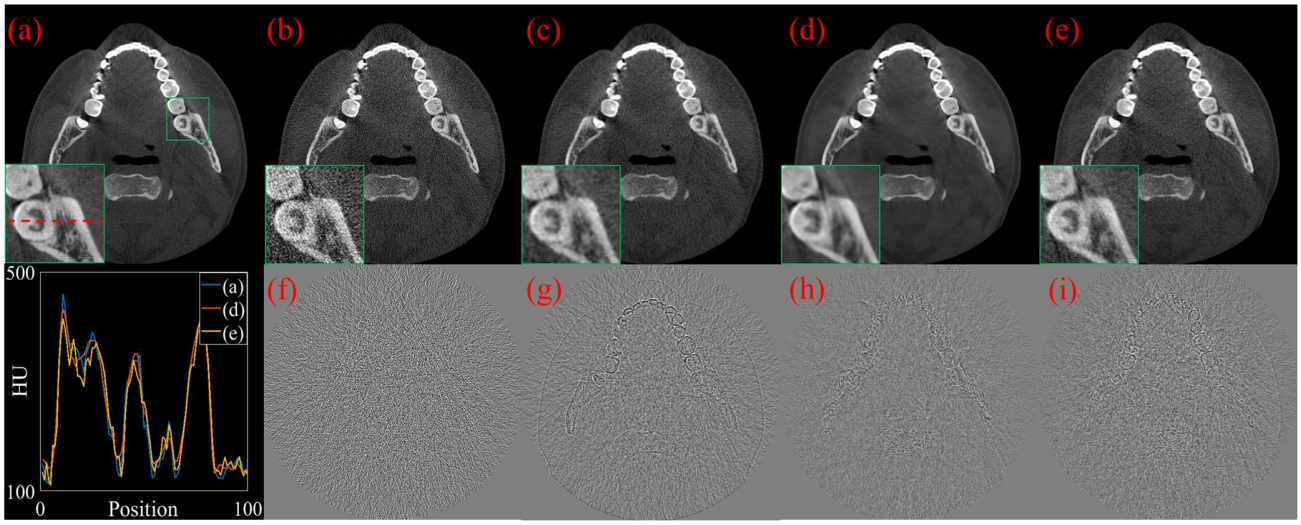


Fig. 3: Comparison of FBP reconstructions of projections from different methods. (a) From ground-truth projection. (b) Directly from low-dose projections. (c) From Hanning-filter denoised projection. (d) From results of supervised denoising network. (e) From DDLN denoised data. (f), (g), (h), (i) are corresponding residual errors between (b), (c), (d), (e) and (a). Lower left corner: the profiles of the red line in zoom-in from (a), (d) and (e).

simulated LDCT data with realistic phantoms. Very encouraging results in both projection domain and image domain are obtained. This framework can be extended to various imaging modality and unsupervised deep learning methods.

Reference:

[1]H. Lu, L. Xiang, et al., "Analyical noise treatment for low-dose CT projection data by penalized weighted least-square smoothing in the K-L domain." Chinese Journal of Medical Physics, vol. 4682, no. 4, pp. 146-152, 2002.

[2]A. Manduca, L. Yu, et al., "Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT." Medical Physics, vol. 36, no. 11, p. 4911-4919, 2009.

[3]J. Wang, H. B. Lu, et al., "Sinogram noise reduction for low-dose CT by statistics-based nonlinear filters." Medical Imaging: Image Processing SPIE, vol. 5747, p. 2058-2066, 2005.

[4]K. Ichikawa, H. Kawashima, et al., "A three-dimensional cross-directional bilateral filter for edge-preserving noise reduction of low-dose computed tomography images." Comput. Biol. Med., vol. 111, p. 9, 2019, Art. no. UNSP 103353.

[5]D. Kang, P. Slomka, et al., "Image denoising of low-radiation dose coronary CT angiography by an adaptive block-matching 3D algorithm." Medical Imaging: Image Processing SPIE, vol. 8669, 2013.

[6]Z. Zhuang, C. Yang, et al., "Fast Low-Dose CT Image Processing Using Improved Parallelized Nonlocal Means Filtering." International Conference on Medical Biometrics (ICMB), 2014.

[7] Y. Zhang, J. Rong, et al., "Low-Dose Lung CT Image Restoration Using Adaptive Prior Features from Full-Dose Training Database." IEEE Transactions on Medical Imaging, vol. 36, no. 12, p. 2510-2523, 2017.

[8]Y. Liu, J. Ma, et al., "Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction." Physics in Medicine & Biology, vol. 57, no. 23, p. 7923-7956, 2012.

[9]S. Niu, G. Yu, et al., "Nonlocal low-rank and sparse matrix decomposition for spectral CT reconstruction." Inverse Problems, vol. 34, no. 2, 2017.

[10] R. Zhang, H. Y. Dong, et al., "A Gaussian Mixture MRF for Model-Based Iterative Reconstruction with Applications to Low-Dose X-ray CT." IEEE Transactions on Computational Imaging, vol. 35, no. 3, pp. 359-374, 2016.

[11] Q. Xu, H. Yu, et al., "Low-Dose X-ray CT Reconstruction via Dictionary Learning." IEEE Transactions on Medical Imaging, vol. 31, no. 9, p. 1682-1697, 2012.

[12] H. Chen, Y. Zhang, et al., "Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network (RED-CNN)." IEEE Trans Med Imaging, vol. 36, no. 12, p. 2524-2535.

[13] E. Kang, W. Chang, et al., "Deep Convolutional Framelet Denosing for Low-Dose CT via Wavelet Residual Network." IEEE Trans Med Imaging, vol. 37, no. 6, p. 1358-1369, 2018.

[14] Q. Yang, P. Yan, et al., "Low Dose CT Image Denoising Using a Generative Adversarial Network with Wasserstein Distance and Perceptual Loss." IEEE Trans Med Imaging, vol. 37, no. 6, p. 1348-1357, 2018.

[15] C. Zhang, Y. Xing, "CT artifact reduction via U-net CNN." Medical Imaging SPIE, vol. 10574, 2018.

[16] J. He, Y. Yang, et al., "Optimizing a Parameterized Plug-and-Play ADMM for Iterative Low-Dose CT Reconstruction." IEEE Trans Med Imaging, vol. 38, no. 2, p. 371-382, 2019.

[17] D. Wu, K. Kim, et al., "A Cascaded Convolutional Neural Network for X-ray Low-dose CT Image Denoising." Arxiv: 1705: 04267, 2017.

[18] H. S. Park, J. Baek, et al., "Unpaired image denoising using a generative adversarial network in X-ray CT.", IEEE Access, vol. 7, p. 110414-110425, 2019.

[19] D. Wu, K. Kim, et al., "Iterative Low-Dose CT Reconstruction With Priors Trained by Artificial Neural Network." IEEE Transactions on Medical Imaging, vol. 36, no. 12, pp. 2479-2486, 2017.

[20] K. Liang, L. Zhang, et al. "A Self-supervised Deep Learning Network for Low-Dose CT Reconstruction.", IEEE Nuclear Science Symposium and Medical Imaging Conference, 2018.

[21] L. Fu, T. Lee, et al., "Comparison between Pre-log and Post-log Statistical Models in Ultra-Low-Dose CT Reconstruction." IEEE Trans Med Imaging, vol. 36, no. 3, p. 707-720, 2017.

[22] M. Mäkitalo, A. Foi, "Poisson-Gaussian Denoising Using the Rxact Unbiased Inverse of the Generalized Anscombe Transformation." IEEE International Conference on Acoustics, Speech and Signal Processing, 2012.

[23] J. Xin, Y. Xing, "A noise variance estimation approach for CT." Conference on Developments in X-Ray Tomography SPIE, 2012.

[24] M. Arjovsky, S. Chintala, et al, "Wasserstein GAN." Arxiv: 1701.07875, 2017.

X-RAY CT SCATTER CORRECTION BY A PHYSICS-MOTIVATED DNN WITH OPPOSITE VIEW PROCESSING

Berk Iskender and Yoram Bresler

Coordinated Science Laboratory and Dept. of ECE, University of Illinois at Urbana-Champaign

ABSTRACT

Photon scattering in X-ray CT creates streaks, cupping, shading artifacts and decreased contrast in the reconstructions. We describe a novel physics-motivated deep-learning-based method using opposite views to estimate and correct scatter. The method incorporates an initial reconstruction, and the sum and the difference of opposite scatter-corrupted projections. Numerical experiments show significant improvement over a recent projection-by-projection-based deep neural network de-scattering method.

***Index Terms*—** X-ray scatter estimation, computed tomography (CT), scatter correction, DNN, deep learning, monte carlo.

1. INTRODUCTION

Scatter in x-ray CT manifests itself in the reconstruction as streaks, cupping, shading artifacts and decreased contrast. Hardware-based scatter mitigation (de-scattering) methods, such as those based on a detector grid or on primary modulation can be successful in specific settings but require modifications in the acquisition hardware and increase in scan time or dose. Hence, many efforts have been devoted to develop software-based methods for scatter estimation, which can be categorized as Monte-Carlo (MC)-based, analytical, kernel-based, and data-driven.

MC methods sample photon paths stochastically for a given object. They operate iteratively on an initial reconstruction to estimate and correct the scatter until convergence. Due to the high computation cost of MC the trade-off between accuracy and run time is often not practical. Analytical methods for estimating scatter include a slice-by-slice approach [1] modelling scatter as a distance-dependent blurring effect and a recent iterative method [2], which approximates the solution to the Linear Boltzmann Transport Equation. While faster than MC, it too trades-off accuracy for run time depending on discretization and number of iterations.

Kernel-based scatter estimation methods use convolution of a weighted total projection or of an initial primary estimate with specific kernels to model the spreading due to scatter [3, 4]. They are computationally efficient, but prior assumptions and predefined kernels restrict their effectiveness.

Finally, data-driven approaches in the projection domain utilize neural nets to estimate scatter [5, 6]. Recent work also includes a correction method for Cone Beam CT (CBCT) with two-step registration (CT-CBCT pairs) [7]; A CBCT method using a scatter library of breast CT to estimate scatter [8]; and a two network approach that learns scatter by separating it into high and low frequency components [9]. However, these methods use little or no information of initial reconstruction, which limits their effectiveness.

In this paper, we present a new scatter correction algorithm for X-ray CT based on a DCNN that unlike previous DNN-based approaches, uses simultaneously both an initial reconstructed image and the raw projections. A similar approach was used in our recent work [10] to which we refer here as "Deep Scatter Estimation with Initial Reconstruction" (DSEIR). The work presented here improves on DSEIR by using a property of opposite projections in CT. The processing pipeline, network architecture and loss function for training are all inspired by the physics of X-ray scatter. For simplicity, the algorithm is presented for the parallel beam imaging geometry, but the approach extends to other CT geometries.

2. PROBLEM STATEMENT

The approach in this work concentrates on scatter correction and the results were obtained using a monochromatic parallel beam source setting. However, it can be extended to the cases of a polychromatic source (with the attendant beam hardening) and to other geometries. The latter extension can be done by rebinning to the parallel beam geometry for the scatter estimation, while still performing the reconstruction itself in the native geometry.

By Beer's law, in a 2D setting and with an energy-integrating detector, the *primary* (scatter-free) measurement at position $t \in \mathbb{R}$ at angle $\theta \in [0, 2\pi)$ is

$$p(t, \theta) = I_0 e^{-g(t,\theta)} \tag{1}$$

where I_0 is the vacuum (or air) fluence measurement, $g(t, \theta) = (\mathcal{R}f)(t, \theta)$ is the line integral projection of the linear attenuation coefficient map $f(\mathbf{x})$ of the object, and $\mathcal{R}$ is the 2D

This work was supported in part by the National Science Foundation (NSF) under Grant IIS 14-47879.

Radon transform. The entire projection at angle θ is denoted by g_θ, with analogous notation for other functions of (t, θ). Due to the additive contribution of scattered X-rays the *total* measurement at angle θ is $\tau_\theta = p_\theta + s_\theta$, where s_θ is the *scatter term*. This term, which is a nonlinear function of the object, causes artifacts in the reconstruction.

3. PROPOSED METHOD

Algorithm, loss function, & network structure. All measurements in the algorithm are normalized by the mean bright field, I_0, to make the method independent of I_0. The normalized total projection $\bar{\tau}_\theta \triangleq \tau_\theta / I_0$, scatter $\bar{s}_\theta \triangleq s_\theta / I_0$, and primary projection $\bar{p}_\theta \triangleq p_\theta / I_0$, are denoted by a bar.

We take advantage of the property of the parallel-beam geometry that for a given object, (normalized) line integral projections in opposite directions are identical, up to a flip (sign reversal) in t: $\bar{g}(t, \theta) = \bar{g}(-t, \theta + \pi)$. We call such a pair of projections π-*opposite* projections, and denote the t flipped version of $\bar{g}_\theta$ by $\hat{\bar{g}}_\theta$. Hence, $\bar{g}_\theta = \hat{\bar{g}}_{\theta+\pi}$. By (1) we also have $\bar{p}_\theta = \hat{\bar{p}}_{\theta+\pi}$, and therefore

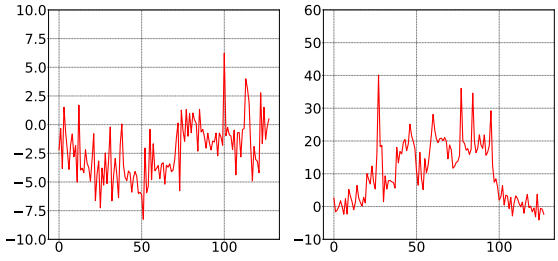
$$\bar{\tau}_\theta - \hat{\bar{\tau}}_{\theta+\pi} = \bar{s}_\theta - \hat{\bar{s}}_{\theta+\pi} \triangleq \Delta \bar{s}_\theta \tag{2}$$

It follows that the difference $\Delta \bar{s}_\theta$ of π-opposite scatter components can be determined *exactly* from the available (scatter-corrupted) total measurements. To fully determine each scatter component, we need only estimate the sum $\bar{b}_\theta \triangleq \bar{s}_\theta + \hat{\bar{s}}_{\theta+\pi}$. The reason that estimation of the sum $\bar{b}_\theta$ might be easier than estimation of each of its constituent scatter components, is that $\bar{b}_\theta$ is typically smoother than the latter, and therefore may be easier to learn by a DCNN. On the other hand, the difference $\Delta \bar{s}_\theta$ has high bandwidth, but as stated before, it need not be estimated, because it is determined exactly from the measured data. An example of $\bar{b}_\theta$ and $\Delta \bar{s}_\theta$ is shown in Fig. 1, illustrating the smoothness of the former relative to the latter. Another useful observation, is that from the relation of the π-opposite primaries, it also follows that $\bar{p}_\theta = (\bar{\tau}_\theta + \hat{\bar{\tau}}_{\theta+\pi} - \bar{b}_\theta)/2$.

Accordingly, the algorithm operates on sums and differences of normalized π-opposite total projections, $\bar{\tau}_\theta + \bar{\tau}_{\theta+\pi}$ and $\bar{\tau}_\theta - \bar{\tau}_{\theta+\pi}$ and produces estimates of the sum $\bar{b}_\theta$ of normalized π-opposite scatter and of the primary $\bar{p}_\theta$. The block diagram of the algorithm is illustrated in Fig. 2. Let the set of normalized total measurements be $\{\bar{\tau}_\theta, \theta \in \Theta\}$, with $\bar{\tau}_\theta \in \mathbb{R}^d$, and $\Theta = \{2k\pi/K, k = 0, 1, \ldots K-1\}$, with K even denoting the set of K view angles. Given $\{\bar{\tau}_\theta, \theta \in \Theta\}$, an initial (scatter-corrupted) reconstruction $\tilde{f} \in \mathbb{R}^{d^2}$ of the linear attenuation coefficient is computed as

$$\tilde{f}(\mathbf{x}) = (\mathcal{R}^{-1} \tilde{g})(\mathbf{x}) \tag{3}$$

where $\mathcal{R}^{-1}$ denotes FBP and $\tilde{g}(t, \theta) = \max\{-\ln \bar{\tau}(t, \theta), 0\}$ is an initial estimate of the line integral projection, which is constrained to be positive - a physics-based constraint. This constraint may become active on low attenuation paths, where due to scatter $\tau(t, \theta)$ might exceed I_0.



(a) (b)

Fig. 1: An example of sum and difference of scatter for opposite views (a) $\Delta s_\theta = s_\theta - \hat{s}_{\theta+\pi}$. (b) $b_\theta = s_\theta + \hat{s}_{\theta+\pi}$.

The DCNN, $\mathcal{N}_\gamma$, operates view-by-view and is used $K/2$ times, for $\theta = k(2\pi/K), k = 0, 1, \ldots, K/2 - 1\}$ It takes three inputs: (i) sum $\bar{\tau}_\theta + \hat{\bar{\tau}}_{\theta+\pi}$ of normalized pre-log π-opposite total measurements at view angle θ; (ii) difference $\bar{\tau}_\theta - \hat{\bar{\tau}}_{\theta+\pi}$ at the same angle; and (iii) the initial reconstruction estimate rotated by θ, $\tilde{f}_\theta \in \mathbb{R}^{d^2}$. The network produces an estimate $\bar{b}^*_\theta \in \mathbb{R}^d$ of the sum of the normalized π-opposite scatter components

$$\bar{b}^*_\theta = \mathcal{N}_\gamma \left(\tilde{f}_\theta, \bar{\tau}_\theta + \hat{\bar{\tau}}_{\theta+\pi}, \bar{\tau}_\theta - \hat{\bar{\tau}}_{\theta+\pi} \right). \tag{4}$$

The normalized primary projection is then estimated as $\bar{p}^*_\theta = \max\{0, \bar{\tau}_\theta + \bar{\tau}_{\theta+\pi} - \bar{b}^*_\theta\}/2$, with a non-negativity physical constraint. Note that $\bar{\tau}_\theta$ at the input to the DCNN is not clipped to 1 since $\bar{\tau}_\theta > 1$ is physically possible, and provides useful information for estimation of the scatter. Finally, the reconstruction estimate is obtained using FBP with Shepp-Logan filtering, $f^*(\mathbf{x}) = (\mathcal{R}^{-1} g^*)(\mathbf{x})$.

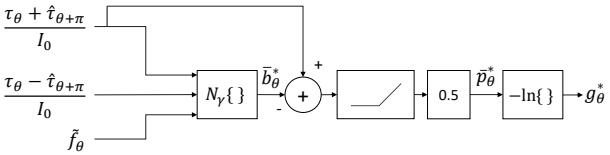


Fig. 2: Block diagram of the algorithm

The network is trained by minimizing the loss function

$$L(g_\theta, g^*_\theta) = \min_\gamma ||h * (g_\theta - g^*_\theta)||_2^2 + \lambda ||g_\theta - g^*_\theta||_1 \tag{5}$$

which includes a filtered l_2-norm and an l_1-norm regularizer, where $g^*_\theta = -\ln \bar{p}^*_\theta$ is the estimated line integral projection, $g_\theta = -\ln \bar{p}_\theta$ is the true line integral projection, and $h[n] = 0.5\delta[n+1] - 0.5\delta[n-1]$. Since h is high-pass, the l_1-norm term (with $\lambda > 0$) encourages recovery of the DC value of the projections. The loss function is tailored to minimize errors in the reconstructed image rather than in the estimated primary in two ways. First, the post-log quantities g are used in the loss function rather than the primary measurements p.

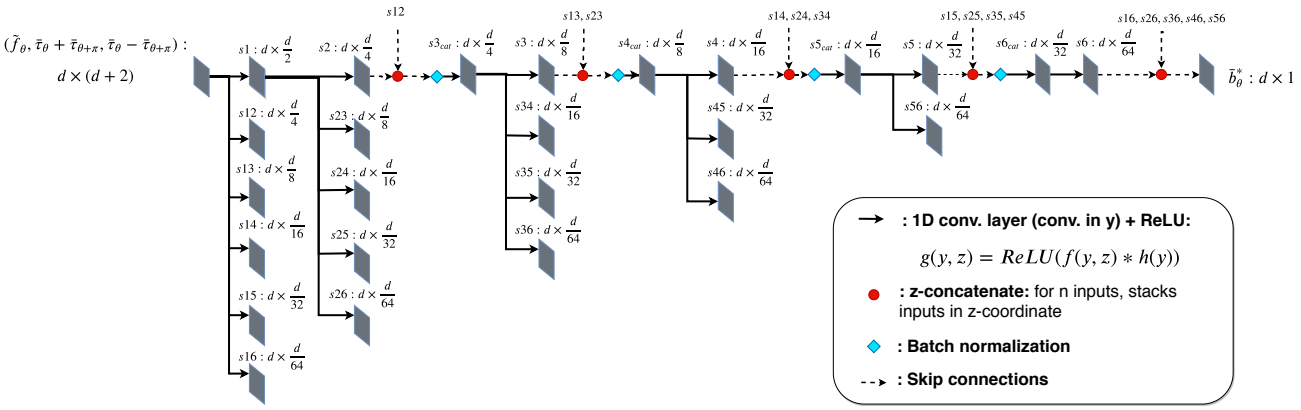


Fig. 3: Proposed network architecture

Second, the particular h enables to express a frequency-weighted norm of f, $||Qf||_2$ where Q is a filter with a angularly-invariant frequency response, in terms of the projections g_θ of f, and thus use the first term in (5) to penalize errors in the image domain. This eliminates the need to include the FBP in the cost function and in the backpropagation for training the network. Suppose f is obtained from sampled projections $g_d[n,\theta] = g(nT,\theta)$ using FBP with a ramp filter apodized by window $w[n]$. Denote the discrete-time Fourier transforms with respect to n of $g_d[n,\theta]$ and $w[n]$ by $G(\nu,\theta)$ and $W(\nu)$, respectively. Owing to the perceptual significance of edges, Q could be a high-pass filter. Assuming K uniformly spaced view angles, using Parseval's identity and the projection-slice theorem, it can be shown that

$$\|Qf\|_2^2 = \frac{1}{2\pi}\sum_{k=0}^{K-1}\int_{-\pi}^{\pi}|G(\nu,k\frac{2\pi}{K})H(\nu)|^2 d\nu \qquad (6)$$

where $H(\nu) = W(\nu)Q(\nu)|\nu|^{0.5}$. If $Q(\nu) = |\nu|^{0.5}$, we have $H(\nu) = W(\nu)|\nu|$. Since the magnitude is computed in (6), $H(\nu) = W(\nu)\nu$ can be chosen. This removes the discontinuity in $\frac{dH}{d\nu}$ at $\nu = 0$ and results in a short implementation of $h[n]$. Particularly, for $W(\nu) = sinc(\nu)$ (Shepp-Logan filter) we have $H(\nu) = \sin(\nu)$ and $h[n] = 0.5\delta[n+1] - 0.5\delta[n-1]$.

The proposed network architecture operates on rotated versions of the initial reconstruction $\tilde{f}_\theta$, one angle θ at a time, by performing 1D convolutions along the coordinate perpendicular to the X-ray propagation, and contracts along the other dimension. This structure is inspired by the slice-by-slice approach [1], where the object is divided into layers perpendicular to the primary X-ray propagation, and scatter is modeled as an incremental blurring effect using the Klein-Nishina formula and computed as the convolution of the output of the respective layer with pre-specified kernels to produce the estimate for the next layer. The second and third inputs to the network, $\bar{\tau}_\theta + \hat{\bar{\tau}}_{\theta+\pi}$ and $\bar{\tau}_\theta - \hat{\bar{\tau}}_{\theta+\pi}$, are concatenated with $\tilde{f}_\theta$, as two additional channels. At the m-th step, for $m = 1, \ldots \log_2 d$, the network performs separate convolutions compressing the input by successive factors of 2 to a set of outputs with number of channels as $\{2^{-m}d, \ldots, 1\}$ using different layers, and conveys the output of those to the following steps via skip connections, concatenating all outputs from previous steps that have the same number of channels (e.g. s_{14} is concatenated with every s_{m4}, $m < 4$ and s_4 at the end of the 4th step in the diagram, each having $2^{-4}d$ channels). After concatenation in every $m \geq 2$ step, the number of channels are reduced to $2^{-m}d$ by another layer.

Computational Cost at Inference (Test Time). Convolutions dominate the cost of the DCNN, which is bounded by $3Ld^3$, where L is the length of the 1D filters. The DCNN is used $K/2$ times, where K is the number of measurements uniformly spaced in $[0, 2\pi)$, resulting in a bound of $3LKd^3/2$. The cost of FBP is cd^3 where $c \approx 1$, as is the cost of d rotations of $\tilde{f}$. The total cost can be approximated by $3LKd^3/2$.

4. DATA GENERATION & EXPERIMENTS

Framework. The primary projections p_θ of phantoms were obtained in a voxelized environment using (1). Total measurements τ_θ were simulated using GATE , a simulation environment utilizing the well-known GEANT4 MC libraries, in a monoenergetic 200 keV source setting with a 128×128 pixels detector plane of physical size of 128×128 cm^2. In the phantoms, objects of rectangular prisms and cylinders with rectangular and circular cross-sections were used with centers and dimensions randomized, allowing intersections. Aluminium, steel and water were used as materials. 27 synthetic phantoms were used for training, each with $K = 360$ views (a total of $27\times360 \approx 10^3$ 1D projections) each having $P = 2\times10^6$ photons. Measurements were averaged along the x-axis to generate reduced variance 1D counterparts. The data was divided into training and validation sets phantom-wise, grouping all

measurements belonging to a phantom into same set.

The proposed algorithm was trained and validated on the first $K/2$ views using (5) and compared with two DCNN-based methods: (i) a recent projection-by-projection-based method [5] (DSE), which uses a modified UNET structure; and (ii) our own recent method DSEIR [10] described in the introduction, both trained on all K views. The network for [5] was adapted to 1D projections to prevent overfitting, and is dubbed here as 1D-DSE and the same workflow was utilized. The DCNN in the proposed algorithm and in DSEIR has $\sim 75k$ parameters, about half those of 1D-DSE with $\sim 185k$ parameters. All FBPs used the Shepp-Logan filter. $\lambda = 0.05$ was selected. The total average run time of the algorithm for a phantom with $K = 360$ views was 0.13 s. However, the two FBPs and 180 image rotations, implemented in Python on the CPU, respectively account for 0.06 s and 0.07 s, whereas the DCNN run time for all views was only 2ms on the NVIDIA GeForce GTX TITAN X GPU (less than 1% of total). Reconstruction quality, as compared to the reference image (FBP of numerically computed line integrals of a test phantom) is quantified using three metrics: PSNR (in dB) calculated as the ratio of peak reconstruction value to root mean square error; MAE (mean absolute error); and SSIM (structural similarity index). The networks were implemented in Pytorch, and the Adam optimizer was used for all methods for training.

Results. In Fig.4, reconstruction obtained from p_θ (reference), difference figures and comparison of 1D line profiles from reconstructions using p_θ and p^*_θ for different methods for one test phantom are shown. The display range for reference reconstruction is set to $\approx 50\%$ of the maximum value in HU (to provide visibility for all objects) and the range for difference figures is set to $\approx \pm 20\%$ of the actual range of the reference reconstruction, centered at 0 HU. Magnitude errors due to scatter are clearly visible in Fig.4(b) and 4(f). which cause a major drop in PSNR. The 1D-DSE method decreases these errors but introduces streaks and does not perform well especially on steel regions, reflected in low SSIM and high MAE. Our DSEIR method [10] performs close to the proposed method by providing less streaking artifacts. However, better performance on steel regions enables the proposed method to have a higher PSNR.

Table 1: Reconstruction Accuracy

	Uncorr.	1D-DSE	DSEIR	Proposed
PSNR	31.05	37.19	43.83	**45.90**
SSIM	0.92	0.90	**0.98**	**0.98**
MAE	42.4	38.3	**14.5**	15.3

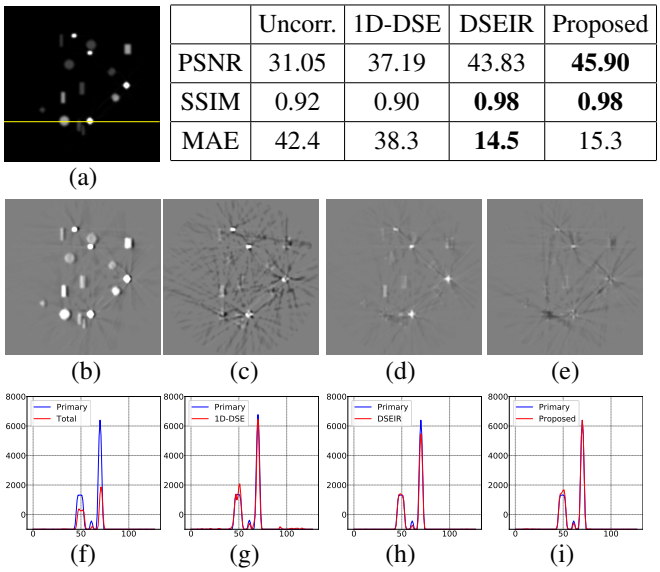


Fig. 4: Reconstruction results for a test phantom. 2nd row reconstruction error, HU window [-800, 800], 3rd row line profiles (along yellow lines in 1st column). (a) Using primary measurements, p_θ, HU window [-1000, 2700] (Gold reference). (b,f) Using total measurements, τ_θ. (c,g) Using primary measurements p^*_θ estimated by the 1D-DSE method [5]. (d,h) Using primary measurements p^*_θ estimated by our recent DSEIR (e,i) Using primary measurements p^*_θ estimated by the proposed π-opposite projections method.

5. REFERENCES

[1] C. Bai et al., "A slice-by-slice blurring model and kernel evaluation using the Klein-Nishina formula for 3D scatter compensation in parallel and converging beam spect," *Phys. Med. Biol.*, vol. 45, no. 5, 2000.

[2] A. Maslowski et al., "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter–Part I: Core algorithms and validation," *Med. Phys.*, vol. 45, no. 5, 2018.

[3] B Ohnesorge et al., "Efficient object scatter correction algorithm for third and fourth generation CT scanners," *Europ. radiology*, vol. 9, no. 3, 1999.

[4] W. Zhao et al., "Patient-specific scatter correction for flat-panel detector-based cone-beam CT imaging," *Phys. Med. Biol.*, vol. 60, no. 3, 2015.

[5] J. Maier et al., "Deep Scatter Estimation (DSE): Accurate real-time scatter estimation for X-ray CT using a deep convolutional neural network," *Jour. of Nondest. Eval.*, vol. 37, no. 3, 2018.

[6] Joscha Maier et al., "Real-time scatter estimation for medical ct using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation," *Med. Phys.*, vol. 46, no. 1, 2019.

[7] S. Xie et al., "Scatter artifacts removal using learning-based method for CBCT in IGRT system," *IEEE Access*, vol. 6, 2018.

[8] L. Shi et al., "Library based x-ray scatter correction for dedicated cone beam breast CT," *Med. Phys.*, vol. 43, no. 8, 2016.

[9] H. Lee and J. Lee, "A deep learning-based scatter correction of simulated X-ray images," *Electronics*, vol. 8, no. 9, 2019.

[10] B. Iskender and Y. Bresler, "A physics-motivated DNN for X-ray CT scatter correction," in *IEEE Int. Symp. Biomed. Imag. (ISBI 2020)*. IEEE, to appear.

Projection domain decomposition for photon counting CT

Gilad Shechter[1]

***Abstract*—Photon counting detector (PCD) CT prototypes have a few advantages over energy integrating detector (EID) CT systems. In addition to their superior spectral separation, using smaller detector pixels in these systems allows reaching a spatial resolution of up to about 30 cm^{-1}. Using small pixels, however, is increasing the noise of the individual readings. This noise, together with the non-linearity of the projection domain decomposition can result in a severe bias that is propagated to the spectral images. In this work, we briefly describe some means that help to avoid this bias and require only a small computational effort. We also show some results for non K-edge spectral imaging.**

***Index Terms*—Computed tomography**

I. INTRODUCTION

Projection domain decomposition [1] is used widely today in spectral CT systems based on different strategies like kVp switching [2], dual-layer (DL) detector [3], and photon counting detector (PCD). The equivalent paths decomposed in these systems are then reconstructed to create basis spectral images such as photo/scatter or water/iodine images. Spectral results of clinical relevance, e.g. virtual mono-energetic (monoE) images are then derived from the basis images. Multi-bin PCD detector of direct conversion offers also decomposition of a three-element image basis for K-edge imaging [4]. In addition, the spectral separation between the PCD energy bins is typically much larger than that in EID CT. This helps to reduce dramatically the noise level in the non K-edge spectral results compared to EID systems. As a result, residual low-frequency noise left in the images after using image domain de-noising algorithms [5-6], becomes in many cases much smaller in PCD systems. Using small detector pixels in PCD systems helps to further reduce the image noise for a fixed spatial resolution of the image. This behavior is formulated mathematically in [7], and has the following intuitive explanation: The smaller cross-section area of the beam, obtained by convolving the focal spot and detector active areas with each other along the beam, enables reaching a fixed image spatial resolution using smoother filtering during the reconstruction. Using PCD detector pixels of about 0.5 mm x 0.5 mm also allows reaching a spatial resolution of up to about 30 cm^{-1}. Using the small detector pixel size is challenging due to the charge sharing effect, see e.g. [8]. In this work, we focus on another challenge that stems from using small pixels. Using small pixels, the noise of the individual sinogram readings used as an input to the projection domain decomposition is increased. This decomposition is non-linear, see e.g. [9], Appendix A. Therefore, feeding it with highly noisy input can lead to severe noise induced bias (NIB) added to the decomposed readings. In Eq. (1), we give a formulation of this bias.

(1)
$$E\{D(c_{\{b\}}^{n}(i))\} \neq D(E\{c_{\{b\}}^{n}(i)\}) \ .$$

Here, $c_{\{b\}}^{n}(i)$ are the noisy counts of all energy bins b corresponding to one sinogram reading i, i.e. to the same ray, D stands for the non-linear decomposition, and $E\{\ \}$ stands for an expectation value over different noise realizations. Inspired by the Ergodic hypothesis, we can replace the expectation values taken in (1) over different noise realizations of the same reading by an average taken across neighbor readings. The bias expressed in (1) can propagate from the decomposed sinograms to the spectral images, and deteriorate many spectral results. For example the spectral results based on iodine to calcium separation. In the next section, we present a few means that help to remove this bias.

The outline of this paper is as follows: In section II, we describe the parameters selected for the simulation. Then, we review the steps and the parameters used during processing the simulated data. We pay here a special attention to the decomposition step that can lead to NIB. In section III, we show some spectral results, and in section IV, we summarize our work.

II. METHODS

A. Simulation

The mathematical pixelated phantom we use for the simulation is shown in Fig. 1. To the original NCAT body phantom, we add 10 mg/cc of iodine within the blood vessels. To liver pixels that fall out of the blood vessels, we add 0.5 mg/cc of iodine. The length of the dashed red line is about 375 mm. We simulate a helical scan of this phantom using a 64 rows detector, assuming a pixel size of 0.5 x 0.5 mm square. We simulate the scan in a dual focal spot mode, using a helical pitch of 1.0, 262

Paper submitted for review on January 19, 2020.
[1]Gilad Shechter, Philips medical technologies LTD., Haifa, Israel (E-mail: gilad.shechter@philips.com).

mAs, and 1200 readings per rotation for each spot position. The source-to-detector (rotation center) distance is taken as 1040 (570) mm, respectively. The focal spot size is 0.6(0.7) mm along $x(z)$, respectively. In the simulation, we neglect scattered radiation. We also neglect the pile-up effects and their complications [10]. We also assume no electronic noise. To get our spectral model we convolve the spectrum that is emitted from the tube with the detector response modeled according to measurements [11]. We use a two-bin configuration, setting the threshold between the electronic two bins at 72 keV. According to Cramer-Rao lower bound calculations, using more bins helps to increase a bit the spectral separation and therefore to improve the SNR of the decomposed readings by about 10-15%. This improvement is however much smaller than that obtained later on using the image domain de-noising algorithms [5-6]. Using a two-bin configuration helps on the other hand to increase the number of counts of the individual readings, and therefore to reduce the NIB. It also simplifies the use of a decomposition look-up table discussed briefly in the next section.

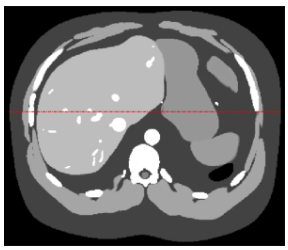


Figure 1. MonoE image of the phantom at 40 keV.
WW (WL) = 500(0) HU.

In Fig. 2, we show the two-bin spectrum. To appreciate better the spectral separation, we harden our modeled spectrum by letting it pass through an absorber that is equivalent to the phantom used. For this purpose, we average angularly the ray paths that intersect the phantom center. This way, we find this absorber to be equivalent to about 380 mm of the water photoelectric absorption mechanism plus 280 mm path of the water scatter mechanism. The signal is taken simply as the counts collected by each of the two-sided energy bins, and is expressed in units of 1/sec./mA/solid angle/keV. Applying a zero moment (center of mass) calculation on the curves in Fig. 2, we find the spectral separation between the two bins is about (22.3) keV. In addition, only about 51% of the area under the two curves is an overlap region. These values are quite larger than the corresponding values in typical EID CT systems.

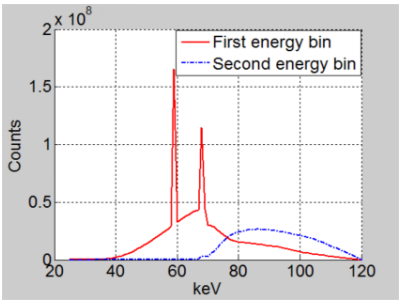


Figure 2. Spectra of the two-bins, hardened by the phantom.

B. Processing

Pre-decomposition de-noising

As a first step, we apply a pre-decomposition de-noising. This way, we manage to reduce not only the noise of the sinogram, but also the NIB. Reducing the NIB is similar to some extent to the reduction of NIB done for EID sinograms. See, e.g. [12]. While in the PCD, sinograms the counts are decomposed to equivalent paths, in EID, the sinogram signals are transformed to line integrals by using the non-linear log operation. Various de-noising methods can be used before the decomposition. One of these methods is based on a total variation minimization [12]. Another method is based on detecting and removing outlayers by comparing the differences between the counts of the two bins belonging to neighbor readings [13].

Decomposition

Performing the decomposition for a configuration of more than two energy bins using a maximal likelihood approach is described in [4]. Once only two bins are used, this approach is reduced to solving numerically a problem of two measured variables and two unknown variables. Here, the measured variables are the counts in each energy bin, and the unknown variables are the two decomposed equivalent paths corresponding to the two basis elements of the decomposition. These basis elements can be taken as the energy-dependent attenuation profiles of a pair of materials, e.g. Iodine and water. Here, we use however, a different basis that is defined according to the photoelectric and scatter attenuation mechanisms in water.

Various means to reduce the NIB

So far, we assumed no regularization during the decomposition step. Yet, adding a regularization term that penalizes huge decomposed paths offers the suppression of the residual NIB left after the pre-decomposition de-noising. Another way to avoid diverging decomposed equivalent paths is to use pre-calculated look-up tables (LUTs). Here, huge path values are replaced by extrapolations performed over the counts (or on the logged counts). Using such LUTs helps also to fasten the decomposition time. A third mean to reduce NIB can be any of the NIB correction methods described in [14]. Consider for example the method captured in Eq. (2).

(2)

$$p_d^{corrected}(i) = p_d(i) - p_d^{low-freq.}(i) + D\left(c_b^{low-freq.}(i)\right);$$
$$c_b^{low-freq.}(i) = c(i) \otimes ker_{smooth};$$
$$p_d^{low-freq.}(i) = p_d(i) \otimes ker_{smooth}\ .$$

Let $p_d(i)$ be the sinogram of equivalent paths describing one the decomposition basis elements. Taking advantage of the low-frequency nature of the bias within the sinogram domain, we correct only the low frequencies of this sinogram. Here, we use

the fact that for getting the low frequencies of $p_d(i)$, we can filter out the high frequencies of the decomposition input, thus reducing its noise. As in (1) $D(\,)$ stands for the decomposition operator. Here ker_{smooth} is a local smoothing 2D FIR filter spanned along the detector rows and columns within the sinogram. While $p_d^{low-freq.}(i)$ is heavily contaminated by the bias, $D\left(p_c^{low-freq.}(i)\right)$ is not, due to filtering out the high-frequency noise before decomposition.

Having the bias-free decomposed equivalent path, we reconstruct them into images using fan-to-parallel re-binning followed by focal spot interleaving, radial (equidistant) re-binning, convolution (filter), z-filtering and back-projection. The reconstructed slice width (slice sensitivity profile FWHM) determined by the z filtering is chosen as 1 mm. The convolution filter selected is shown in Fig. 3. Using a field of view of 430 mm and 1024x1024 matrix size, the Nyquist frequency associated with the image pixel size is 11.9 cm^{-1}. In agreement with Fig. 3, the filter MTF is reaching a value of 0.5 right at this frequency. To end this section, in Fig. 4 we list the processing steps used.

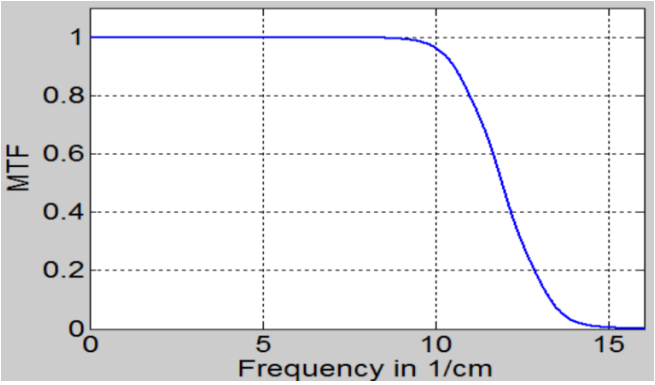


Figure 3. Reconstruction filter used.

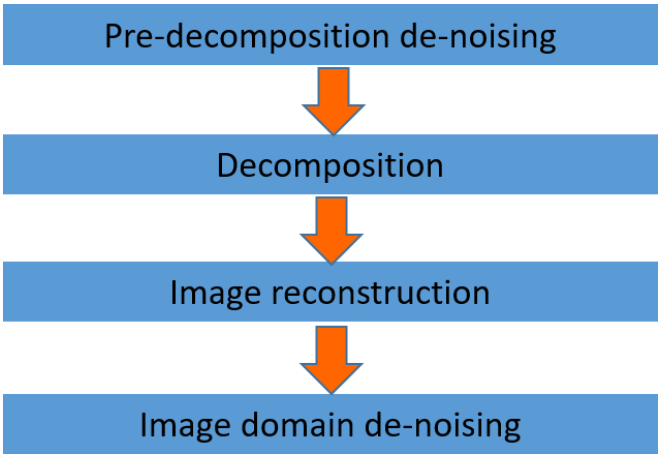


Figure 4. Processing flow-chart.

III. Results

In Fig. 5, we show photo/scatter images reconstructed from a noiseless simulation. For quantifying the NIB, we consider these images as the gold standard reference. In Fig. 6, we omit these images from corresponding noisy images obtained after applying a pre-decomposition de-noising. Some of the methods to reduce the NIB described in the previous section that are missing in (a) are used in (b). It is clear that in order to get rid of the bias, pre-decomposition de-noising must be combined with other means.

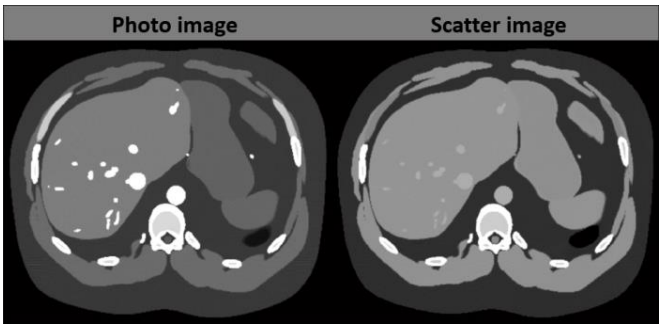


Fig. 5 Noiseless simulation images. WW=200 HU, WL=-875 (-75) HU for the photo (scatter) image.

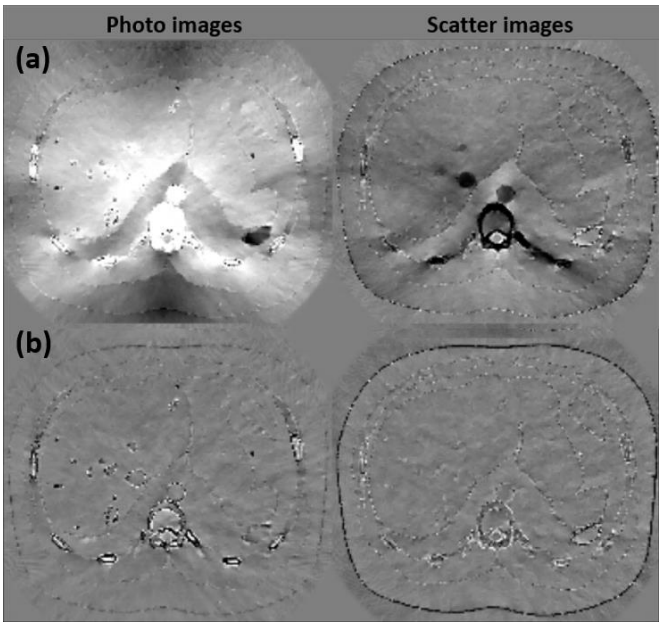


Figure 6. Difference images. WW=100 HU.

We categorize the residual patterns shown in (b) into two classes: patterns that correlate visually with the object structure, and those that do not. The patterns belonging to the first class obtained in different noise realizations hardly correlate with each other see Fig. 7. Therefore, we conclude that these patterns are dominated by residual low-frequency noise remaining after image de-noising, and not by consistent NIB. Due to the superior spectral separation of the PCD system, the amplitude of these patterns is much smaller than that observed in EID systems. The patterns belonging to the second class (see e.g. the darkening of the cortical bone of some of the ribs in the photo image), are originating from a small over-smoothing performed during the pre-decomposition de-noising and not due to NIB.

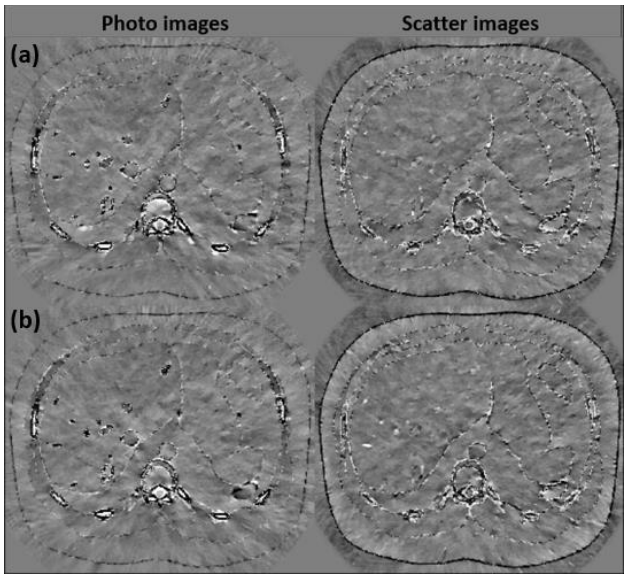


Fig. 7. Difference images in two different noise realizations. The upper row is as row (c) in Fig. 6, yet here, WW=50 HU.

Next, we demonstrate the iodine to calcium classification offered using some of the means to remove NIB described in the previous section. For this purpose, we select four groups of neighbor pixels contained in different liver blood vessels. In Fig. 8, we color these pixels in red. In Fig. 9, we draw a scatter plot of the values of these pixels taken from the photo and scatter images. In the absence of NIB, the pixels fall on the average right on the iodine line as expected. Yet, fluctuations of residual low-frequency noise are responsible for the spread of the values of these pixels along the scatter (horizontal) axis. In the case presented, this spread is yet small enough to promise an accurate iodine/calcium classification.

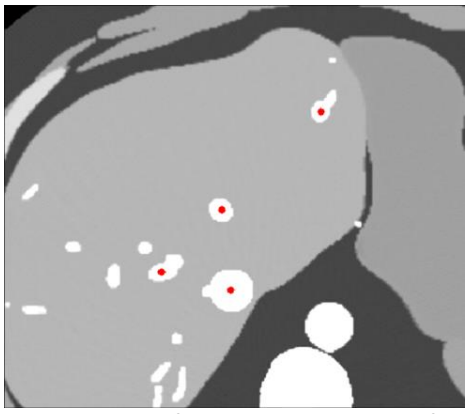


Figure 8. Selected pixels for iodine/calcium classification. The selected pixels are colored in this zoom in on a noiseless monoE image at 70 keV. WW (WL) = 300 (0) HU.

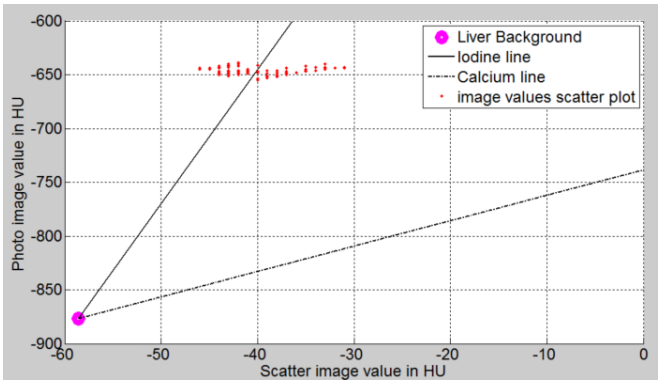


Figure 9. HU-HU scatter plot of the selected image pixels. All these pixels belong to the liver blood vessel.

IV. Summary

We have described a few techniques for suppressing the NIB in projection domain decomposition for PCD CT. We have shown that using only pre-decomposition de-noising is not powerful enough to suppress this bias. Yet, in combination with some of the techniques described, we have shown the ability to remove completely this bias.

References

[1] "Energy-selective reconstructions in X-ray computerized tomography", Robert E. Alvarez and Albert Macovski, Phys. Med. Biol., **21**, pp. 733-744, (1976).

[2] "Monochromatic CT image representation via fast switching dual kVp", X. Wu, Proc. SPIE Medical Imaging, **7258** (2009).

[3] "A dual-energy CT based on a double layer detector", Ami Altman and Raz Carmi, AAPM Annual Meeting, (2013).

[4] "K-edge imaging in x-ray computed tomography using multi-bin photon counting detector", Ewald Roessl and Roland Proksa, Phys. Med. Biol., **52**, pp. 4679-4696, (2007).

[5] "Structure propagation restoration for spectral CT", Liran Goshen, US patent 2015379694(A1), (2015).

[6] "Anti-correlated noise filter", Kevin Brow, Liran Goshen, Asher Gringauz, and Stanislav Zabic, patent WO2016103088 (A1), (2016).

[7] "Pre-sampling, algorithm factors, and noise: Considerations for CT in particular and for medical imaging in general", Marc Kachelriess and Willi. A. Calender, Med. Phys. **32** (5), pp. 1321-1334, (2005).

[8] Christer Ullberg et al, Photon counting, dual energy X-ray imaging at CT count rates - measurements and implications of in-pixel charge sharing correction, SPIE Phys. of Med Imag. 1057319 (2018).

[9] "Quantization of liver tissue in dual-kVp computed tomography using linear discriminant analysis", J. Eric Tkaczyk et al., Phys. Med. Imag, Proc. SPIE, 72580G, pp. 1-12, (2009).

[10] "Mean and variance of single photon counting with deadtime", Daniel F. Yu and Jeffrey A. Fessler, Phys. Med. Biol. **45** pp. 2043-2056, (2000).

[11] "Characterization of CZT performed at ESRF", Heiner Daerr et al. from Royal PHILPS (Grenoble), (2014).

[12] "Projection data de-noising", Kevin M. Brown and Stanislav Zabic, US patent 9031299(B2), (2011).

[13] "Spectral imaging", Kevin M. Brown et al, US patent 20170278278(A1), (2015).

[14] "X-Ray imaging device for an object", Bernhard Brendel, Gilad Shechter, Liran Goshen and Thomas Koehler, US patent 10275859(B2), (2016).

Filtered Backprojection Reconstruction for Inverse Geometry CT with Sparsely Distributed Sources

Xiao Jiang, and Lei Zhu*

***Abstract*—Inverse geometry CT (IGCT) has demonstrated advantages on improved image quality with optimized radiation dose using densely distributed sources and a small detector. In practical implementations, a sparse distribution of sources is preferred on an IGCT system to lower the manufacturing cost. However, rebinning based reconstruction methods cannot apply for IGCT with sparsely distributed sources due to the insufficient samplings. Although the existing direct filtered-backprojection (FBP) method for IGCT avoids data rebinning, it significantly reduces the field-of-view size and requires strictly equidistant sources. In this paper, we propose a more flexible and practical FBP reconstruction for IGCT with sparsely distributed sources. The projections of IGCT are considered as data from multiple scans on the same object with severe truncation. A practical algorithm of weighting calculation is then developed on projections to suppress non-uniform data redundancy. IGCT images are finally reconstructed using FBP on weighted projections. Both simulation and phantom studies reveal excellent numerical accuracy and image spatial resolution of the proposed method.**

***Index Terms*—filtered backprojection, inverse geometry CT, multi-source**

I. Introduction

Inverse geometry CT (IGCT) employs a distributed source array and a small detector, leading to reduced photon scatter as well as cone-beam artifacts [3, 4]. The distributed sources also enable the technique of virtual bowtie filtration [5], which not only lowers imaging dose but also relaxes the requirement of detector dynamic range [6]. Many researches related to IGCT have been published, with focus on both hardware integration [6-9] and relevant algorithms [10-12]. A dense distribution of x-ray sources is required in the original concept of IGCT. In this work, we aim to develop a practical reconstruction algorithm for IGCT with sparsely distributed sources.

Source array plays an important role in IGCT architecture. Conventional source with thermionic cathodes is not suitable for integration in a small area, such as 100-by-20 sources in a 25cm-by-5cm area[13], and the carbon nanotube (CNT) [14], a potential substitution of the conventional source, is not technically mature yet for wide use on human imaging with acceptable cost. Detailed discussion of state-of-art multi-source technology is beyond the scope of this work, and an extensive review can be found in Ref. [15]. To lower the manufacturing cost, a sparse distribution of x-ray sources is preferred on IGCT systems. For example, on the IGCT system developed by the GE corporation [6, 9], which is currently the only gantry-based IGCT prototype system to our best knowledge, the sources have a sparse distribution of 16-by-2 sources in an area of 37.5cm-by-10cm.

The sparse distribution of x-ray sources poses new challenges of IGCT reconstruction. Rebinning based methods [13, 16] are designed for densely distributed sources with sufficient samplings, and their performances significantly degrade especially for IGCT with longitudinally sparsely distributed sources, such as the system in Ref. [6]. As proposed in Ref. [17], direct FBP reconstruction avoids data rebinning, by pre-calculating gain vectors to merge all the projections into an entire reconstruction. To avoid data truncation and enable precise projection filtering, this method views detector and sources conversely, resulting in significant decrease of the field of view (FOV) since the scanned objects should be covered by each source (detector pixel). Smaller detector is therefore needed to enlarge FOV, which in turn, requires the denser sources. Moreover, this method is sensitive to the geometric error due to the strict requirement of equidistant sources. Thus, this method is not suitable for the IGCT system with sparsely distributed sources, either.

In this paper, we propose a flexible and practical FBP reconstruction for IGCT with sparsely distributed sources. Simulation studies show that proposed method outperforms previous FBP method in numerical accuracy, and phantom studies show that proposed method presents images with spatial resolution comparable to fan-beam CT images.

II. Methods

A. Formulation of the Proposed Reconstruction

In an IGCT system with sparsely distributed sources, the number of sources is much smaller than the number of detector pixels, and each source provides part of the full size sinogram. Though the collection of all the projections could cover the full-size sinogram, the sampling patterns of different sources do not match due to the different source locations, therefore

Research reported in this publication was supported by the National Natural Science Foundation of China (No. 81671681), the Ministry of Science and Technology of China Key Research and Development Projects (Grant No. 2016YFC0101400) and by the Fundamental Research Funds for the Central Universities (No. WK2030040089).

Xiao Jiang is with the Department of Engineering and Applied Physics, University of Science and Technology of China, Hefei, China (e-mail: jx2015@mail.ustc.edu.cn).

*Corresponding author, Lei Zhu, is with the Department of Engineering and Applied Physics, University of Science and Technology of China, Hefei, China (e-mail: leizhusg@ustc.edu.cn).

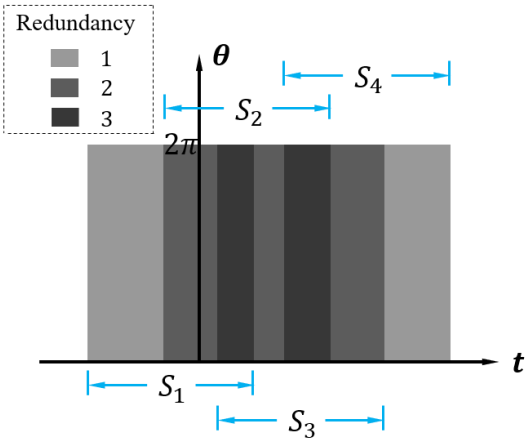


Fig. 1. Illustration of the data redundancy in a single circular scan of IGCT. t and θ represent signed distance from projection ray to the isocenter and angle between projection ray and negative y-axis, respectively.

conventional FBP reconstruction for single source cannot be used directly.

Considering the irregular sampling patterns, we conceive to independently reconstruct images from each source via conventional FBP and then merge them into the final reconstruction. However, due to the data truncation and non-uniform data redundancy, direct reconstruction leads to severe ring artifacts, and this undesirable phenomenon could be eliminated by the delicately designed weightings. In the Ref. [17], the authors propose to view the sources and detector conversely and iteratively calculate the weightings using a reference object. In this paper, an analytical and practical algorithm is used to design the weightings that applies to raw projection directly. Using the designed weightings, the proposed reconstruction could be formulated as:

$$F(x,y) = \sum_{i=1}^{N} f\left(W_i(s,\alpha)p_i(s,\alpha)\right) \tag{1}$$

where N is the number of sources, p_i is the projection of the ith source, F is the reconstructed image, s, α are coordinates of the detector pixels and rotation angle, respectively, and $f(p)$ denotes the offset-source FBP reconstruction[18, 19] using the projection p.

B. Weighting Calculation

Now we focus on how to design weighting function $W_i(s,\alpha)$ that normalizes the redundancy and avoids data truncation simultaneously. Firstly, we analyze the data redundancy in the sinogram domain. As illustrated in Fig. 1, during a single circular scan, each source covers a rectangle area in sinogram, which spans from 0 to 2π in θ direction, and the t coordinates are determined by the source location. The overlap between different rectangles represent data redundancy. Obviously that the projections have non-uniform redundancy, which is also intuitively reflected on the different greyscale in Fig. 1. Since the non-uniform redundancy only occurs along the t direction, 1D weightings is enough to remove it, i.e., $W_i(s,\alpha) = W_i(s) = w_i(t(s))$. For 2 sources, the redundancy could be simply eliminated via two functions, $w_1(t), w_2(t)$:

$$w_1(t) = \begin{cases} 1, & T_{1l} \le t < T_{2l} \\ 1 - \dfrac{t-T_{2l}}{T_{1r}-T_{2l}} + \dfrac{1}{2\pi}\sin\left(2\pi\dfrac{t-T_{2l}}{T_{1r}-T_{2l}}\right), & T_{2l} \le t < T_{1r} \\ 0, & T_{1r} \le t \le T_{2r} \end{cases} \tag{2}$$

$$w_2(t) = \begin{cases} 0, & T_{1l} \le t < T_{2l} \\ \dfrac{t-T_{2l}}{T_{1r}-T_{2l}} - \dfrac{1}{2\pi}\sin\left(2\pi\dfrac{t-T_{2l}}{T_{1r}-T_{2l}}\right), & T_{2l} \le t < T_{1r} \\ 1, & T_{1r} \le t \le T_{2r} \end{cases} \tag{3}$$

where $[T_{1l}, T_{1r}], [T_{2l}, T_{2r}]$ represent the coverage of S_1, S_2 along the t direction, respectively. Since $w_1(t)+w_2(t) = 1$, these weightings could normalize the data redundancy. It is also seen that both $w_1(t)$ and $w_2(t)$ are second-order differentiable and smoothly vary to 0 at the detector edge, which ensures no data truncation. The same strategy could be applied on weighting assignments for arbitrary number of sources. We recursively weight the first k sources and the $(k+1)$th source, i.e., view the combination of the first k sources and the $(k+1)$th source as S_1 and S_2, respectively, and the corresponding weightings are denoted as $w_{1k}(t)$ and $w_{2k}(t)$, which are the transformation of $w_1(t)$ and $w_2(t)$ based on the coverage of the first k sources and the $(k+1)$th source. According to the proposed algorithm, the weighting W_i assigned to S_i could be analytically written as:

$$W_i(s) = \begin{cases} \prod_{k=1}^{N-1} w_{1k}(t(s)), & i = 1 \\ w_{2(i-1)}(t(s)) \prod_{k=i}^{N-1} w_{1k}(t(s)), & 2 \le i \le N-1 \\ w_{2(i-1)}(t(s)), & i = N \end{cases} \tag{4}$$

$$t(s) = \frac{SAD \cdot s + (SID - SAD) \cdot x}{\sqrt{SID^2 + (s-x)^2}} \tag{5}$$

where the x is the source offset distance, N is the number of sources. Note that $w_{1k}(t) + w_{2k}(t) = 1$, it can be easily verified that the W_i is also second-order differentiable and $\sum_{i=1}^{N} W_i = 1$. Compared with the method in Ref. [17], propose method discard the requirement that each source (detector pixel) should cover the whole FOV. As long as the collection of projections cover the whole sinogram, i.e., redundancy≥ 1, sources could be arranged as sparse as possible. Moreover, once we make use of the redundancy of the conjugated rays, i.e., $p(t,\theta) = p(-t,\theta+\pi)$, the sparsity of the source could be further increased, for which proposed method could still calculate the weightings based on the same strategy.

III. Results

A. Simulation Study

In simulation study, we simulate projections of the Shepp-Logan phantom in four IGCT systems with different source sparsity, which indicates different data redundancy. Poisson noise equivalent to 10^5 incident photons per ray is added on the raw projections. Fig. 2 displays the source distributions and corresponding calculated weightings. The source spacing used

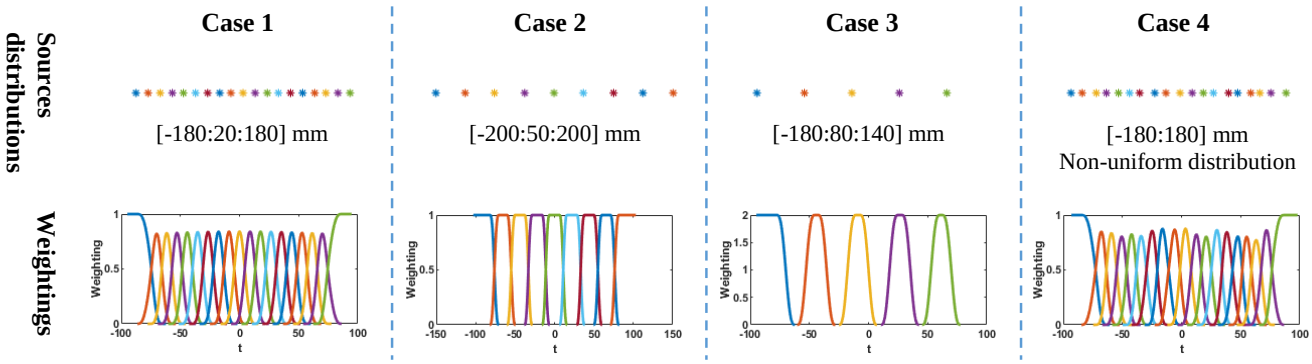


Fig. 2. Four IGCT systems used in simulation studies. The source distribution is illustrated in the first row, with corresponding weightings in the second row.

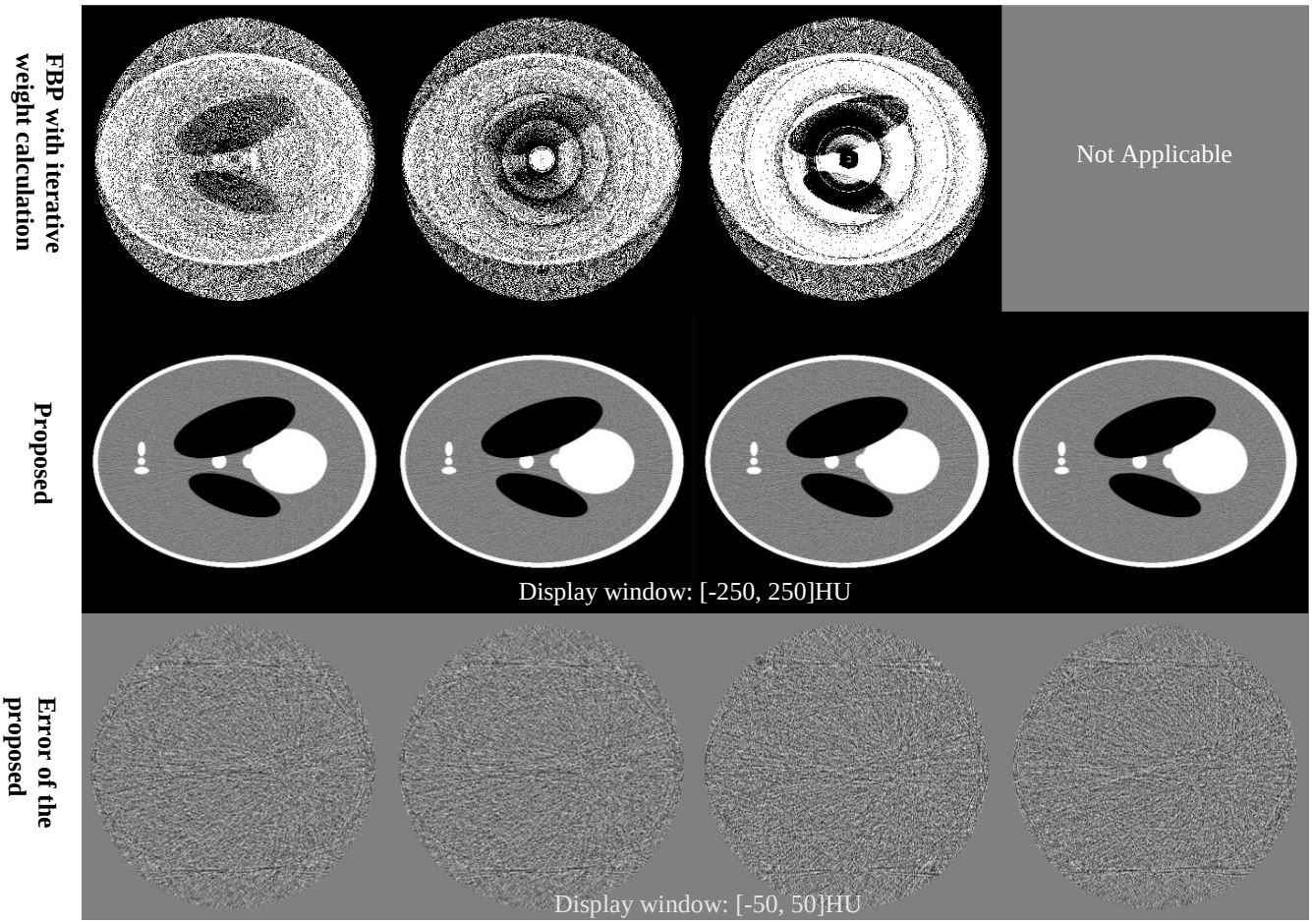


Fig. 3. Reconstructions of Shepp-Logan phantom in four IGCT systems indicated by Fig. 2. The first and second row are reconstructions via FBP with iterative weight calculation [17] and the proposed method, respectively. The last row is the difference images between the second row and the ground-truth.

in previous paper are usually shorter than $1mm$, to highlight the effectiveness of the proposed method in sparse source system, we adopt $20mm$ and $50mm$ source spacing in the case 1 and case 2. Case 3 represents an even sparser distribution that no data redundancy among sources, for which proposed method could generate effective weightings via making full use of redundancy between conjugate rays. In case 4, we explore the reconstruction of non-uniform source distribution which may occurs in the practical use when the sources are approximately equidistant due to slight geometry error. Reconstructions of the previous FBP method[17] are also added for comparison, with results in the first row of Fig .3. Due to the data truncation of each source and resulting inaccuracy in projection filtering, previous method presents severely contaminated images, while the proposed method provides satisfactory reconstruction even the sources are non-uniform distributed. The error images in the

last row of Fig .3 displays only noise pattern without obvious object structure, indicating precise numerical accuracy and splendid edge preservation of the proposed reconstruction.

B. Catphan Phantom Study

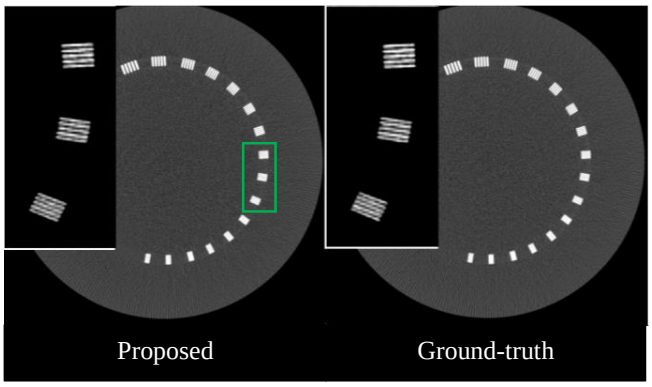


Fig. 4. Reconstructions of the line-pair module with inserted zoom-in images (for line pairs of 13-15 lp/cm). Display window: full-size: [-400 1200] HU; zoom-in: [800 1400]HU.

We further evaluate proposed method in a tabletop x-ray system. The system has 7 degrees of freedom, capable of mimicking the IGCT scan via moving the single source, and the IGCT geometry in case 1 is simulated on this system to scan the Catphan phantom, with the results shown in the Fig. 4. We compare the IGCT images with the conventional fan-beam images which is commonly referred as ground-truth. The zoom in images of the box area shows that proposed method achieves almost the same visibility of line-pairs as the ground-truth, which reveals the excellent performance of the proposed method in image spatial resolution.

IV. Conclusion and Discussion

In this paper, we propose a practical FBP reconstruction for IGCT with sparsely distributed sources. This algorithm avoids the drawbacks of existing FBP methods since it relaxes the requirement of equidistant sources and that the scanned object need not be covered by each source (detector pixel). Both simulation and phantom studies reveal excellent numerical accuracy and image spatial resolution of the proposed method.

Though we only perform evaluations in 2D situation, this method could be easily generalized to 3D situation, which may formulate as summation of offset-source FBP reconstructions. In the future, we will investigate proposed method in cone-beam IGCT systems, and more studies on anthropomorphic phantoms is need to further assess the proposed method. Noted that our algorithm does not rely on the sparsity of source distribution explicitly, which means that this method could apply for arbitrary IGCT system with data redundancy. However, time consumption may become a great concern as the number of sources increases, since the filtering and backprojection time is proportional to the number of sources. Derivative backprojection filtration (DBF) is a potential substitution of FBP, which accelerates filtering and backprojection via cancelling the zero-padding step. Related works will be carried on in the following researches.

References

[1] H. Xu and T. Zhuang, "A novel volume CT with x-ray on a trough-like surface and point detectors on circle-plus-arc curve," in *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2001, vol. 3, pp. 2363-2366: IEEE.

[2] T. G. Schmidt, R. Fahrig, N. J. Pelc, and E. G. Solomon, "An inverse‐geometry volumetric CT system with a large‐area scanned source: A feasibility study," *Medical physics*, vol. 31, no. 9, pp. 2623-2627, 2004.

[3] T. G. Schmidt, "What is inverse-geometry CT?," *Journal of cardiovascular computed tomography*, vol. 5, no. 3, pp. 145-148, 2011.

[4] S. Ryu, C. Lee, and J. Baek, "A new CT concept: inverse geometry CT," *Biomedical Engineering Letters*, vol. 4, no. 1, pp. 25-31, 2014.

[5] B. De Man *et al.*, "Multi-source inverse geometry CT: a new system concept for X-ray computed tomography," presented at the Medical Imaging 2007: Physics of Medical Imaging, 2007.

[6] B. De Man *et al.*, "Multisource inverse‐geometry CT. Part I. System concept and development," *Medical physics*, vol. 43, no. 8Part1, pp. 4607-4616, 2016.

[7] T. G. Schmidt *et al.*, "A prototype table‐top inverse‐geometry volumetric CT system," *Medical physics*, vol. 33, no. 6Part1, pp. 1867-1878, 2006.

[8] J. Baek *et al.*, "A multi-source inverse-geometry CT system: initial results with an 8 spot x-ray source array," *Physics in Medicine & Biology*, vol. 59, no. 5, p. 1189, 2014.

[9] V. B. Neculaes *et al.*, "Multisource inverse‐geometry CT. Part II. X‐ray source design and prototype," *Medical physics*, vol. 43, no. 8Part1, pp. 4617-4627, 2016.

[10] T. G. Schmidt, R. Fahrig, and N. J. Pelc, "Noise simulations for an inverse-geometry volumetric CT system," in *Medical Imaging 2004: Physics of Medical Imaging*, 2004, vol. 5368, pp. 420-428: International Society for Optics and Photonics.

[11] R. Bhagtani and T. G. Schmidt, "Simulated scatter performance of an inverse‐geometry dedicated breast CT system," *Medical physics*, vol. 36, no. 3, pp. 788-796, 2009.

[12] J. Baek, B. De Man, D. Harrison, and N. J. Pelc, "Raw data normalization for a multi source inverse geometry CT system," *Optics express*, vol. 23, no. 6, pp. 7514-7526, 2015.

[13] T. G. Schmidt, R. Fahrig, and N. J. Pelc, "A three‐dimensional reconstruction algorithm for an inverse‐geometry volumetric CT system," *Medical physics*, vol. 32, no. 11, pp. 3234-3245, 2005.

[14] X. Calderón-Colón, H. Geng, B. Gao, L. An, G. Cao, and O. Zhou, "A carbon nanotube field emission cathode with high current density and long-term stability," *Nanotechnology*, vol. 20, no. 32, p. 325707, 2009.

[15] V. B. Neculaes, P. M. Edic, M. Frontera, A. Caiafa, G. Wang, and B. De Man, "Multisource X-ray and CT: Lessons learned and future outlook," *IEEE Access*, vol. 2, pp. 1568-1585, 2014.

[16] S. R. Mazin and N. J. Pelc, "Fourier rebinning algorithm for inverse geometry CT," *Medical physics*, vol. 35, no. 11, pp. 4857-4862, 2008.

[17] J. Baek and N. J. Pelc, "Direct two‐dimensional reconstruction algorithm for an inverse‐geometry CT system," *Medical physics*, vol. 36, no. 2, pp. 394-401, 2009.

[18] L. A. Feldkamp, L. Davis, and J. W. J. J. a. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, no. 6, pp. 612-619, 1984.

[19] G. T. Gullberg, C. R. Crawford, and B. M. J. I. t. o. m. i. Tsui, "Reconstruction algorithm for fan beam with a displaced center-of-rotation," *IEEE transactions on medical imaging*, vol. 5, no. 1, pp. 23-29, 1986.

Single-Scan Dual-Energy CT Image Reconstruction via Normalized Gradient Magnitude Image (NGMI) Regularization

Donghyeon Lee, Hwanhui Jo, Uijin Jeong and Seungryong Cho

***Abstract*— Enabling dual-energy imaging capability in the existing single-energy CT systems has been under our keen interest. In this paper, we propose a stationary multi-slit filter based dual-energy CT imaging method and associated image reconstruction algorithm. The multi-slit filter placed between the x-ray source and the patient provides both high- and low-energy projection data, though partitioned into tiled segments, at each projection. A compressed-sensing-inspired iterative algorithm has been developed for image reconstruction. Joint sparsity of the imaged object or patient has been widely utilized mostly in the form of regularization. Particularly, we find that a normalized gradient magnitude image (NGMI) regularization helps improving image quality substantially. We demonstrate the feasibility of the proposed method by use of Monte Carlo simulation.**

***Index Terms*—Cone-beam computed tomography, dual-energy CT, image reconstruction, low-dose CT**

I. INTRODUCTION

Dual-energy computed tomography (DECT) can provide material-specific information that is not available in conventional single-energy computed tomography (SECT). There are increasing evidences and applications that such material-specific image information improves diagnostic performance [1]. Commercialized DECT systems typically use expensive technologies such as fast kV-switching, dual-source, dual-layer detector, or photon counting detector [2]–[4]. Additionally, they are all based on multi-row diagnostic CT systems with a fast rotating gantry. Not much commercialized progress has been reported in cone-beam CT (CBCT) systems that use flat-panel detector for dual-energy imaging. Research, however, seems to be active toward dual-energy CBCT imaging. For example, a C-arm CT system with rather a slow kV-switching has been implemented and has shown its feasibility [5].

This work was supported in part by Korean National Research Foundation [NRF-2016M3A9E9941837, NRF-2018M3A9E9024949, and NRF-2019M2A2A4A05031487].

Donghyeon Lee is with the Korea Advanced Institute of Science and Technology, Daejeon, 34141 Rep. of KOREA (telephone: +82)042-350-3868, e-mail: leedhyun85@kaist.ac.kr).

Hwanhui Jo is with the Korea Advanced Institute of Science and Technology, Daejeon, 34141 Rep. of KOREA (telephone: +82)042-350-3868, e-mail: okpc426@kaist.ac.kr).

Uijin Jeong is with the Korea Advanced Institute of Science and Technology, Daejeon, 34141 Rep. of KOREA (telephone: +82)042-350-3868, e-mail: uj.jeong@kaist.ac.kr).

Seungryong Cho is with the Korea Advanced Institute of Science and Technology, Daejeon, 34141 Rep. of KOREA (telephone: +82)042-350-3828, e-mail: scho@kaist.ac.kr). S. Cho is in correspondence of this paper.

Enabling dual-energy imaging capability in the existing single-energy CT systems, both in CBCT and diagnostic CT systems, has been under our keen interest. On top of dual-energy imaging, a low-dose CT scanning is highly desirable. We have earlier proposed a beam-filter-based sparse sampling approach for such a single-scan low-dose dual-energy CT, and successfully demonstrated its feasibility in a benchtop CBCT setting [6]. Beam-filter-based dual-energy imaging, by the way, is also adopted in a commercialized diagnostic CT system (Twin-beam CT, Siemens Healthineers) but is not necessarily low-dose driven [7]. The mean energy of the filtered x-ray spectrum is relatively shifted to a higher-energy compared to the unfiltered one. Therefore, dual-energy data acquisition is available at each projection by use of the multi-slit filter. The energy separation between the high- and low-energy spectra by use of multi-slit filter may not be as large as using different kVp settings, but has shown to provide clinically acceptable material decomposition performance [6], [8].

We have been working on beam-filter based sparse sampling approach for low-dose CT imaging [6], [9]–[11]. Although CT imaging in general does not directly follow the compressed-sensing requirements, image sparsity has been exploited in various iterative image reconstruction algorithms. When designing a sparse sampling scheme in CT, it has been shown that more uniform number of ray measurements per image pixel across the entire image field-of-view (FOV) and less correlated measurements per image pixel are desirable [10]. A static multi-slit filter turns out to yield more non-uniform and more correlated measurements compared to other sparse sampling schemes such as sparse-view sampling and reciprocating multi-slit filter, which makes image reconstruction more challenging [10]. Static multi-slit filter often results in strong ring artifacts in the reconstructed images. Reciprocating the beam filter during a scan can greatly increase the sampling uniformity and data incoherence, but such a motion can be mechanically challenging particularly in a fast gantry-rotation system.

In this work, we propose a static multi-slit filter design and an image reconstruction algorithm that can much reduce the ring artifacts. The filter, proposed in this work, guarantees the sampling uniformity for low-energy data. For high-energy image reconstruction, we propose a normalized gradient magnitude image (NGMI) regularization that effectively utilizes joint sparsity between dual-energy images. The use of joint sparsity between dual-energy images has been investigated in various forms [2], [6], [12], [13]. We propose to use NGMI by multiplying an attenuation normalizing factor in each GMI. This normalizing process turns out to greatly

reduce energy dependence of the GMI, thus leading to an improved material decomposition performance.

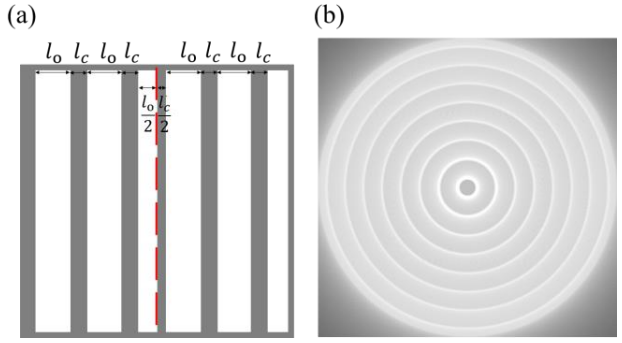


Fig. 1. (a) A front view of the proposed multi-slit filter. The red line is the intersection between the filter and the X-ray path toward the center of detector. (b) Sampling density in the field of view.

II. METHOD

A. Proposed Sampling Scheme

The filter design makes use of sampling density concept [6], [10]. Sampling density (SD) in this work refers to the weighted intersection lengths of the rays that pass through a given pixel in the image domain. It gives an idea how many ray measurements are made on a particular pixel. In this study, we used a ray-driven approach for computing SD. The intersection length of the *m*-th ray with the *n*-th image pixel having two sides is calculated as below:

$$d_{m,n} = \sqrt{(x^1_{m,n} - x^2_{m,n})^2 + (y^1_{m,n} - y^2_{m,n})^2}, \qquad (1)$$

where the points $(x^1_{m,n}, y^1_{m,n})$ and $(x^2_{m,n}, y^2_{m,n})$ represent the first and second intersection points with the corresponding vertical side and horizontal side of the pixel. Sampling density for the *n*-th pixel is the sum of the intersection lengths of rays, and is calculated as

$$SD_n = \sum_{m \in L_n} d_{m,n}, \qquad (2)$$

where L_n is a set of indexes of all rays, which pass through the *n*-th pixel.

We designed a multi-slit filter with two objectives in mind. The first objective is that unfiltered regions should cover at least one half regions of the detector particularly including the central line that corresponds to the rotation axis projection. This would lead to sufficient data for image reconstruction of low-energy CT. In other words, low-energy CT image reconstruction can be done by conventional (not necessarily CS-inspired) iterative algorithms or even by analytic algorithm such as backprojection-filtration algorithm [14]. Note that we assume there is no geometrical shift during a scanning, which may not be the case for real scanners. More adjustment of the filter design and data correction may be needed in such practical applications. Second, the filter strips that form multi-slits should be distributed as uniformly as possible across the horizontal direction. The uniform filter distribution helps to reduce x-ray scatter effects in projections. However, as the number of filter strips increases, penumbra effects caused by edges of the filters would also increase. Therefore, an appropriate number of filter strips should be empirically determined. In this numerical study, we used 6 strips as shown in Fig. 1 as an example. A front view of the designed filter is shown in Fig. 1 (a). The unfiltered width l_o to the filtered width l_c ratio is 6:4 in this work. We designed the multi-slit filter to cover one half regions of the detector by locating half-width of the strip on one side and half-width of the open region on the other side in the middle. The figure 1 (b) shows the sampling density map in the image FOV when using unfiltered projection ray measurements. We made the unfiltered region large enough to conservatively offset the penumbra effects, which causes that the brightest rings in FOV have redundant data for reconstruction. Although the sampling density map for the filtered ray measurements is not shown here, it is quite more non-uniform than the unfiltered one.

B. Image Reconstruction Scheme

We use a discrete-to-discrete imaging model with the reconstructed image $\vec{f}$ and the measured data $\vec{g}$. The cost function for image reconstruction is written by:

$$\vec{f}^* = \text{argmin}_{\vec{f}} \parallel A\vec{f} - \vec{g} \parallel \quad \text{s.t.}\, \vec{f} \geq 0, \qquad (3)$$

where $\vec{f}^*$ represents the satisfied solution, and A is the system matrix. The non-negativity constraint is added. We utilized the algebraic reconstruction technique (ART) to solve this problem [15]. For improving image quality in a given task, regularization $R(\vec{f})$ is often added as following.

$$\vec{f}^* = \text{argmin}_{\vec{f}} \parallel A\vec{f} - \vec{g} \parallel + R(\vec{f}) \quad \text{s.t.}\, \vec{f} \geq 0. \qquad (4)$$

In this study, we used the image total variation (TV) as a regularization for low-energy image reconstruction [16].

C. Normalized Gradient Magnitude Image (NGMI)

In DECT, minimization of GMI difference between the high-and low-energy CT images as a regularization was earlier proposed [6]. The GMI regularization from dual-energy images is defined as following:

$$\parallel GMI_{low} - GMI_{high} \parallel_2 = \sqrt{\sum \left(\sqrt{\left(f^{low}_{i,j,k} - f^{low}_{i-1,j,k}\right)^2 + \left(f^{low}_{i,j,k} - f^{low}_{i,j-1,k}\right)^2 + \left(f^{low}_{i,j,k} - f^{low}_{i,j,k-1}\right)^2} - \sqrt{(f^{high}_{i,j,k} - f^{high}_{i-1,j,k})^2 + (f^{high}_{i,j,k} - f^{high}_{i,j-1,k})^2 + (f^{high}_{i,j,k} - f^{high}_{i,j,k-1})^2} \right)^2}, \qquad (5)$$

where i, j, and k represent the pixel indices, and $\vec{f}^{low}$ and $\vec{f}^{high}$ are the reconstructed low- and high-energy images. Roughly speaking, a gradient operator in GMI can be interpreted as a high frequency emphasizing filter. The minimization of GMI difference takes a role that supports the algorithm to reconstruct high-frequency information of the high-energy image based on the low-energy image. However, the GMI highly depends on the x-ray source energy spectrum since CT pixel values of the corresponding image would be subject to the energy spectrum. The minimization of GMI difference between two images acquired at different energies may introduce unwanted edge enhancing effects including noise amplification. In this study, we normalized the GMI of high-energy CT image with respect to the low-energy CT to

remove such energy dependence. The proposed regularization term is given by

$$\| GMI_{low} - w^{(n)} GMI_{high} \|_2, \quad w^{(n)} = \frac{\|GMI_{low}\|}{\|GMI_{high}^{(n)}\|} \tag{6}$$

, where w is a normalizing factor. We utilized the gradient descent scheme with respect to each pixel to minimize the proposed NGMI. The total cost function for reconstruction of dual-energy images are shown below:

$$\vec{f}^{*low} = \underset{\vec{f}^{low}}{\operatorname{argmin}} \| A\vec{f}^{low} - \vec{g}^{low} \| + \alpha \| \vec{f}^{low} \|_{TV} \quad \text{s.t.} \vec{f}^{low} \geq 0 \tag{7}$$

$$\vec{f}^{*high} = \operatorname{argmin}_{\vec{f}^{high}} \| A\vec{f}^{high} - \vec{g}^{high} \| + \beta \| GMI_{low} - w^{(n)} GMI_{high} \|_2 \quad \text{s.t.} \vec{f}^{high} \geq 0. \tag{8}$$

where α, β are weighting parameters. The high-energy image reconstruction process is performed after the low-energy image reconstruction is done.

D. Simulation conditions

We simulated a fan-beam CT system and obtained dual-energy data via GATE, Geant4-based simulation tool for medical imaging. Simulated scanning parameters are summarized in table. 1. We used a poly-energetic spectrum at 80kVp simulated from an x-ray spectrum simulation program (SRS 78, spectrum processor, IPEM). For a material of the multi-slit filter, a 1mm molybdenum was used. A molybdenum filter effectively separates the energy spectrum because it has enough high Z number and has adequate k edges. The x-ray spectrum for the unfiltered and the filtered region are shown in Fig. 2. The 1mm molybdenum filter absorbs about 78% of the x-ray photons at the tube voltage of 80kVp.

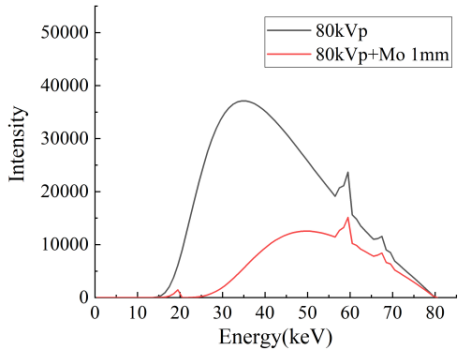


Fig. 2. Energy spectrum for the unfiltered region (80kVp) and the filtered region (80kVp+Mo 1mm).

E. Image quality assessment

We calculated contrast-to-noise ratio (CNR) to quantify the improvement of the image quality. Region-of-interest (ROI) and background region were selected for each contrast structures as shown in Fig. 4. CNR is defined by:

$$\text{CNR} = \frac{|\mu_{background} - \mu_{ROI}|}{\sigma_{background} + \sigma_{ROI}}, \tag{9}$$

where $\mu_{background}$ and μ_{ROI} refer to the mean of a ROI and a background region. $\sigma_{background}$ and σ_{ROI} are standard deviation of a ROI and a background region.

F. Material decomposition

Most of sparse DECT reconstruction methods including the proposed one are only compatible to image-domain material decomposition because material decomposition should be performed after reconstruction [2], [6], [12], [13]. We adopted an empirical calibration method for material decomposition proposed by P. Stenner et al. [17]. Please refer to [17] for more detailed method. We predetermined coefficients for two material bases by simulating two cylindrical phantom composed of acrylic and aluminum. For exact calibration, low-energy data from the whole FOV was used. For high-energy data, we made a filter, which covers the whole FOV and used it instead of the multi-slit filter. The thickness of the filter is the same as the multi-slit filter. Scatter is negligible in this work because of fan-beam geometry.

TABLE I
SCAN SYSTEM PARAMETERS

S. N.	Parameter	Value
1	Distance of source to detector	220 mm
2	Distance of source to isocenter	150 mm
3	Distance of the filter to source	50 mm
4	Field of view (FOV)	54.9 mm
5	Detector array size (pixels × pixels)	512 x 1
6	Image array size (pixels × pixels)	512 × 512
7	Number of views per rotation	720

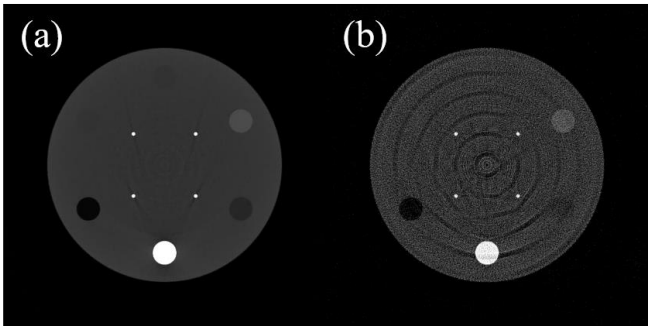


Fig. 3. (a) Low-energy CT image reconstructed by solving Eq. 7, and (b) high-image CT image reconstructed by ART algorithm without regularization. Display window: [0, 0.20] for the low-energy CT image and [0, 0.15] for the high-energy CT image.

III. RESULTS AND DISCUSSION

Figure 3 shows the reconstructed images of the simulated phantom. Figure 3 (a) is the reconstructed image from the low-energy data, and (b) is the reconstructed image from the high-energy data or filtered data. We used Eq. (7) to reconstruct the low-energy image, but for the high-energy image only ART algorithm was used without regularization. The low-energy image was successfully reconstructed and all the contrast structures in the simulated phantom are clearly

shown. However, several unwanted ring artifacts and salt-and-pepper noise are shown in the high-energy image because of insufficient and non-uniform sampling.

Figure 4 shows the high-energy reconstructed images from the conventional GMI regularized reconstruction and the proposed reconstruction method. Both reconstruction methods used the same value for β, 0.2. The GMI regularization almost removed the ring artifacts, but small patterns and noise still remain. On the other hand, the reconstruction method with the NGMI regularization substantially reduced the patterns and noise.

We calculated CNR for the simulated four contrast structures: air, aluminum, polyethylene, and Teflon. The region in the yellow circle stands for a ROI, and the region in the blue dotted circles stands for a background region of the ROI. The results for all contrast structures are summarized in table. 2. Material decomposition was also performed with the reconstructed low- and high-energy images. The coefficients for each material, which are prepared separately by the calibration phantom, was applied. The result images are shown in Fig. 5. The material decomposition was successful in both materials.

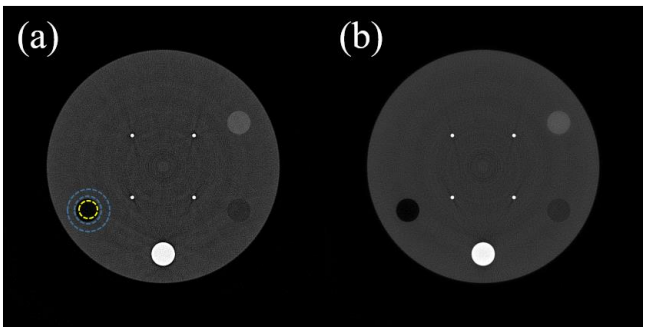


Fig. 4. (a) High-energy CT image reconstructed with the GMI regularization and (b) high-image CT image reconstructed with the NGMI regularization. Display window: [0, 0.15]

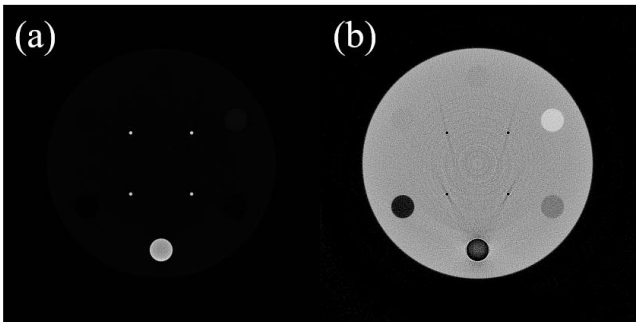


Fig. 5. Material decomposed images: (a) aluminum-based image and (b) acrylic-based image. Display window: [0, 1.5]

IV. Conclusion

In this study, we proposed a new multi-slit filter and the NGMI regularization for dual-energy CT imaging, and demonstrated its feasibility in the Monte Carlo simulation environment. The low-energy CT image was successfully reconstructed from the unfiltered data. Most image artifacts in the high-energy CT image were suppressed by use of the proposed NGMI regularization. Further experimental studies with the proposed method will be investigated in the future study.

TABLE 2

CNRs for Contrast Structures in the Simulated Phantom

Method	Air	Aluminum	Polyethylene	Teflon
GMI	2.14	5.53	0.40	1.20
NGMI	3.47	8.71	0.69	1.87

References

[1] M. Patino, A. Prochowski, M. D. Agrawal, F. J. Simeone, R. Gupta, P. F. Hahn, and D. V. Sahani, "Material separation using dual-energy CT: Current and emerging applications," *Radiographics*, vol. 36, no. 4, pp. 1087–1105, 2016.

[2] K. Kim, J. C. Ye, W. Worstell, J. Ouyang, Y. Rakvongthai, G. El Fakhri, and Q. Li, "Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty," *IEEE Trans. Med. Imaging*, vol. 34, no. 3, pp. 748–760, 2015.

[3] L. Yu, A. N. Primak, X. Liu, and C. H. McCollough, "Image quality optimization and evaluation of linearly mixed images in dual-source, dual-energy CT," *Med. Phys.*, vol. 36, no. 3, pp. 1019–1024, 2009.

[4] J. Doerner, M. Hauger, T. Hickethier, J. Byrtus, C. Wybranski, N. G. Hokamp, D. Maintz, and S. Haneder, "Image quality evaluation of dual-layer spectral detector CT of the chest and comparison with conventional CT imaging," *Eur. J. Radiol.*, vol. 93, no. May, pp. 52–58, 2017.

[5] K. Müller, S. Datta, M. Ahmad, J. H. Choi, T. Moore, L. Pung, C. Niebler, G. E. Gold, A. Maier, and R. Fahrig, "Interventional dual-energy imaging - Feasibility of rapid kV-switching on a C-arm CT system," *Med. Phys.*, vol. 43, no. 10, pp. 5537–5546, 2016.

[6] D. Lee, J. Lee, H. Kim, T. Lee, J. Soh, M. Park, C. Kim, Y. J. Lee, S. Cho, D. Lee, H. Kim, T. Lee, J. Soh, M. Park, and C. Kim, "A Feasibility Study of Low-Dose Single-Scan Dual-Energy Cone-Beam CT in Many-View Under-Sampling Framework," *IEEE Trans. Med. Imaging*, vol. 36, no. 12, 2017.

[7] I. P. Almeida, L. E. J. R. Schyns, M. C. Öllers, W. van Elmpt, K. Parodi, G. Landry, and F. Verhaegen, "Dual-energy CT quantitative imaging: a comparison study between twin-beam and dual-source CT scanners," *Med. Phys.*, vol. 44, no. 1, pp. 171–179, 2017.

[8] T. Lee, C. Lee, J. Baek, and S. Cho, "Moving Beam-Blocker-Based Low-Dose Cone-Beam CT," *IEEE Trans. Nucl. Sci.*, 2016.

[9] M. Petrongolo and L. Zhu, "Single-Scan Dual-Energy CT Using Primary Modulation," *IEEE Trans. Med. Imaging*, vol. 37, no. 8, pp. 1799–1808, Aug. 2018.

[10] S. Abbas, T. Lee, S. Shin, R. Lee, and S. Cho, "Effects of sparse sampling schemes on image quality in low-dose CT," *Med. Phys.*, 2013.

[11] T. Wang and L. Zhu, "Dual energy CT with one full scan and a second sparse-view scan using structure preserving iterative reconstruction (SPIR)," *Phys. Med. Biol.*, vol. 61, no. 18, pp. 6684–6706, Sep. 2016.

[12] M. Petrongolo and L. Zhu, "Single-Scan Dual-Energy CT Using Primary Modulation," *IEEE Trans. Med. Imaging*, vol. 37, no. 8, pp. 1799–1808, 2018.

[13] T. Wang and L. Zhu, "Dual energy CT with one full scan and a second sparse-view scan using structure preserving iterative reconstruction (SPIR)," *Phys. Med. Biol.*, vol. 61, no. 18, pp. 6684–6706, 2016.

[14] J. Min, R. Pua, C. Kim, M. Park, J. Lee, S. J. Ye, and S. Cho, "A weighted rebinned backprojection-filtration algorithm from partially beam-blocked data for a single-scan cone-beam CT with hybrid type scatter correction," *Med. Phys.*, vol. 46, no. 3, pp. 1182–1197, 2019.

[15] R. Gordon, R. Bender, and G. T. Herman, "Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theor. Biol.*, vol. 29, no. 3, pp. 471–481, 1970.

[16] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, 2008.

[17] P. Stenner, T. Berkus, and M. Kachelriess, "Empirical dual energy calibration (EDEC) for cone-beam computed tomography," *Med. Phys.*, vol. 34, no. 9, pp. 3630–3641, 2007.

Prior-knowledge based Linogram Reconstruction Neural Network for X-ray Computed Tomography

Tao Zhang, Hewei Gao, Yuxiang Xing, Li Zhang, and Zhiqiang Chen*

Abstract—In this work, we propose a general Linogram reconstruction neural network (Linogram-Net) for X-ray Computed Tomography (CT). It's an end-to-end neural network, in which existing physical principles about Linogram reconstruction theory are embedded as prior knowledge. To validate the effectiveness of our proposed Linogram-Net, we conduct some experiments on a special CT imaging configuration with Linogram sampling. Preliminary results show that, as a learning-based approach, the weighting layer of projection domain can be trained to improve reconstructed CT image from a limited-angle scan. To our best knowledge, it is the first attempt to investigate the feasibility of implementing Linogram reconstruction into an end-to-end trainable neural network and explore the potential of such a network to learn the parameters of Linogram reconstruction. Overall, our proposed network is lightweight and flexible, which can be easily extended to various reconstruction tasks such as filter design, sparse-view imaging and low-dose reconstruction.

Index Terms—Linogram Reconstruction, Prior Knowledge, Neural Network, Computed Tomography

I. Introduction

DEEP Learning methods have generated overwhelming interests in computer tomography (CT) imaging over the past several years. The parameters of neural network can be trained according to the target task, showing great potential in ill-posed inverse problems such as limited-angle [1], low-dose [2] and sparse-view CT reconstruction. However, in general, it's difficult to figure out the contribution of each layer of network to CT imaging. As a result, the design of network becomes a big challenge. To solve this problem, the existing knowledge of physical principles were introduced into the neural network structure [3], [4], [5].

In this work, we propose an end-to-end Linogram reconstruction neural network (Linogram-Net) for X-ray CT. Linogram-Net transfers the power of neural network to image reconstruction while making use of prior knowledge on direct Fourier reconstruction. It can achieve accurate reconstruction the same as a traditional analytic reconstruction method, when the projection data is sufficient. In addition, as a learning-based approach, the weighting layer of projection domain can be trained to improve reconstruction when missing data occurs such as the limited-angle reconstruction, through minimizing the loss function of image domain.

To our best knowledge, it is the first attempt to investigate the feasibility of implementing Linogram reconstruction into an end-to-end trainable neural network. Overall, our proposed network is lightweight and flexible, which can be easily extended to various reconstruction tasks such as filter design, sparse-view imaging and low-dose reconstruction.

This work was supported in part by the National Natural Science Foundation of China (Grant No. 11525521) and by the National Key R&D Program of China (Grand No. 2016YFB1101203)

T. Zhang, H. W. Gao, Y. X. Xing, L. Zhang and Z. Q. Chen are with Department of Engineering Physics, Tsinghua University, Beijing, 100084, China, and also with Key Laboratory of Particle Radiation Imaging (Tsinghua University), Ministry of Education, Beijing, 100084, China

*Author to whom the correspondence should be addressed. Email: czq@tsinghua.edu.cn

II. Linogram Reconstruction Theory

The Linogram sampling grids in Fourier space are illustrated in Fig. 1. As a direct Fourier method, Linogram reconstruction algorithm can be used in the situation where sampling points of Fourier space are on concentric squares [6].

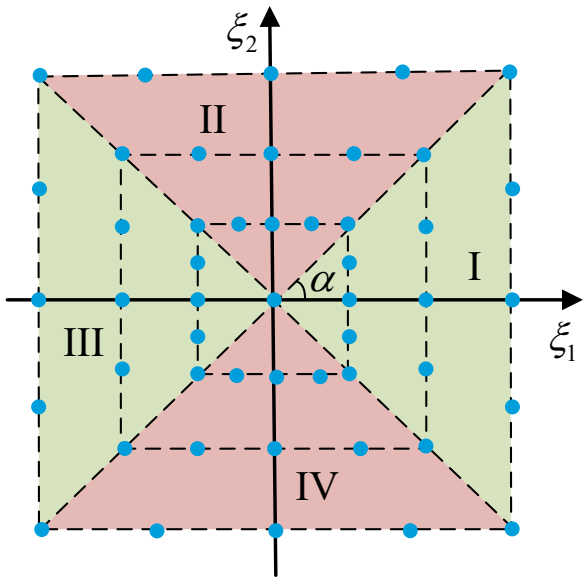


Fig. 1. Linogram sampling grids in Fourier space.

To satisfy such sampling condition, the Fourier transform of object $f(x, y)$ should be like:

$$\hat{f}_1(\xi_1, \lambda\xi_1) = \int\int \mathrm{d}x\mathrm{d}y f(x,y) e^{-2\pi j\xi_1(x+\lambda y)} \tag{1a}$$

$$\hat{f}_2(\eta\xi_2, \xi_2) = \int\int \mathrm{d}x\mathrm{d}y f(x,y) e^{-2\pi j\xi_2(\eta x+y)} \tag{1b}$$

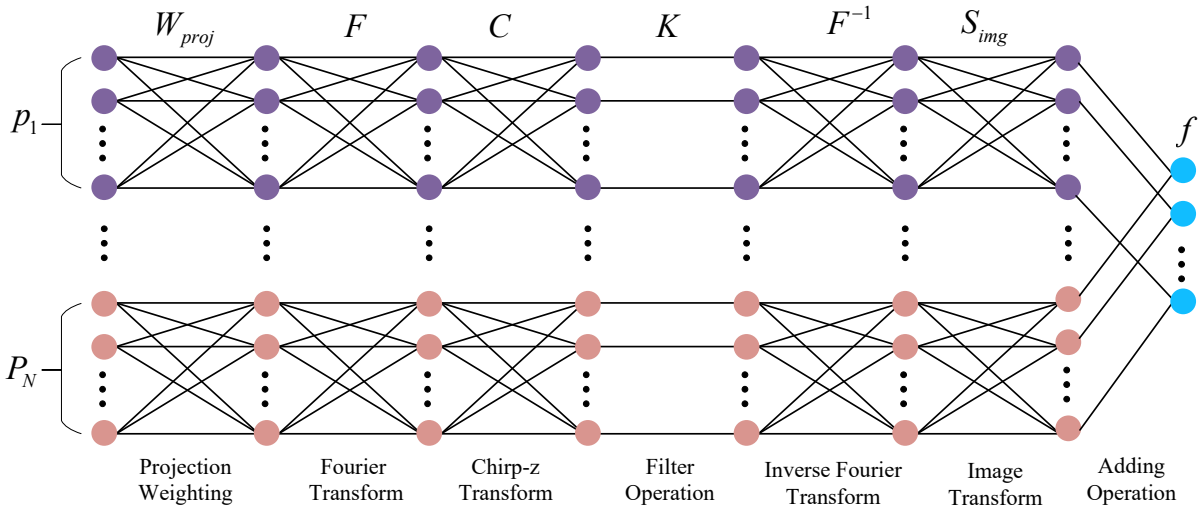


Fig. 2. Architecture of Linogram-Net. It consists of multiple reconstruction paths (each corresponds to a segment of Linogram-type CT scan), whose results will be combined through adding operation to achieve a final reconstruction.

here, $\hat{f}_1$ corresponds to the sampling points in region I and III, while $\hat{f}_2$ corresponds to the regions II and IV. Thus we have $\lambda <= \tan(\alpha)$ and $\eta <= \cot(\alpha)$. And $f(x, y)$ can be obtained, ideally, by the sum of inverse Fourier transform of $\hat{f}_1$ and $\hat{f}_2$, in which chirp-z transform can be applied to avoid interpolations in Fourier space.

Generally, the acquired data from circular or helical CT scans can not directly meet the Linogram sampling condition, thus rebinning operations are needed in projection domain. It is known that in practical applications multiple segments of Linogram-type CT scan are usually conducted in order to obtain the sufficient sampling points to fill the Fourier space. As a result, data redundancy problem could occur, in which there are overlapping regions between $\hat{f}_1$ and $\hat{f}_2$. Thus projection weighting strategy should be introduced to deal with redundancy problem. In order to introduce a unified solution for the reconstruction problems above, in section II, we propose a general prior-knowledge based Linogram reconstruction neural network and explore the potential of such a network to learn the parameters of reconstruction.

III. Linogram-Net

As shown in Fig. 2, the inputs of Linogram-Net are projections from multiple segments of Linogram-type CT scan. So our Linogram-Net consists of multiple reconstruction paths, each of them corresponds to one scan segment. And the output f is the final image combined from results of the multiple reconstruction paths, which can be expressed as:

$$f = \sum_{n=1}^{N} S_{img_n} F_n^{-1} K_n \bullet C_n F_n W_{proj_n} p_n \qquad (2)$$

here n is the index of Linogram-type CT scan segment, N is the total number of scan segments and p_n is the projection data acquired from scan segment n; C_n, F_n and F_n^{-1} represent chirp-z transform matrix, Fourier transform matrix and inverse Fourier transform matrix respectively. K_n is the filter kernel and $\bullet$ represents element-wise multiplication. W_{proj_n} is the weight in projection domain which is designed for rebinning operation, redundancy weighting or other pre-processing of projection data. S_{img_n} is the transform matrix in image domain, responsible for post-processing of reconstructed image from p_n.

Under accurate reconstruction requirements, when projection rays cover at least 180 degrees for each and every points of the scanned points, the parameters of Linogram-Net layers can be initialized by the known solution of analytic Linogram reconstruction theory and one can directly obtain accurate reconstruction using Linogram-Net. When the projection data is insufficient, violating the conditions of an accurate reconstruction, the parameters of some layers can be adjusted in data-driven fashion to improve the quality of reconstructed image.

Taking the limited-angle scan case as example, the projection weighting layer of Linogram-Net can be trained through minimizing the loss function in image domain while the parameters of other layers remain unchanged in the whole training stage. With appropriate optimization strategy selected, the parameters of projection weighting layer in projection domain can converge to a optimized solution. As a result, the output image of trained Linogram-Net is much more closer to the ground truth, compared with that of non-trained Linogram-Net in which projection weighting layer is initialized by empirical parameters.

IV. Evaluations

A. Experimental Setups

We evaluate the proposed Linogram-Net in Symmetric-Geometry Computed Tomography (SGCT). With linearly distributed sources and detectors, SGCT is a special case of straight-line CT [7], [8], whose data sampling fulfills

a Linogram pattern essentially. A dual-SGCT scan mode is applied here, where the source and detector arrays of SGCT scan segment I are particular to that of SGCT scan segment II. In this work, we simulate a limited-angle case where only some points have 180 degrees projection data while the angular coverage of others is less than 180 degrees.

The data used in our experiments are 80 abdominal 3D CT scans collected by National Institutes of Health Clinical Center [9], [10], [11]. This dataset provides 3D CT volumes reconstructed from multi-detector Computed Tomography. To obtain projection of a dual-SGCT scan, we forward-project them according to the SGCT imaging geometry.

In this study, Linogram-Net is implemented in Tensorflow platform. And the loss value of Linogram-Net is computed by the sum of squares of difference between the output image of Linogram-Net and corresponding ground truth.

B. Projection Weight Learning in the Limited-angle Case

In this section, we explore the potential of our network to learn the projection weight for improving Limited-angle reconstruction. In the training stage of Linogram-Net, the limited-angle projection data are fed to the network, and the parameters of projection weighting layer are optimized to minimize image domain loss while that of other layers are fixed. Here, we adapt the 5-folds cross validation strategy. The train and validation loss curves of fold-4 and fold-5 cross validation are plotted in Fig. 3, which reveals that projection weight converges to a optimized solution through the training stage. What's more, an example

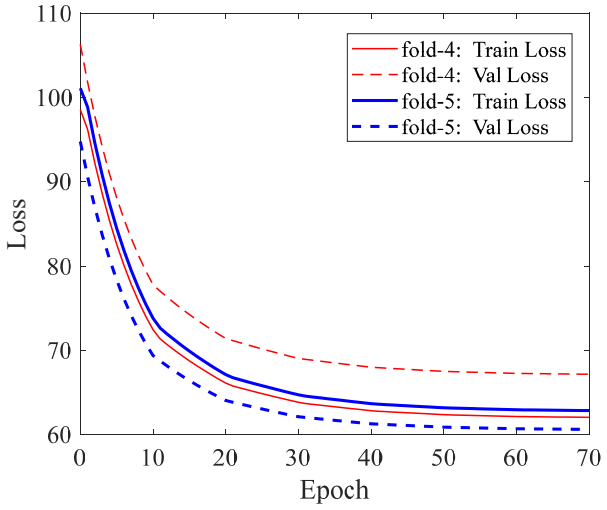


Fig. 3. Train and validation Loss curves of fold-4 cross validation experiment and fold-5 cross validation experiment.

of cross validation is shown in Fig. 4. It demonstrates that the learned projection weight can increase the image quality and suppress the limited-angel artifacts very well, revealing that Linogram-Net has the power to adjust Linogram reconstruction parameters in projection domain through minimizing the loss function in image domain.

V. Conclusions

In this work, we proposed a prior-knowledge based end-to-end trainable neural network for Linogram reconstruction in X-ray CT. This general framework transfers the power of neural network to reconstruction while making use of existing known knowledge, in which the weighting layer of projection domain can be trained to improve the image quality when missing data occurs such as limited-angle reconstruction. Overall, our proposed network is lightweight and flexible, which can be easily extended to various reconstruction tasks such as filter design, sparse-view imaging and low-dose reconstruction.

References

[1] T. A. Bubba, G. Kutyniok, M. Lassas, M. März, W. Samek, S. Siltanen, and V. Srinivasan, "Learning the invisible: A hybrid deep learning-shearlet framework for limited angle computed tomography," Inverse Problems, vol. 35, no. 6, p. 064002, 2019.

[2] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction," Medical physics, vol. 44, no. 10, pp. e360–e375, 2017.

[3] C. Syben, M. Michen, B. Stimpel, S. Seitz, S. Ploner, and A. Maier, "PYRO-NN: Python Reconstruction Operators in Neural Networks," Medical Physics, vol. 46, no. 11, pp. 5110–5115, 2019.

[4] T. Würfl, M. Hoffmann, V. Christlein, K. Breininger, Y. Huang, M. Unberath, and A. K. Maier, "Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems," IEEE transactions on medical imaging, vol. 37, no. 6, pp. 1454–1463, 2018.

[5] A. K. Maier, C. Syben, B. Stimpel, T. Würfl, M. Hoffmann, F. Schebesch, W. Fu, L. Mill, L. Kling, and S. Christiansen, "Learning with known operators reduces maximum error bounds," Nature machine intelligence, vol. 1, no. 8, pp. 373–380, 2019.

[6] H. Gao, L. Zhang, Y. Xing, Z. Chen, and J. Cheng, "An improved form of linogram algorithm for image reconstruction," IEEE Transactions on Nuclear Science, vol. 55, no. 1, pp. 552–559, 2008.

[7] T. Zhang, Y. Xing, H. Gao, Z. Chen, and L. Zhang, "Inverse-geometry ct with linearly distributed source and detector: stationary configuration and direct filtered-backprojection reconstruction," in 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, vol. 11072. International Society for Optics and Photonics, 2019, p. 110722T.

[8] T. Zhang, Y. Xing, L. Zhang, X. Jin, H. Gao, and Z. Chen, "Stationary computed tomography with source and detector in linear symmetric-geometry: Direct filtered-backprojection reconstruction," Medical Physics, 2020.

[9] H. Roth, A. Farag, E. Turkbey, L. Lu, J. Liu, and R. Summers, "Data from tcia pancreas-ct," The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU, 2016.

[10] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in International conference on medical image computing and computer-assisted intervention. Springer, 2015, pp. 556–564.

[11] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle et al., "The cancer imaging archive (tcia): maintaining and operating a public information repository," Journal of digital imaging, vol. 26, no. 6, pp. 1045–1057, 2013.

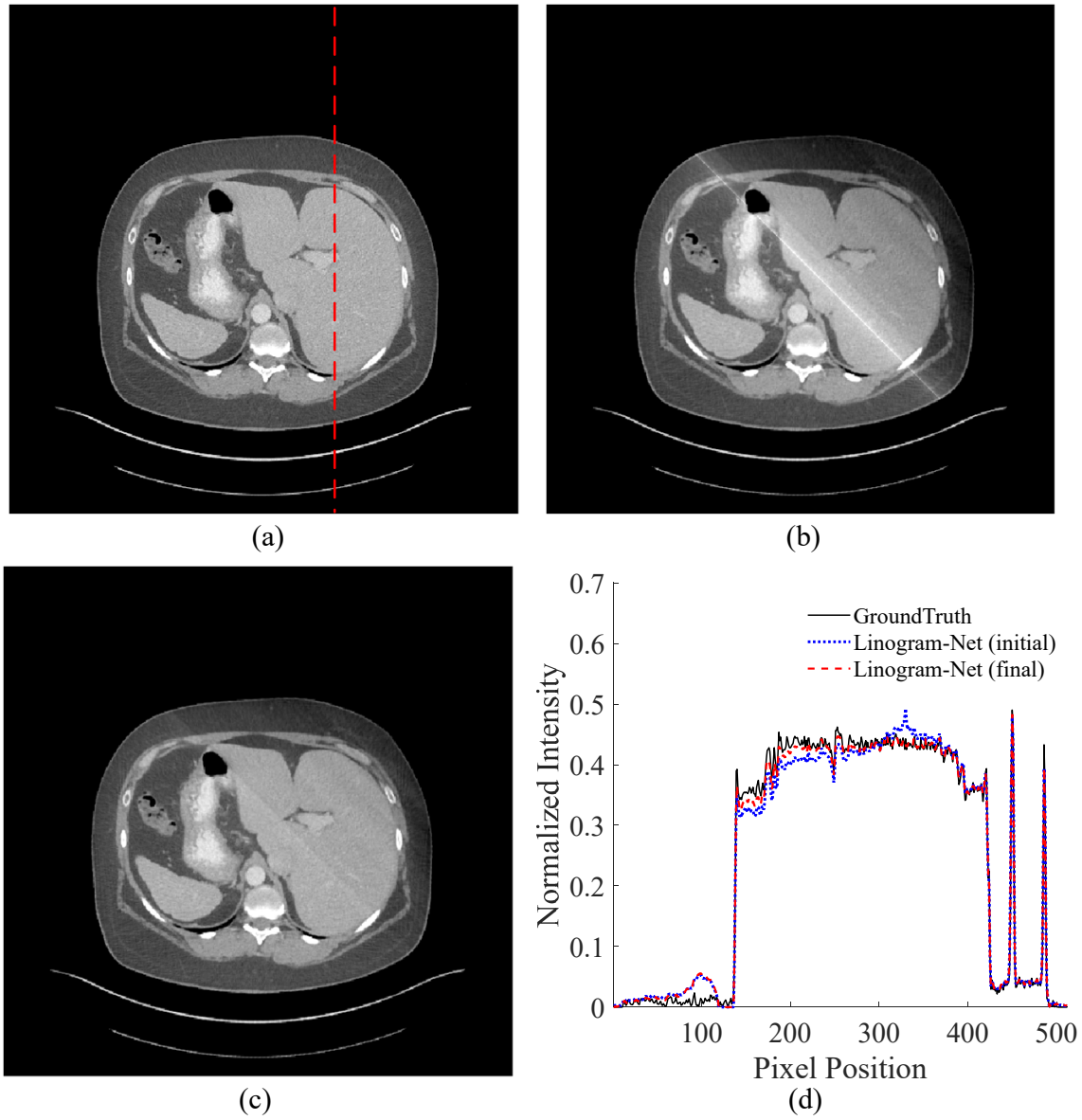


Fig. 4. Reconstructed images in limited-angle case, display window: [0.3, 0.5]. (a) Ground Truth; (b) output by the initial Linogram-Net, in which the parameters of projection weighting layer are initialized with "Parker-Weights"-like strategy; (c) output by the trained Linogram-Net; and (d) 1D profiles of (a), (b) and (c) along the line indicated in (a).

Segmented static CBCT as initialization of fully dynamic tomography reconstruction of foot and ankle

Nargiza Djurabekova[1], Andrew Goldberg[1,2], David Hawkes[1], Guy Long[3], Luke Shepherd[1] and Marta M. Betcke[1]

***Abstract*—In our approach to studying the dynamics of the foot and ankle, we acquire two consecutive weigh-bearing conebeam CT scans. The first scan is a regular, static, full data scan, while the other scan is dynamic. This means that we can have one perfect reconstruction and several imperfect reconstructions of binned dynamic data.**

We use the principal component analysis on individually segmented bones to replace the imperfect reconstructions from the binned dynamic data with their perfect segmentation from the static scan and demonstrate the initial results on three hindfoot bones.

***Index Terms*—Dynamic CT, cone Beam CT, bones, PCA, rigid motion**

I. Introduction

THE precise motion of the foot and ankle is difficult to study, due to the sheer number of bones in the structure as well as them obscuring one another in both tactile inspection as well as when studying radiographs (Fig. 1). Hence, at present, for more complex conditions a static weight-bearing cone beam CT scan is acquired to aid the diagnosis. By static scan we mean the process of collecting X-ray projections while the scanned foot and ankle are immobile. With a static weight-bearing cone beam CT scan, we get access to information of the exact shape of each bone as well as its relative position and orientation with respect to other bones. Our aim is to study the dynamics of the foot and ankle structure with minimal dose hence a protocol providing both the static and dynamic scan at minimal dose is prudent. Furthermore, we will demonstrate that knowing the shape and original placement of bones can benefit the reconstruction of the dynamic data. This is due to the assumed rigidity of bones which in turn means that their motion is restricted to exclusively rotations and translations of the individual bones

Knowing how the foot and ankle move can have direct impact on orthopaedic diagnoses as well treatment. Joint replacement could be developed for the 70000 people in the UK suffering from hindfoot arthritis (data from Clinical Practice Research Datalink or CPRD)[1] once there is an improved biomechanical model of the motion. Previous attempts without the knowledge of the exact range of motion led to half the trial surgeries being unsuccessful [2]. Just computing the motion field simultaneously with the reconstruction, has led to some promising results [3][4]. However reconstructing takes time and the resulting bones and spaces between them are imprecise, so using some form of a shape prior could be beneficial.

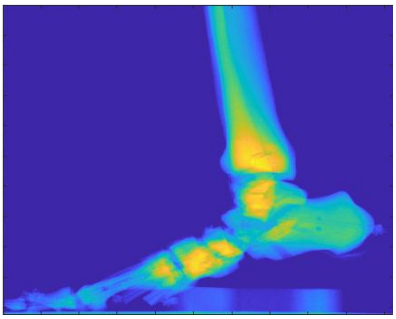


Fig. 1: Sample X-ray projection to demonstrate how some bones obscure others, resolution: 1024×1024.

In this paper we outline how static, load-bearing CBCT scans can be used to initialize a reconstruction of fully dynamic data, i.e. a scan being taken while the foot is moving inside a CBCT machine. The code can be found on github [1].

A. Principal component analysis

Karl Pearson is credited with introducing the concept of principal component analysis (PCA) as early as 1901 [5], although the same idea was rediscovered in many different fields under a variety of different names. The main idea of PCA is to find orthogonal principal components (PCs) that describe the directions of the highest variance within the data (fig. 2).

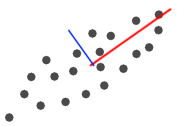


Fig. 2: Example of a PCA applied to a 2D data point cloud. The first principle component (red) represents the (scaled) direction of the highest variance in the data, while the second PC (blue) represents the direction with less variance.

The method we use for computing PCs is called the covariance method (see e.g. [6]), which we briefly describe below.

[1] University College London
[2] The Wellington Hospital London
[3] CurveBeam

[1] https://lukemshepherd.github.io/vox/

{Given a set X of n points in $\mathbb{R}^d$, we compute their mean $\bar{X} \in \mathbb{R}^d$ componentwise

$$\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}, \quad i = 1, \ldots, d.$$

Subtracting the mean, we obtain data centered at 0

$$X'_{ij} = X_{ij} - \bar{X}_i$$

and the corresponding covariance matrix, $C \in \mathcal{R}^{d\times d}$

$$C = Cov(X') = \frac{1}{n-1} X' X'^T.$$

The normalised eigenvectors E_{vec} of the covariance matrix are the principal components directions and their associated eigenvalues correspond to their magnitudes.

After segmenting the bones, each individual bone is represented with a point cloud of the indices of the grid points inside its support and its PCA is computed as described above. In this paper we assume that such segmentation can be obtained for both scenarios: the static scan reconstruction as well as the sparse dynamic data reconstruction. Instead of segmenting, we simulate both scenarios on an a-priori segmented foot phantom using the linearity of the forward operator to compute the projections of the individual bones and superpose them to obtain the projection of the entire structure. PCA of bones is already a widely used tool in clinical analysis [2], making the approach outlined in the paper fit well into clinical practice.

B. Quaternions

Quaternions are a mathematical concept that extends the complex numbers to the 3D space. They are often written in the form

$$q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k},$$

where **i, j** and **k** are orthogonal unit vectors that denote the 3D coordinate system along the x, y and z axis respectively, and fulfilling the mathematical axiom

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1.$$

The concept was introduced already in 1843 by a mathematician by the name of Willian Rowan Hamilton [7], hence quaternion algebra being sometimes refered to as Hamiltonian. Unit quaternions, in particular, have become a useful tool used in all things related to rotation and orientation in 3D space. Some of the benefits of using quaternions over a regular rotation matrix are their efficient storage and their ability to prevent gimbal lock. Not only that but, unlike the usual Euler rotation, quaternions always find the shortest way to rotate from one vector to the next. This has made them extremely useful in finding mid-rotation positions for animating objects between keyframes.

[2]personal communication with Dr Francois Lintz, orthpaedic surgeon, Clinique de L'Union.

A more intuitive, for our purposes, formulation for qs can be written as follows. Let v and w be two vectors between which we are trying to find the exact rotation. Then

$$q = \cos\frac{\theta}{2} + \bar{u}\sin\frac{\theta}{2},$$

where $\bar{u}$ is the unit vector resulting from $v \times w$ and θ is the angle between them. In other words, we can rotate v about $\bar{u}$ by some non-zero angle θ to get w. The reason for dividing the θ by 2 is due to how quaternions are applied to perform rotation:

$$v = qwq^{-1},$$

where q^{-1} is the conjugate, defined as

$$q^{-1} = \cos\frac{\theta}{2} - \bar{u}\sin\frac{\theta}{2}.$$

II. Methodology

A. Set up

A combination of Blender [8] and MATLAB [9] was used to create two arrays of size $256 \times 256 \times 256$ containing 3 solid bones (tibia, talus and calcaneus) stacked on top of each other in two different positions, representing two frames of stop motion. We refer to these two perfectly segmented sets of bones as phantoms, to differentiate them from reconstructions we will refer to later.

Then the ASTRA toolbox [10] was used on the individual bones of the second frame to simulate projections chosen based on stratified sampling. This sampling method mimics binning projections based on an idealised quasi-periodic motion of an object (see [4]).

The reason for simulating projections on individual bones is to avoid the need to segment the poor reconstruction of all three bones in this proof-of-concept. As long as the bones are correctly located in space, the radiographs of each bone from the same angle can just be summed, due to the linearity of the X-ray operator A (*). If f_b is the vectorised projection for each bone, x_b represents the individual bone and A is the system matrix, we have that

$$f = \sum_{b=1}^{B} f_b + noise = A\sum_{b=1}^{B} x_b + noise \overset{*}{=} \sum_{b=1}^{B} Ax_b + noise,$$

where B is the total number of bones.

Then, to reconstruct each bone, we use FISTA [11] to solve the following variational minimization problem

$$\hat{x} = \arg\min_x ||Ax - f|| + \alpha TV(x),$$

where $x = \sum_{b=1}^{B} x_b$, TV is the regularization term that uses total variation and α is a regularization parameter. TV was chosen due to its simplicity as well as region constancy assumption, as our simulated bones are assumed to be constant inside.

B. Step-by-step procedure

In this section we describe the step by step guide of taking a fully segmented bone from one frame and placing it into frame

two, in place of a quick preliminary dynamic reconstruction. We illustrate the method first on just one bone (tibia) in two positions, frame 1 and frame 2. In principle, all we want to accomplish is to find the translation vector t and the rotation matrix R between the two frames and the apply those to frame 1 to move the bone rigidly into frame 2, however in practice we take a slightly different approach. In the following we assume each bone to be a point cloud of grid points laying inside of the bone which are described by their indices aka Cartesian coordinates.

1) Let the matrix of indices of bone in frame 1 be called **A** and **B** for frame 2. The indices can be found only with the right threshold of which values belong to a bone and which do not. In the phantom case (frame 1) where the bones are homogeneous, this is trivial. In the case of the reconstruction, the threshold is chosen to be 0.5 of the phantom threshold value, to avoid including unnecessary noise and artefacts. Then the mean of **A** is μ_A and the mean of **B** is μ_B.
2) Using the covariance method described in section I-A and the means μ_A and μ_B, we center **A** and **B** at 0 and and compute their PCs (see fig. 3).
3) Treating the bones respective PCs as local axis, we use quaternions from sec. I-B to align the PCs of the phantom to the PCs of the reconstruction. This is done in stages.
   First, we align PC1s i.e. the first principal vectors. For this, we compute the angle and the normalised cross product between the them. This gives us the quaternion for the first step in the rotation. We apply this rotation to all the PCs (see fig. 4(b)).
   Now with all the PCs in new positions, we find another quaternion that would align PC2s and apply it to all the PCs (fig. 4(c)). Note that aligning along PC2 does not affect the previous alignment because after the first alignment both PC2s are in a plane orthogonal to aligned PC1s. Finally, all the PCs are aligned also using the final quaternion for PC3s (fig. 4(d)).
4) With the three quaternions acquired in the previous step, we can now rotate the phantom bone in position 1 into the position of the imperfect reconstruction of the bone in frame 2. Lastly, we move the now aligned bones so that their center is the original center of the imperfect reconstruction, i.e. we add μ_B to all the indices of the bone (see fig. 5).

III. Results and Discussion

Assuming we have 3 moving bones (tibia, talus and calcaneus), we choose two stop-motion frames to work with. The first frame we set to be the initial static position, while second frame gets subsampled to mimic the binning of dynamic data. The proposed two step reconstruction algorithm is performed in MATLAB R2019a on arrays of size $256 \times 256 \times 256$. The first step involves reconstructing individual bones within the array using FISTA and 3 projections sampled around the perimeter using stratified random sampling. The second step uses the PC method described in section II to align each of the bones from the static, initial frame of motion to the poorly reconstructed frame. In figure 6, we demonstrate the difference in PCs between the rotated and translated phantom and the reconstruction obtained with FISTA as well as the subsequent thresholding with a 0.5 threshold. The difference is quite small, however, it may cause overlap between bones in places where they are meant to only be touching.

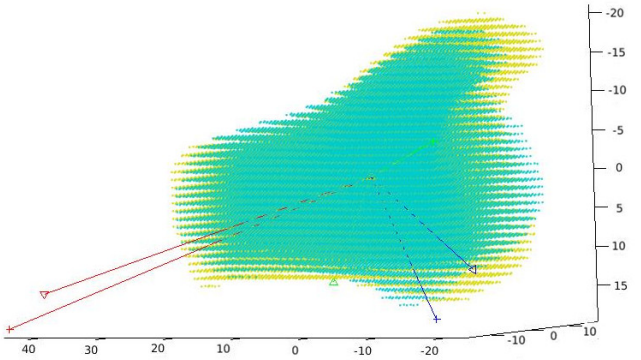


Fig. 3: Finding the PCs of the centred bone point clouds **A** and **B**. Here the PCs are scaled using the corresponding eigenvalues. The PCs with endpoint annotated as $+$s correspond to **A** and those with triangles to **B**.

With this in mind, we apply the proposed method to individual bones and add them back into a single array. The results are demonstrated in figure 7. We can see that the reconstruction using PCs does not exactly match the rotated and translated phantom. The main reason of the discrepancy is related to the density of the point cloud representing the individual bones. As we rotate points in space, they are likely to not land into any precise pixel, meaning that their indices become floats rather than integers. To remedy this, the indices have to be rounded, which can create holes in an otherwise solid rotated object. To reduce the number of these holes, we interpolate between indices of the original bones, effectively creating a denser point cloud. Practically, this is achieved by increasing the size of the bones by 2 using linear interpolation and then respectively dividing the indices by 2. That way, we essentially double the bone resolution in each space dimension which suffices to fill out most of the holes that occurred from rounding floating point numbers. We expect that more sophisticated morphological operations could effectively eliminate this problem entirely. However, even this simple fix yields very promising results, with the PC aided reconstruction matching the ideal reconstruction (rotated phantom) quite well.

Acknowledgment

The authors would like to thank NVIDIA Corporation for the GeForce Titan Xp GPU used to create the reconstructions presented here. For the original segmented bones we used as the phantom, we thank Dr Francois Lintz from Clinique de L'Union and PRof Sorin Siegler from Drexel University.

References

[1] "Clinical practice research datalink - uk data driving real-world evidence," https://www.cprd.com/.

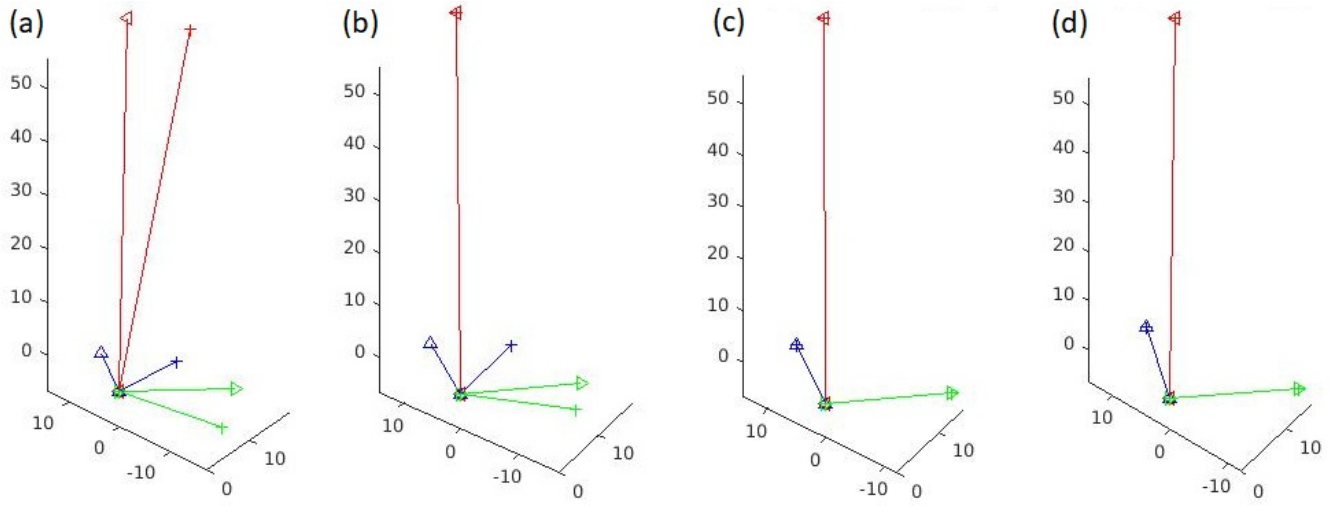


Fig. 4: Alignment of PCs. PCs with + refer to frame 1 (or **A**) and with triangles to frame 2 (or **B**).(a) - PCs of the two frames before any alignment. (b) - Aligning all PCs based on PC1s. Followed by aligning all PCs based on PC2s in (c) and finally based on PC3s in (d).

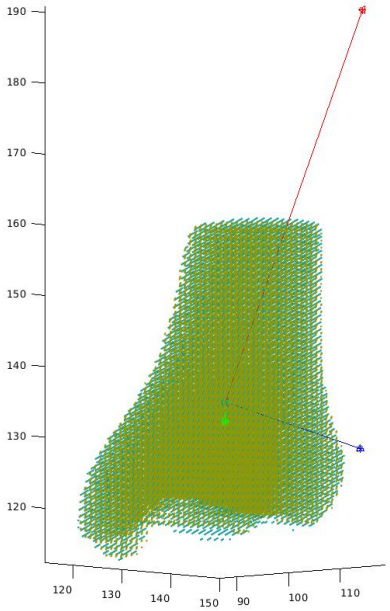


Fig. 5: Final alignment and centering of aligned **A** and **B** at μ_B via adding μ_B directly to **B** and **A** instead of centering **A** at μ_A first and then adding the translation vector $t = \mu_B - \mu_A$ to place it over **B**.

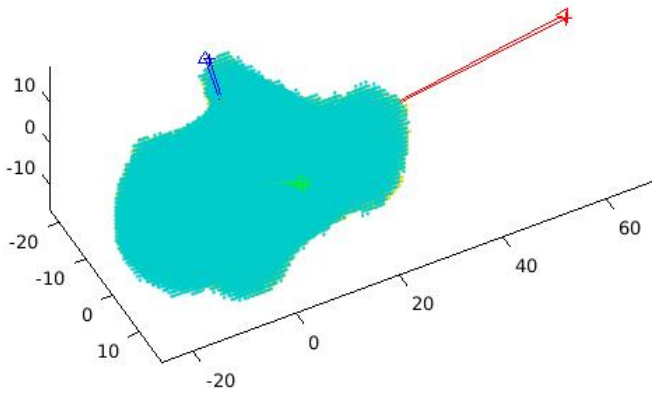


Fig. 6: The difference between the PCs of reference frame 2 (obtained via the exact rotation and translation of the phantom) and the PCs of the imperfect FISTA reconstruction.

Fig. 7: A single slice that demonstrates the result of the proposed method on the three hindfoot bones: the imperfect reconstruction using FISTA (left); the result of the proposed method (middle) and the reference aka phantom obtained using exact rotation sand translation (right).

[2] S. Fournol, "L'arthoplastie total sous-talienne. resultats et bilan d'une serie de 100 prothses," vol. 15, pp. 67–71, 1999.

[3] F. Lucka, N. Huynh, M. Betcke, E. Zhang, P. Beard, B. Cox, and S. Arridge, "Enhancing compressed sensing photoacoustic tomography by simultaneous motion estimation," 02 2018.

[4] N. Djurabekova, A. Goldberg Obe, A. Hauptmann, D. Hawkes, G. Long, F. Lucka, and M. Betcke, "Application of proximal alternating linearized minimization (palm) and inertial palm to dynamic 3d ct," 05 2019.

[5] K. P. F.R.S., "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901. [Online]. Available: https://doi.org/10.1080/14786440109462720

[6] L. Smith, "A tutorial on principal components analysis," 2002. [Online]. Available: http://hdl.handle.net/10523/7534

[7] S. W. R. Hamilton, "On quaternions, or a new system of imaginaeis in algebra (a letter from sir william rowan hamilton to john t. graves)," vol. XXV, pp. 489–95, 1844.

[8] "Blender - a 3d modelling and rendering package," Blender Foundation, Blender Institute, Amsterdam, 2017, http://www.blender.org.

[9] MathWorks, "Matlab 2017a," 2017, https://uk.mathworks.com/products/matlab.html?s_tid=srchtitle.

[10] W. van Aarle, W. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the astra toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25 129–25 147, Oct 2016.

[11] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009. [Online]. Available: https://doi.org/10.1137/080716542

Total Deep Variation for SparseCT Reconstruction

Erich Kobler, Baiyu Chen, Alexander Effland, Thomas Pock, Daniel K. Sodickson, and Ricardo Otazo

***Abstract*—SparseCT is a promising compressed sensing technique for low-dose examinations that undersamples projection data along the gantry angle and slice dimensions using a multislit collimator. The thereby scanned volume is typically reconstructed using a model-based iterative approach with a sparsity enforcing regularizer on image gradients. However, these simple regularizers are not able to represent the complex nature of CT images, and as a result, introduce blocky and blurring artifacts. In this work, we propose to learn a total deep variation (TDV) regularizer designed by deep learning principles to denoise SparseCT undersampled reconstructions given fully-sampled pendants. The resulting TDV regularizer is then used in the model-base iterative reconstruction and tested using in vivo abdominal patient data with retrospective 4-fold SparseCT undersampling.**

***Index Terms*—CT, compressed sensing, SparseCT, deep learning**

I. Introduction

SparseCT is a technique for CT dose reduction that implements compressed sensing with a multislit collimator (MSC) for undersampled projection data acquisition and iterative reconstruction with a sparsity enforcing regularizer [1]. The MSC is composed of several narrow slits along the slice dimension (z) and can move linearly to have a different z-undersampling pattern for each gantry angle θ, which improves incoherence for compressed sensing reconstruction [2]. Image reconstruction is performed by an iterative algorithm using a sparsity-enforcing regularizer on image gradients [3]. The incorporated regularizer is usually a very simple function, such as total variation (TV) [4] or total generalized variation (TGV) [5], which is not able to fully represent the complex nature of CT images, and might lead to reconstruction artifacts, such as stair-casing, residual streaking artifacts, or image blurring.

The new generation of image reconstruction techniques is based on deep learning, where convolutional neural networks (CNNs) trained on several datasets are applied in the reconstruction process. Deep learning has been extensively applied to the low-dose CT case for denosing a filtered back projection (FBP) reconstruction of either low mA data [6]–[8] or undersampled data [9]. Later, deep learning approaches were incorporated in model-based iterative reconstruction [10]–[13], where data consistency is integrated by penalizing the deviation to the acquired sinogram data using a well-known model of the acquisition operator. Hence, the main contribution of deep learning for model-based reconstruction is the improvement of the regularizer.

In this work, we present an extension of the total deep variation (TDV) regularizer [14] to 3D CT reconstruction. We train the regularizer based on an image-domain volumetric denoising problem and apply the learned regularizer to model-based iterative compressed sensing reconstruction. The proposed technique is qualitatively evaluated on retrospectively undersampled SparseCT in vivo abdominal data.

Erich Kobler (e-mail: erich.kobler@icg.tugraz.at), Alexander Effland and Thomas Pock are with Graz University of Technology, Graz, Austria.

Baiyu Chen and Daniel K. Sodickson are with NYU School of Medicine, New York, NY USA.

Ricardo Otazo (e-mail: otazotoj@mskcc.org) is with Memorial Sloan Kettering Cancer Center, New York, NY, USA

We acknowledge support from the European Research Council under the Horizon 2020 program, ERC starting grant HOMOVIS, No. 640156 the National Institutes of Biomedical Imaging and Bioengineering under Grant U01 EB018760.

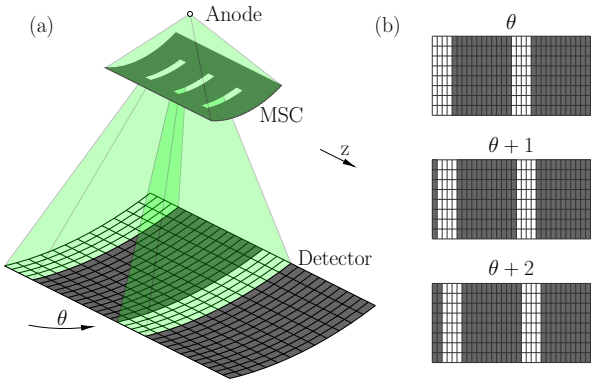


Fig. 1. (a) Schematic sketch of SparseCT undersampling, where the MSC blocks the majority of the X-ray radiation. (b) Illustration of the motion pattern of the MSC. Here, the W4S16 mask using a width of 4 detector rows and a stride of 16 rows moves with a speed $v = 1$ rows per gantry angle.

II. SparseCT undersampling

SparseCT undersamples a helical CT acquisition by using a moving MSC along the z-dimension as illustrated in Fig. 1. Linear motion of the MSC along the z-dimension while the gantry is rotating results in a combined $\theta - \mathrm{z}$ undersampling approach. Ideally, SparseCT can result into a different z-undersampling pattern for each gantry angle θ. However, in practice the focal spot has finite size and creates penumbra on both sides of each undersampled beam. As a result, each undersampled beam (the beam through one slit of the MSC) irradiates several continuous detector rows. To avoid this overlap between adjacent undersampled beams the distance between two consecutive slits of the MSC needs to be increased. On the right of Fig. 1 an example of a W4S16 sampling pattern is depicted, where 4 detector-rows are irradiated by each undersampled beam and 16 is the distance between two consecutive slits (in detector-row units).

III. Regularized SparseCT reconstruction

In this section, we elaborate on the reconstruction of the attenuation volume given the measured sinogram data of a CT

scanner using SparseCT undersampling. In detail, we consider the scanned volume $\Omega \subset \mathbb{R}^3$, which is discretized by $n = n_z \cdot n_y \cdot n_x$ voxels. The reconstructed volume is denoted by $x \in \mathbb{R}^n$. Likewise, $z \in \mathbb{R}^m$ with $m = r \cdot a \cdot d_c \cdot d_r$ represents the acquired sinogram data for r rotations of the gantry each measuring a angles, d_c detector channels, and d_r detector rows. Here, we consider a CT scanner with $a = 2304$ angles per gantry rotation and a detector using $d_c = 736$ channels and $d_r = 64$ rows. Each detector element measures a radiation intensity $\iota_i \in \mathbb{R}$, which is determined by the attenuation $\xi : \Omega \to \mathbb{R}$ in the cone Γ_i spanned by the X-ray point source and the rectangular shaped detector element, i.e.

$$\iota_i = \iota_0 \exp\left(-\int_{\Gamma_i} \xi(\vec{\omega}) \mathrm{d}\vec{\omega}\right).$$

The X-ray source intensity is denoted by $\iota_0 \in \mathbb{R}$. A transformation into the negative logarithmic domain yields

$$z_i = -\log\left(\frac{\iota_i}{\iota_0}\right) = \int_{\Gamma_i} \xi(\vec{\omega}) \mathrm{d}\vec{\omega} \approx a_i^T x,$$

where $a_i \in \mathbb{R}^n$ approximates the volumetric integral using lookup tables proposed by [15]. Stacking all post-log sinogram measurements into a vector z yields

$$z = M(Ax + n) \tag{1}$$

with $A = (a_1, \dots, a_m)^T \in \mathbb{R}^{m \times n}$. We follow [3], [16] to account for the quantum and electronic noise in the acquisition process and model the noise by a Gaussian distribution $n \sim \mathcal{N}(0, W^{-1})$, where W is diagonal with statistical weights that are determined by the flux measurements $W_{ii} = \iota_i$. To retrospectively apply SparseCT undersampling we utilize a binary diagonal mask $M \in \{0,1\}^{m \times m}$ to the fully-sampled sinogram data z, where the diagonal entries indicate if the corresponding detector element was measured. Using this model, the regularized reconstruction problem is given by

$$\min_{x \in \mathbb{R}^n} \lambda \mathcal{R}(x) + \frac{1}{2}\|M(Ax - z)\|_W^2, \tag{2}$$

where $\lambda > 0$ balances the ℓ^2-data consistency term and the regularization term $\mathcal{R} : \mathbb{R}^n \to \mathbb{R}$. Typical choices for the regularizer in medical imaging are TV [4], TGV [5], or sparsity promoting non-convex regularizers on image gradients [17], [18]. All these regularization methods solely quantify local image statistics and are prone to staircasing and blurring artifacts. To overcome these issues, we propose in this work to learn a multiscale regularizer suitable in particular for CT reconstruction.

IV. Learning deep regularizers

In [14], the TDV regularizer is rigorously analyzed and the efficient learning of the parameters for denoising is demonstrated. Moreover, the thus obtained regularizer is successfully applied to medical image reconstruction tasks *without* any further training. In what follows, we adopt this TDV regularizer to 3D CT reconstruction. The TDV regularizer is given by

$$\mathcal{R}(x, \vartheta) = \sum_{i=1}^{n} r(x, \vartheta)_i, \tag{3}$$

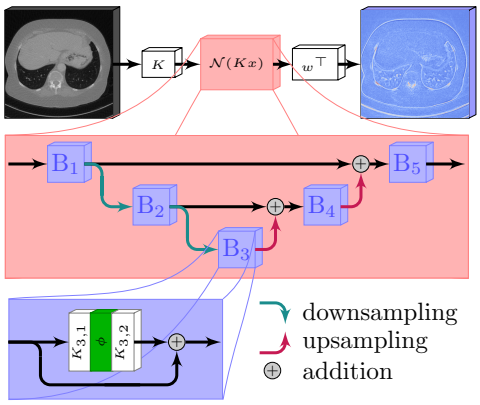


Fig. 2. Illustration of the TDV regularizer, which assigns to every voxel a regularization energy $r(x, \vartheta)$. The CNN $\mathcal{N}$ consists of residual blocks B_i operating on three scales in a U-Net type architecture.

which is the sum of the voxel-wise deep variation $r(x, \vartheta) = w^T \mathcal{N}(Kx) \in \mathbb{R}^n$ illustrated in Fig. 2. In detail, $K \in \mathbb{R}^{nc \times n}$ represents a learned 3D convolution kernel with zero-mean constraint. The non-linear function $\mathcal{N} : \mathbb{R}^{nc} \to \mathbb{R}^{nc}$ is a CNN and $w \in \mathbb{R}^c$ is a learned weight vector. Following [14], $\mathcal{N}$ is a U-Net type network with residual building blocks using the activation function $\phi = \frac{1}{2\alpha}\log(1 + \alpha x^2)$. We encode all trainable parameters of TDV by $\vartheta \in \Theta$, where Θ is the set of admissible parameters reflecting the zero-mean constraint on K. To enable an efficient implementation of the regularizer, all 3D convolutions are computed by first applying a 2D convolution in x/y-plane followed by a 1D convolution along the z-axis. The support of the resulting convolution filter is $3 \times 3 \times 3$. We use $c = 16$ feature channels to reduce the memory footprint.

To estimate the parameters ϑ of TDV we consider a sampled discrete optimal control problem. To this end, let $(x_0^i, y^i)_{i=1}^N$ be a collection of N training pairs consisting of an initial degraded volume x_0^i and a corresponding fully-sampled reference y^i. Then, the optimal control training problem is

$$\inf_{T, \vartheta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \|x_S^i - y^i\|_\epsilon \tag{4}$$

subject to

$$x_{s+1}^i = \frac{1}{1 + \frac{T}{S}}\left(x_s^i + \frac{T}{S}\left(x_0^i - \nabla_1 \mathcal{R}(x_s^i, \vartheta)\right)\right) \tag{5}$$

for $s = 1, \dots, S-1$ and $i = 1, \dots, N$. Here $\|\cdot\|_\epsilon$ is the Huber-norm, $T \in \mathbb{R}^+$ is the learned stopping time, and $S \in \mathbb{N}$ denotes the number of temporal steps. The iterative scheme (5) is motivated by a semi-implicit discretization of a gradient flow on the energy $\mathcal{R}(x, \vartheta) + \frac{1}{2}\|x - x_0^i\|_2^2$, which is equivalent to a proximal gradient method [19].

We use the sinogram data of 10 fully-sampled clinical 3D in vivo abdominal CT scans of different patients acquired by a Siemens Definition AS scanner using a routine clinical dose. Two of these scans include metal artifacts. We select those two scans for testing, while the remaining scans are used for training. To account for the limited memory of a GPU, we split the fully-sampled sinogram data of each scan into chunks

consisting of 3 full gantry rotations that overlap by 1 rotation, resulting in 133 training samples. We compute the reference volumes y^i by solving the regularized inverse problem (2) to account for the noise in the clinical sinogram data. In detail, we choose the commonly used regularizer $\mathcal{R}(x) = \sum_{j=1}^{3n} \phi((Dx)_j)$ and $\alpha = 1000$ for $\lambda = 0.0125$ using the fully-sampled data (i.e. M is the identity matrix) and an accelerated ordered-subset approach with 8 subsets. Here, D refers to the voxel-wise finite difference operator. The initial reconstructions x_0^i are computed by performing 40 conjugate gradient (CG) steps to solve (2) with $\lambda = 0$ for SparseCT data, where M reflects the W4S16 undersampling with a speed of $v \in \{1, 4\}$ detector rows between two gantry angles.

Given the training samples, we use the Adam optimizer [20] with step size 10^{-4} and default momentum to estimate the $45,936$ parameters of the TDV regularizer and the stopping time T by minimizing (4) with $S = 10$. In each of the $50,000$ training iterations, 6 patches of size $18 \times 96 \times 96$ are randomly sampled from the training data.

The resulting TDV regularizer with fixed learned parameters ϑ can be applied to denoise an initial corrupted reconstruction by applying (5). Moreover, the variational structure of TDV enables an application in the model-based CT reconstruction setting with balances parameter $\lambda > 0$ as follows:

$$x_R = \arg\min_{x \in \mathbb{R}^n} \lambda \mathcal{R}(x, \vartheta) + \frac{1}{2} \|M(Ax - z)\|_W^2. \qquad (6)$$

V. Results

The reconstruction of abdominal data with 4-fold retrospective SparseCT undersampling is presented in Fig. 3 and Fig. 4. The first column shows the initial non-regularized conjugate gradient (CG) reconstruction with undersampling artifacts, the second column shows denoising of the CG initialization regularized by TDV using (5), the third column shows TDV reconstruction results using (6), and the last column shows fully-sampled reference images. Two cases of the SparseCT undersampling that correspond to two different speeds v of the MSC are included: 1) $v = 4$ detector-rows per gantry angle (higher undersampling performance, Fig, 3) and 2) $v = 1$ detector-rows per gantry angle (lower undersampling performance, Fig. 4). In both cases, TDV reconstruction outperforms TDV denoising, and presents images that are closer to the reference images qualitatively and in terms of PSNR and SSIM. TDV denoising presents blurring artifacts. The incorporation of data consistency in TDV reconstruction enables to reinsert the high resolution features from the raw data, which results in significant improvements in image quality. SparseCT with $v = 4$ outperforms $v = 1$ as expected, given the superior undersampling performance in the former scheme. However, the difference is lower in the case of TDV reconstruction with respect to TDV denoising, which shows the ability of the latter to adapt to situations with data of lower acquisition quality.

VI. Conclusion

This work proposed to replace the standard sparsity enforcing regularizer in model-based SparseCT reconstruction by a regularizer that was learned to denoise corrupted and noisy CT images. The proposed training approach for the TDV regularizer enables an effective training since the time and memory consuming data consistency is not used. Using the TDV regularizer for reconstruction is a promising alternative to classical sparsity enforcing regularizers and/or deep learning reconstruction schemes due to its variational structure and performance and can lead to high CT dose reduction with minimal impact on image quality.

In future we work on applying the proposed TDV regularizer to SparseCT acquired in vivo data, however to do so the SparseCT prototype scanner needs FDA approval.

References

[1] T. Koesters, F. Knoll, A. Sodickson, D. K. Sodickson, and R. Otazo, "SparseCT: interrupted-beam acquisition and sparse reconstruction for radiation dose reduction," in *Medical Imaging 2017: Physics of Medical Imaging*. SPIE, 2017, pp. 174 – 180.

[2] B. Chen, E. Kobler, M. J. Muckley, A. D. Sodickson, T. O'Donnell, T. Flohr, B. Schmidt, D. K. Sodickson, and R. Otazo, "Sparsect: System concept and design of multislit collimators," *Med Phys*, vol. 46, no. 6, pp. 2589–2599, 2019.

[3] M. J. Muckley, B. Chen, T. Vahle, T. ODonnell, F. Knoll, A. D. Sodickson, D. K. Sodickson, and R. Otazo, "Image reconstruction for interrupted-beam x-ray CT on diagnostic clinical scanners," *Phys Med Biol*, vol. 64, p. 155007, 2019.

[4] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[5] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J Imaging Sci*, vol. 3, no. 3, pp. 492–526, 2010.

[6] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans Image Process*, vol. 26, no. 9, pp. 4509–4522, 2017.

[7] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans Med Imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[8] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction," *Med Phys*, vol. 44, no. 10, pp. e360–e375, 2017.

[9] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans Med Imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.

[10] E. Kobler, M. Muckley, B. Chen, F. Knoll, K. Hammernik, T. Pock, D. Sodickson, and R. Otazo, "Variational deep learning for low-dose computed tomography," in *Proc IEEE Int Conf Acoust Speech Signal Process*. IEEE, 2018, pp. 6687–6691.

[11] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Trans Med Imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[12] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier, "NETT: Solving inverse problems with deep neural networks," *Inverse Probl*, 2020.

[13] A. Kofler, M. Haltmeier, T. Schaeffter, M. Kachelrieß, M. Dewey, C. Wald, and C. Kolbitsch, "Neural networks-based regularization of large-scale inverse problems in medical imaging," *arXiv preprint arXiv:1912.09395*, 2019.

[14] E. Kobler, A. Effland, K. Kunisch, and T. Pock, "Total Deep Variation for linear inverse problems," *arXiv preprint arXiv:2001.05005*, 2020.

[15] S. Ha and K. Mueller, "A Look-Up Table-Based Ray Integration Framework for 2-D/3-D Forward and Back Projection in X-Ray CT," *IEEE Trans Med Imaging*, vol. 37, no. 2, pp. 361–371, 2017.

[16] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med Phys*, vol. 34, no. 11, pp. 4526–4544, 2007.

[17] S. C. Zhu and D. Mumford, "Prior learning and Gibbs reaction-diffusion," *IEEE T Pattern Aanl*, vol. 19, no. 11, pp. 1236–1250, 1997.

[18] D. Kim, S. Ramani, and J. A. Fessler, "Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction," *IEEE Trans Med Imaging*, vol. 34, no. 1, pp. 167–178, Jan 2015.

[19] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J Imaging Sci*, vol. 2, no. 1, pp. 183–202, 2009.

[20] D. P. Kingma and J. L. Ba, "ADAM: a method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

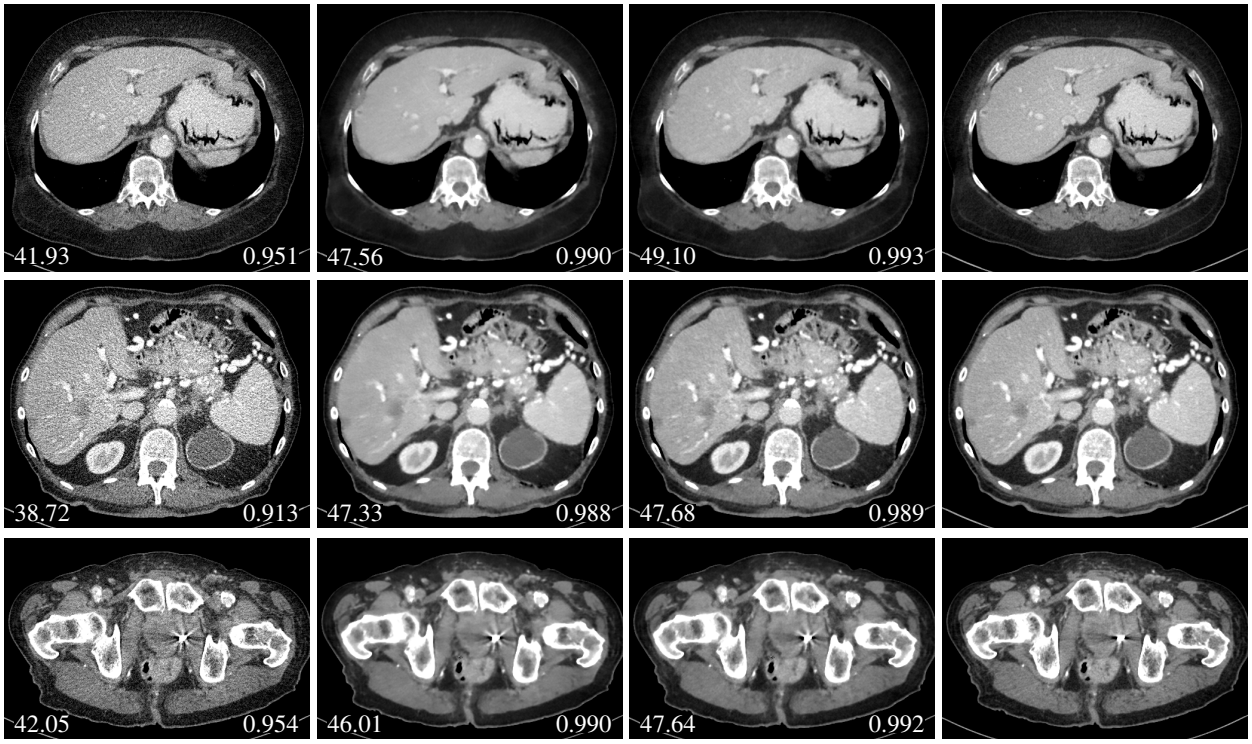


Fig. 3. Reconstruction of 4-fold undersampled SparseCT sinogram data with collimator speed $v = 4$ detector-rows per gantry angle (higher undersampling performance). From left to right: unregularized CG initialization, TDV denoising, TDV reconstruction, fully-sampled reference. Images are displayed in the range 40 ± 200HU. The PNSR scores w.r.t. to the reference y are depicted on the lower left corner and the SSIM score in the lower right corner.

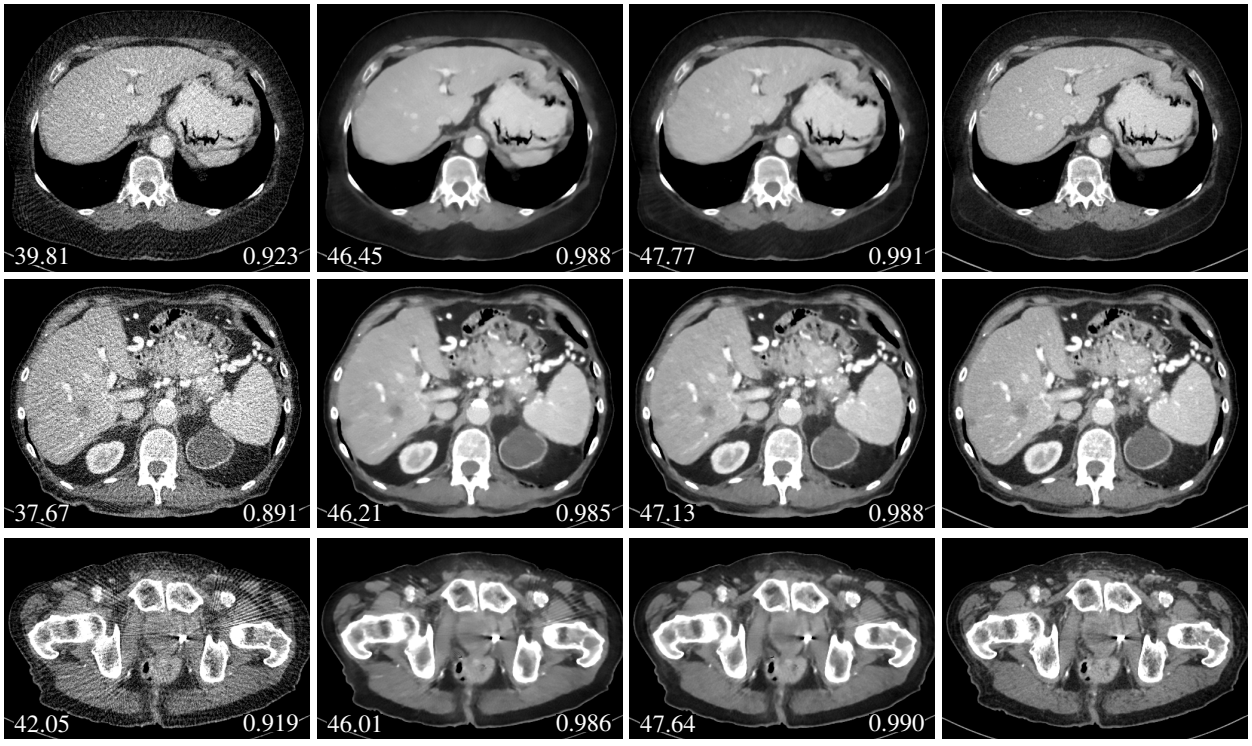


Fig. 4. Reconstruction of 4-fold undersampled SparseCT sinogram data with collimator speed $v = 1$ detector-rows per gantry angle (lower undersampling performance). From left to right: unregularized CG initialization, TDV denoising, TDV reconstruction, fully-sampled reference. Images are displayed in the range 40 ± 200HU. The PNSR scores w.r.t. to the reference y are depicted on the lower left corner and the SSIM score in the lower right corner.

X-ray Computed Tomography with a Robotic Sample Holder

Erdal Pekel*, Martin Dierolf†, Franz Pfeiffer†, Tobias Lasser*

***Abstract*—In this work we present the integration of a robotic arm with seven degrees of freedom as a sample holder within an experimental X-ray computed tomography setup. The robot enables arbitrary rotations of the sample and hence different non-standard acquisition trajectories that are not possible with conventional sample holding mechanisms. Our framework provides a programmatic interface for trajectory inputs and their motion planning, a robotic arm calibration mechanism for accurate target placement utilizing a single RGB camera, and both passive and active collision detection mechanisms based on a set of depth cameras. We elaborate on the challenges of precisely positioning samples into the X-ray beam and propose a method together with a discussion on its current limitations, in particular in view of performing X-ray computed tomography.**

I. Introduction

In this work we introduce a flexible robotic arm with seven degrees of freedom as a sample holder into an X-ray Computed Tomography (CT) setup. The robot adds flexibility to the setup as a sample holder by enabling arbitrary rotation and placement of the sample. This allows non-standard trajectories that are not restricted in their sequence, such as conventional circular or helical trajectories. In addition, the robotic sample holder can avoid occlusions on the projections that would normally be introduced by limitations of static setups where the sample is inherently mounted to non-moving parts (e.g. mounted on a plate). In the following we present a study on the feasibility of the integration of a robotic arm with seven degrees of freedom into an X-ray CT setup; we also discuss it's limitations and the future potential for X-ray CT applications.

II. Robotic Software Framework

In this section we briefly introduce the software architecture for the integration of a robotic arm with seven degrees of freedom (in our case, a FRANKA EMIKA Panda robot) within an existing X-ray computed tomography setup. The implementation comprises a set of independent ROS nodes written in C++ for optimal performance and safety. A set of different collision detection mechanisms is provided for safe robot operation. The positioning of the robot relative to the X-ray detector is calibrated using a commodity industrial RGB camera.

Fig. 1. Laboratory X-ray computed tomography setup, with a robotic arm with seven degrees of freedom holding a sample into the X-ray beam.

A. Technology Stack

At the center of the implementation is the Robot Operating System (ROS Melodic Morenia) [10]. A set of independent nodes that provide different functionality by exposing network endpoints are built on top of communication protocols predefined by the ROS ecosystem. Source code can be found at [9].

Additionally, we make use of:

- *MoveIt* [3, 12] for manipulation and motion planning,
- *Gazebo* [6] for robot simulation,
- *VISP* [8] for camera calibration,

*E. Pekel and T. Lasser are with the Department of Informatics, Technical University of Munich, Germany.

†M. Dierolf and F. Pfeiffer are with the Chair of Biomedical Physics, Department of Physics and Munich School of BioEngineering, Technical University of Munich, Germany.

‡F. Pfeiffer is also with the Department of Diagnostic and Interventional Radiology, School of Medicine and Klinikum rechts der Isar, Technical University of Munich, Germany.

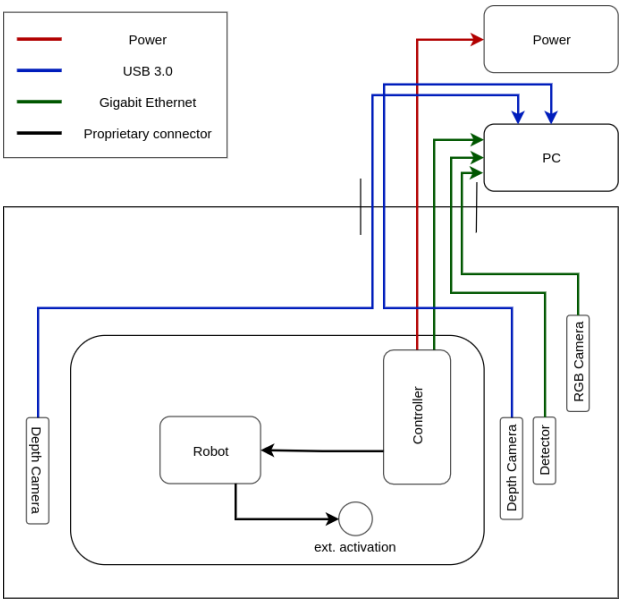


Fig. 2. Schematic of the laboratory X-ray computed tomography setup.

- *OpenCV* [2] for various image analysis tasks,
- *OpenCL* [11] for real-time collision detection,
- *OpenMesh* [1] for 3D mesh file handling,
- *elsa* [7] for the reconstruction of the sample, and
- *Eigen* [5] for linear algebra calculations.

B. Safety Measures

For safe robot operation in the laboratory we implement two collision detection mechanisms. The passive collision detection is implemented by reading files from the file system and passing the geometric shapes defined therein to MoveIt! [12], which uses the FCL library to avoid these objects during the motion planning process. In our implementation this is only useful for objects that do not change their position often, such as optical tables or walls.

In order to avoid collisions with unexpected objects during the robot's movements we record its immediate surroundings with two depth cameras and pass these to a custom GPU-based collision detection algorithm based on the open standard OpenCL [11].

III. Robotic Computed Tomography

In this section we will shortly introduce our laboratory setup, followed by the out-of-the-box placement precision and repeatability of the robot in the context of high resolution CT, and we will outline our approach for improving the aforementioned characteristics. Then we will introduce the results of a first experimental sample acquisition on a tilted circular trajectory. Finally, a 3D reconstruction of the sample is presented. A registration step ensures that the positioning error introduced by the robot is corrected in the sample projection images.

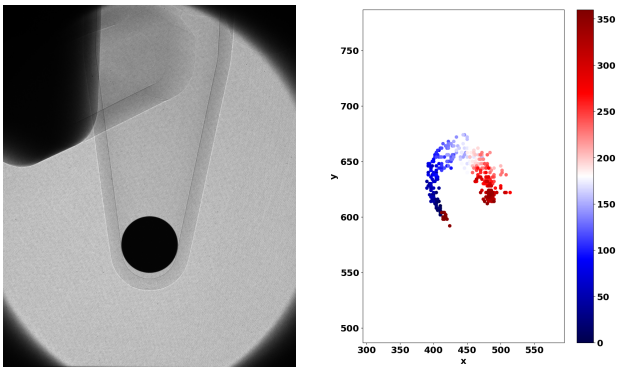


Fig. 3. *Left:* example X-ray projection measured of a simple sphere phantom consisting of a sphere in a falcon tube. One of the fingers of the robotic arm is visible in the top left. *Right:* distribution of the sphere center for several projections acquired along a circular trajectory with the angular position color-coded (axis scaled).

A. Laboratory setup

The laboratory setup consists of an experimental X-ray computed tomography setup with X-ray source and a high-resolution X-ray detector. We augmented this setup with a robotic arm with seven degrees of freedom (FRANKA EMIKA Panda) as a sample holder, two depth cameras (Intel RealSense D435) for collision detection, and a single RGB camera (Basler acA1920-40gc) for calibrating the robot position with the X-ray detector, see Figs. 1 and 2.

B. Target Placement

The target placement precision is critical for computed tomography, especially when paired with high-resolution detectors. The detector pixel size in our experiments was $9\mu m$, so a displacement of $9\mu m$ is enough to shift the projection by one pixel, which in turn affects the reconstruction process.

A major challenge with the robotic integration is the fact that the motion planning should be performed for the last link in the arm's chain, which is offset to the center of the fingers of the robot by $10.34cm$. This implies that the fixed goal tolerance of $10\mu m$ that we specified manually with *MoveIt* does not apply to the sample's position, but only at the last robot arm link, and it propagates relative to that position. The manufacturer's specification of $100\mu m$ for the pose repeatability (ISO 9283) [4] is only reached at an offset of roughly $10cm$ to the sample's position, which increases with the sample size. In our heuristic

analysis of the displacement at the sample's center with the optimizations outlined in the following paragraph, we always observed a total error of $300\mu m$ and more for the sample's position (see Fig. 4 on the right).

In Fig. 4 the default motion planning configuration of the robot is used to place a ball pen's head at the same cartesian goal position starting from a fixed configuration. An X-ray projection of the ball pen's head was acquired and annotated manually after the robot reached its goal; each blue point denotes an instance were the robot reached the goal successfully. The last X-ray projection was used as a background for illustration purposes. Our observation is that with the default motion planning configuration the robot is optimized for responsiveness and speed. When the allowed trajectory execution time is increased from $0.5s$ to $2.0s$ and the joint goal tolerance is tightened from 0.05 to 0.0005 radians, and a different inverse kinematics plugin is employed (*trac_ik*, *Manip1* method) with an increased solver timeout from $0.005s$ to $0.8s$, the same procedure lead to a more precise placement pattern with minimal deviation from the mean of all points (see Fig. 4).

However, for full circular trajectories, where slight actuation of a single joint is not always enough for reaching the next pose and hence the robot has to actuate all joints, the *trac_ik* solver poses a problem, especially when it is restricted in its moving space by surrounding equipment. Candidate solutions are not checked for self-collisions and hence self-collision reflexes of the robot's low-level controller implementation are triggered. Checking the self-collision would require that the actual link volumes are considered by the *trac_ik* solver, for example with actual 3D mesh files which incorporate information about the robot's geometry. The reflex interrupts the trajectory execution for a single goal pose and with it the execution of the circular trajectory as a whole, and thus represents a challenge that our implementation currently cannot recover from. An extension solving this issue is in development.

C. Experimental Setup

For the reconstruction experiments we placed a sphere with $2mm$ radius into a plastic enclosure (falcon tube) which was manually placed between the robot's fingers. A full circular trajectory with 350 poses was sampled for a fixed point in the robot's coordinate frame, tilted around the x-axis by $-30°$. Each pose is targeted individually, hence motion planning and trajectory execution from pose **i** to pose **i+1** is executed independently. A sample detector image ($9\mu m$ pixel size, binned and interpolated to $30\mu m$ in post-processing) from the acquisition can be seen in Fig. 3. Improving the repeatability for a specific pose as shown in Fig. 4 did not, however, lead to the desired accuracy for a series of independent goal poses. In Fig. 3 (right) the

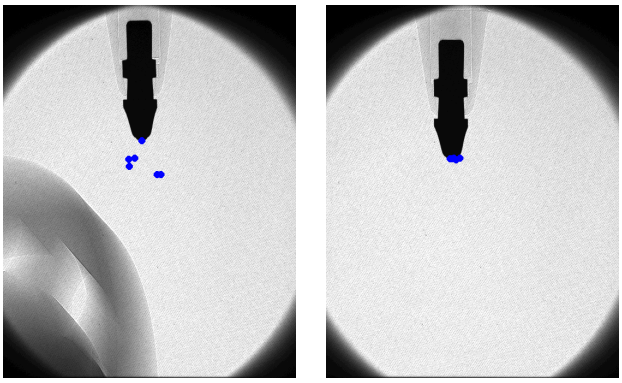


Fig. 4. Comparison of default robot placement precision *(left)* to our optimizations *(right)* for identical robot goal poses, overlaid over an example X-ray projection of a ball pen's head.

centers of the sphere for the 350 acquisitions of the full circular trajectory are plotted in absolute pixel coordinates (axis scaled, original size 806×1209). The color of the centers encodes the angle in the circular trajectory, with the final pose rotated by $359°$ around the robot's initial pose. From the deviation of around 100 pixels of the sphere centers on the scatter plot we can conclude that despite the aforementioned optimizations for single pose repeatability, the target needs to be registered separately in order to be reconstructed properly.

D. Target Registration and Reconstruction

For the purposes of demonstration we chose a sample with a very simple shape, a sphere that can easily be registered from any angle as a circle. We determined the center of the circle by means of Hough Circles with the OpenCV library in a post-processing step, and applied their offset to the detector center. The images from the X-ray detector were cropped in post-processing (see Fig. 5 top).

For tomographic reconstruction, the sinogram contained 350 equi-distant X-ray projections along a tilted circular trajectory, sized 806×310 with a spacing of $30\mu m$. The reconstruction volume was sized $806 \times 310 \times 350$ with isotropic spacing of $30\mu m$. Using our C++ reconstruction framework *elsa* [7], reconstruction was performed using an iterative conjugate gradient solver run until convergence on a Tikhonov regularized weighted least squares problem, with the Josephs method for X-ray transform discretization and parallel beam geometry.

We were able to reconstruct the sphere as seen in Fig. 5 in the middle row. To confirm our findings, we simulated X-ray projections of an equally sized sphere and reconstructed the resulting projections using the same acquisition geometry and settings in an inverse crime scenario, see

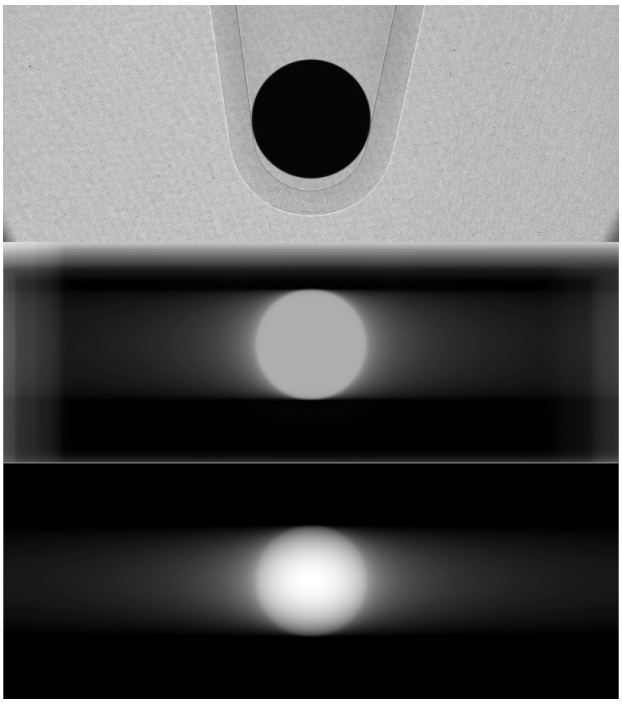


Fig. 5. Initial experimental X-ray CT results. *Top:* example cropped X-ray projection of the sphere sample. *Middle:* slice of the reconstructed volume using 350 X-ray projections. *Bottom:* reconstruction of a synthetic sphere using the same acquisition geometry.

Fig. 5 bottom row. From the comparison to the synthetic example we can conclude that the differences are minor. The reconstruction in the middle shows line artifacts on the borders resulting from the detector shift applied after registering the sample.

To facilitate future accurate acquisitions using this setup, it will be necessary to design a sample holder that first does not interfere with the X-ray projection image of the sample, and second provides multiple high-contrast landmarks for an automated registration step to provide the desired accuracy. This sample holder should also be tailored to the robot's end-effector characteristics, while not limiting the degrees of freedom of the robotic arm.

IV. Conclusion

In this study we demonstrated the integration of a robotic arm with seven degrees of freedom into an experimental X-ray CT setup as a replacement for conventional sample holding mechanisms. We briefly presented the software stack involved in providing safe positioning of the robot. We studied the robot's placement accuracy and repeatability within the *ROS* and *MoveIt* frameworks. From our study we conclude that a separate registration step using the X-ray projection images is required for reaching the desired micrometer-range placement accuracy with the given setup. To demonstrate the feasibility of our approach, we performed a first tomographic reconstruction of a simple sphere phantom, allowing for a simple registration step. Our proposed setup will enable the fast, accurate acquisition of arbitrary non-standard trajectories for X-ray CT for various applications.

References

[1] Mario Botsch et al. "Openmesh-a generic and efficient polygon mesh data structure". In: (2002).

[2] G. Bradski. "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* (2000).

[3] David Coleman et al. *Reducing the Barrier to Entry of Complex Robotic Software: a MoveIt! Case Study*. 2014. arXiv: 1404.3785.

[4] FRANKA EMIKA. *PANDA DATASHEET*. 2020. URL: https://s3-eu-central-1.amazonaws.com/franka-de-uploads/uploads/2019/04/Datasheet.pdf.

[5] Gaël Guennebaud, Benoît Jacob, et al. *Eigen v3*. http://eigen.tuxfamily.org. 2010.

[6] Nathan Koenig and Andrew Howard. "Design and use paradigms for gazebo, an open-source multi-robot simulator". In: *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*. Vol. 3. IEEE. 2004, pp. 2149–2154.

[7] Tobias Lasser, Maximilian Hornung, and David Frank. "elsa - an elegant framework for tomographic reconstruction". In: *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. Vol. 11072. SPIE, 2019, pp. 570–573. DOI: 10.1117/12.2534833. URL: https://elsa.campar.in.tum.de/.

[8] E. Marchand, F. Spindler, and F. Chaumette. "ViSP for visual servoing: a generic software platform with a wide class of robot control skills". In: *IEEE Robotics and Automation Magazine* 12.4 (Dec. 2005), pp. 40–52.

[9] Erdal Pekel. *github repository [online]*. https://github.com/erdalpekel.

[10] Morgan Quigley et al. "ROS: an open-source Robot Operating System". In: *ICRA workshop on open source software*. Vol. 3. 3.2. Kobe, Japan. 2009, p. 5.

[11] J. E. Stone, D. Gohara, and G. Shi. "OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems". In: *Computing in Science Engineering* 12.3 (May 2010), pp. 66–73. ISSN: 1558-366X. DOI: 10.1109/MCSE.2010.69.

[12] Ioan A. Sucan and Sachin Chitta. *MoveIt: [online]*. http://moveit.ros.org/.

DBP Domain Deep Learning for CT Reconstruction

Junyoung Kim*, Yoseob Han*, and Jong Chul Ye, *Fellow, IEEE*

***Abstract*—Currently, most of the deep learning approaches for CT reconstruction are based on the image domain learning. Inspired by the recent understanding of deep neural network as an adaptive signal representation, here we show that there exists another domain known as the differentiated backprojection (DBP) domain, in which neural network can be implemented efficiently. Experimental results for interior tomography and circular conebeam artifact removal show that the DBP domain deep learning outperforms the existing methods with better generalization capability.**

***Index Terms*—Differentiated backprojection, deep learning, interior tomography, circular conebeam artifacts**

I. Introduction

X-RAY computed tomography (CT) is now considered as the first examination before treatment plan is made by doctors. The standard method for image reconstruction is filtered-backprojection (FBP) algorithms where the sinogram data is filtered along the specific 1-D filter lines, after which the backprojection operation is performed to obtain 3-D reconstruction. Currently, FBP algorithms are available not only 2-D parallel and fanbeam geometry, but also complicated conebeam geometry [1].

In recent years, deep learning approaches have been successfully used for CT image reconstruction problems such as low-dose CT [2], [3], sparse-view CT [4], [5], interior tomography [6], [7], and so on. Most of these approaches are based on the image domain processing, where the FBP reconstruction is first performed to obtain noisy images, from which the deep neural network is trained to learn the mapping between the noisy image and clean output. The reason that image domain deep learning is so popular is that the original formulation of the convolutional neural network (CNN) was inspired by the information processing in retina and visual cortex, so there are common beliefs that the CNN should be implemented in such a way.

In our recent theoretical results [8], we showed that the CNN is a novel signal representation, where the rectified linear unit (ReLU) activation patterns automatically chooses distinct linear representation depending on input signal. This perspective of deep neural network tells us that a CNN can be more efficiently implemented in the domains where the concise signal representation exists. For example, this led to the k-space deep learning for the case of magnetic resonance imaging (MRI) [9], where CNN is implemented in the Fourier domain.

In this paper, we show that there exists another domain that deep neural network can be efficiently implemented for CT reconstruction problems. In particular, we introduce a new class of deep learning methods on the differentiated backprojection (DBP). The DBP step converts the CT image reconstruction problems to a deconvolution problem, which makes the CNN learning much simpler.

While this idea was introduced for interior tomography [7], no unified analysis of the DBP domain deep learning was not provided. Moreover, with the introduction of the deep learning, beautiful theories of analytic reconstruction in early 2000 [1], [10]–[12] have been largely forgotten. Therefore, this paper revisits the classical results on analysis reconstruction in particular using DBP, and explains how these ideas can be ideally combined with deep learning.

II. Theory

A. Differentiated Backprojection (DBP)

Let $\boldsymbol{x} = (x, y, z)$ denotes the point in $\mathbb{R}^3$, where x, y and z are the Cartesian coordinates with respect to the patient. The X-ray transform D_f, which maps $f(\boldsymbol{x})$ into the set of its line integrals, is defined as

$$D_f(\boldsymbol{a}, \boldsymbol{\theta}) = \int_{-\infty}^{\infty} dt \; f(\boldsymbol{a} + t\boldsymbol{\theta}) \; , \tag{1}$$

where $\boldsymbol{\theta}$ denotes a vector on the unit sphere $\mathbb{S} \in \mathbb{R}^3$.

For a given X-ray source trajectory $\boldsymbol{a}(\lambda)$, the differentiated backprojection (DBP) on a point $\boldsymbol{x} \in \mathbb{R}^3$ using the projection data from X-ray source trajectory $\boldsymbol{a}(\lambda), \lambda \in [\lambda^-, \lambda^+]$ is defined as [10]–[12]:

$$g(\boldsymbol{x}) = \int_{\lambda^-}^{\lambda^+} d\lambda \; \frac{1}{\|\boldsymbol{x} - \boldsymbol{a}(\lambda)\|} \frac{\partial}{\partial \mu} D_f(\boldsymbol{a}(\mu), \boldsymbol{\theta}) \Big|_{\mu=\lambda} , \tag{2}$$

where $1/\|\boldsymbol{x} - \boldsymbol{a}(\lambda)\|$ is the magnification factor dependent weighting. Then the DBP data in (2) can be represented as [12], [13]:

$$g(\boldsymbol{x}) = \frac{1}{(2\pi)^3} \int d\boldsymbol{\omega} \; \hat{f}(\boldsymbol{\omega}) e^{\iota \boldsymbol{\omega}^\top \boldsymbol{x}} \sigma(\boldsymbol{x}, \boldsymbol{\omega}, \lambda^\pm), \tag{3}$$

where $\iota = \sqrt{-1}$, $\hat{f}(\boldsymbol{\omega})$ is the 3-D Fourier transform of $f(\boldsymbol{r})$, and the signum function is given by

$$\sigma(\boldsymbol{x}, \boldsymbol{\omega}, \lambda^\pm) = \iota\pi[\mathrm{sgn}(\boldsymbol{\omega}^\top \boldsymbol{\alpha}(\lambda^-, \boldsymbol{x})) - \mathrm{sgn}(\boldsymbol{\omega}^\top \boldsymbol{\alpha}(\lambda^+, \boldsymbol{x}))],$$

where $\boldsymbol{\alpha}(\lambda, \boldsymbol{x})$ is the filtering direction known as *chord lines*:

$$\boldsymbol{\alpha}(\lambda, \boldsymbol{x}) \;\; = \;\; \frac{\boldsymbol{x} - \boldsymbol{a}(\lambda)}{\|\boldsymbol{x} - \boldsymbol{a}(\lambda)\|}. \tag{4}$$

J. Kim, and J. C. Ye are with the Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea (e-mail: {junyoung.kim, jong.ye}@kaist.ac.kr). Y. Han is with the Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA (e-mail: yoseobhan@lanl.gov). J.C. Ye is also with the Department of Mathematical Sciences, KAIST. *These authors contributed equally to this work. This work was supported by National Research Foundation of Korea, Grant number NRF-2020R1A2B5B03001980.

The following are already well-known acquisition trajectories for the chord line processing, which we will investigate for our applications.

- *2D Fan-beam CT:* In 2D circular fan-beam CT, the chord line $\mathcal{L}$ that passes through $\boldsymbol{x}$ inside the circle is not unique; hence, we can choose multiple filtering lines. For example, a vertical chord line that goes through $\boldsymbol{x}$ is determined by $\boldsymbol{a}(\lambda^-)$ and $\boldsymbol{a}(\lambda^+)$. Similarly, we can use horizontal chord lines.
- For the circular trajectory, the X-ray source rotates around the object $f(\boldsymbol{x})$ such that the source trajectory is represented by

$$\boldsymbol{a}(\lambda) = \begin{bmatrix} R\cos\lambda & R\sin\lambda & 0 \end{bmatrix}^\top, \quad \lambda \in [0, 2\pi], \tag{5}$$

  where R denotes the radius of the circular scan. Similar to the fan-beam geometry, the *virtual* chord line for circular conebeam trajectory is not unique. The redundancy in the choice of chord lines will be exploited later in mitigating cone-beam artifacts from missing frequency regions in the circular cone-beam trajectory.

Regardless of the trajectories, the expression in (3) suggests that DBP provides a filtered version of $f(\boldsymbol{x})$. Since the spectrum of the filter $\sigma(\boldsymbol{x}, \boldsymbol{\omega}, \lambda^\pm)$ also depends on $\boldsymbol{x}$, the corresponding filter is a spatially varying filter. Therefore, the image reconstruction problem now becomes the image deconvolution problems.

B. Classical DBP Domain Processing

Here, we review the classical results of the backprojection filtration algorithm (BPF) which was derived from (3). Specifically, if the directional vector $\boldsymbol{\alpha}(\lambda^-, \boldsymbol{x})$ is set as a coordinate axis, then the authors in [10]–[12] showed that (3) can be decomposed to a successive application of 1-D Hilbert transform along the chord lines. The backprojection filtration (BPF) method then recovers the object on each chord line by taking the inverse Hilbert transform of the DBP data. More specifically, let $g(s)$ and $f(s)$ denote the DBP data and the unknown image along the chord lines. If we assume that they are confined in $|s| < 1$, then the inversion can be done using the finite Hilbert transform [14]:

$$f(u) = \frac{c}{\pi\sqrt{1-u^2}} - \frac{1}{\pi\sqrt{1-u^2}} \int_{-1}^{1} \frac{\sqrt{1-s^2}\, g(s)}{u-s} ds, \tag{6}$$

where

$$c := \int_{-1}^{1} f(s) ds \ . \tag{7}$$

This BPF algorithm has many advantages compared to the FBP approaches. First, as the image can be reconstructed line-by-line, reconstruction from super-short scan is possible. Moreover, due to the lack of the ramp filtering, the DBP images do not suffer from the detector truncation artifacts. This leads to extensive research of interior tomography using DBP. Finally, even for the non-complete trajectory like circular conebeam trajectory, the DBP allows factorized representation [15]. For example, consider a plane that is parallel to the z-axis and intersects the source trajectory at two locations $\boldsymbol{a}(\lambda^-)$ and $\boldsymbol{a}(\lambda^+)$. Let $\mathcal{P}$ denote such a plane of interest. Then, the main idea of the factorization methods [15], [16] is to convert the 3-D reconstruction problem to a successive 2-D problems on the planes of interest. Specifically, on a plane of interest $\mathcal{P}$, the authors in [15] showed that (3) can be simply represented as

$$g(t, z) = \pi \int_{-\infty}^{\infty} h_H(t-\tau) \left(f(\tau, z_1(\tau)) + f(\tau, z_2(\tau)) \right) d\tau \tag{8}$$

where $h_H(t)$ denotes the Hilbert transform, and $z_1(\tau)$ and $z_2(\tau)$ refer to two sources to voxel lines on $\mathcal{P}$. This implies that the conebeam reconstruction problem can be solved by solving a deconvolution problem in the DBP domain.

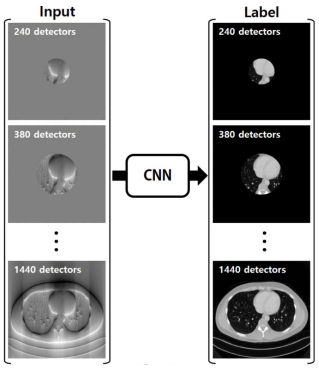


Fig. 1. Flowchart of DBP domain training for ROI tomography.

However, there are several limitations of classical DBP domain processing that prohibits its use in real applications. First, many of the chord lines are not on a cartesian grid, so specialized gridding operation is required to perform 1-D Hilbert filtering, which introduces additional complexity and numerical errors. Second, in the application of the interior tomography problem, the use of the finite Hilbert transform (6) no more works, since we do not know the exact value of DBP data $g(s)$ that are calculated within ROI regions. Finally, the actual implementation of the inverse Hilbert transform in (6) is sensitive to the bias term c in (7). This usually introduces streaking artifacts along the chord line, which makes the reconstruction quality degraded.

C. Advantages of DBP Domain Deep Learning

To address the problem, we proposed a neural network $\mathcal{T}_{\boldsymbol{\Theta}}$ parameterized by $\boldsymbol{\Theta}$, which is trained by solving the following optimization problem:

$$\min_{\boldsymbol{\Theta}} \sum_{i=1}^{N} \| \boldsymbol{F}^{(i)} - \mathcal{T}_{\boldsymbol{\Theta}} \boldsymbol{G}^{(i)} \|_2^2, \tag{9}$$

where $\{(\boldsymbol{F}^{(i)}, \boldsymbol{G}^{(i)})\}_{i=1}^N$ denotes the training data set composed of ground-truth 2-D image and 2-D DBP image along the adjacent (virtual) chord lines. In particular, to address the gridding problem at the same time, the training data are obtained from cartesian grid after the DBP operations. Fig. 1 shows the DBP domain training for the case of ROI tomography. We trained the neural network with four different ROI size DBP data.

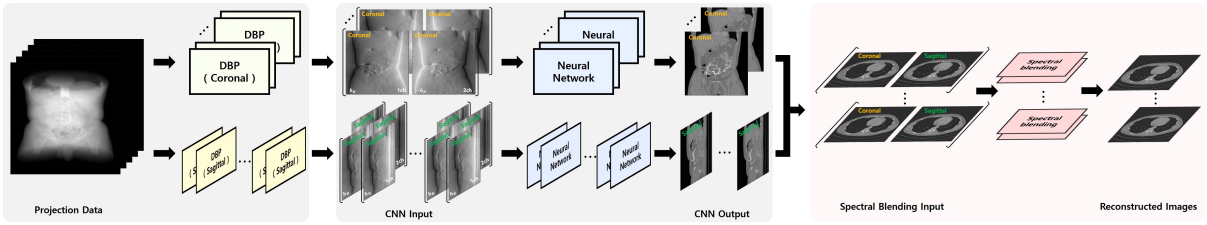


Fig. 2. Overview for differentiated backprojection domain deep learning for conebeam artifact removal.

For the case of conebeam reconstruction, although the factorization method leads to 2-D deep learning approach for each plane of interest, Theorem 2.2 in our previous work [13] showed that the missing frequency region depends on the direction of the plane of interest. However, we also showed in our prior work [13] that at any point $\boldsymbol{x}$, if the missing frequency regions for different filtering directions are appropriately combined, we can minimize the artifacts from the missing frequency regions. Accordingly, we can perform reconstruction for both coronal and sagittal directions, and combine them. This implies that except for the common frequency regions, the missing spectral components in coronal directional processing can be compensated by the results from the sagittal processing, and vice versa.

Fig. 2 shows a flowchart for comebeam artifact removal using DBP domain processing. In particular, the axial images f^{cor} and f^{sag} are obtained from the DBP domain deep neural networks. Then, coronal and sagittal reconstructed images are first transformed to the spectral domain using the 2-D Fourier transform. We apply the bow-tie spectral weighting to suppress the signal in the missing frequency regions and combined them together. This process can be mathematically represented by

$$f^{com} = \mathcal{F}^{-1}\left\{w \odot \mathcal{F}\{f^{cor}\} + (1-w) \odot \mathcal{F}\{f^{sag}\}\right\}.$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier and inverse Fourier transforms, respectively; w and $1-w$ are the bow-tie spectral masks, and $\odot$ denotes the element-wise multiplication.

III. Method

To train the neural network using simulated projection data, we used simulated projection data from ten subject data sets from American Association of Physicists in Medicine (AAPM) Low-Dose CT Grand Challenge were used. Out of ten patient, eight patient data were used as training sets, one patient data was used as validation set, and the other patient data was used for test set. The network backbone corresponds to a modified architecture of U-Net [17]. For quantitative evaluation, peak signal-to-noise ratio (PSNR), normalized mean-squared-error (NMSE), and structural similarity (SSIM) are used.

We performed two experiments. One for the interior tomography, and the other for conebeam artifact removal. The $x-y$ size of images is 512×512 and the z size ranges from 400 to 600, varying from the patient to the patient. The voxel size is 1 mm^3. The 512×512 size artifact-free CT images are reconstructed from the rebinned fan-beam CT data using filtered backprojection (FBP) algorithm. For the generation of the training data, sinogram are numerically obtained using forward fan-beam projection operator for our experiments. The number of detectors is 1440 elements with pitch of 1 mm. The number of views is 1200. The size of images is 512×512.

For the case of ROI reconstruction, the distance from source to origin (DSO) is 800 mm and the distance from source to detector (DSD) is 1400 mm. The truncated DBP data are exactly the same as the full DBP data within the ROI mask. For the case of conebeam artifact removal, the distance from source to origin (DSO) was 500 mm, and the distance from source to detector (DSD) was 1000 mm. The maximum cone angle was set to $35.8°$, and the projection data were generated accordingly.

IV. Results

A. Interior tomography

To confirm the generalization capability of the proposed method network with respect to varying ROI sizes, the trained

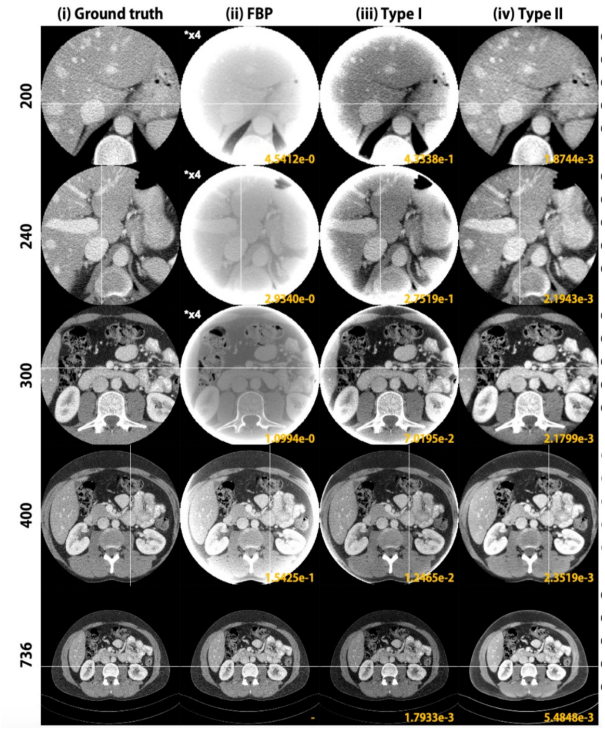


Fig. 3. Row direction: (i) ground-truth, and (ii) FBP images. Reconstruction images by (iii) neural network trained with FBP images, and (iv) our method. Row direction: interior images from 200, 240, 300, 400, and 736 detectors. The NMSE values are written at the corner. A window range is (-150, 300)[HU].

network using simulated projection data was applied to the real projection data from AAPM for different ROIs from 200 to 736 detectors. In Fig. 3, we provide reconstruction results using the network trained with FBP images (type I) and the DBP images (type II). The results clearly shows that our method is generalized better for various ROI sizes.

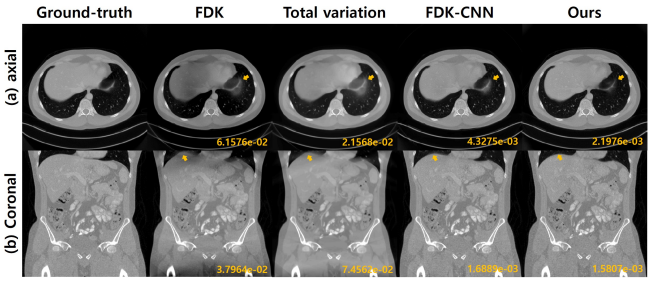


Fig. 4. (a) Axial and (b) coronal reconstruction results from cone-beam geometry. The number written to the images is the NMSE value. Intensity range of the CT image is (-1000, 800) [HU].

B. Conebeam artifact removal

Fig. 4 shows conebeam reconstruction images of AAPM data along the axial and coronal planes. As shown in Fig. 4, the FDK reconstructions around the midplane are similar to the ground-truth images, but poor intensity and the cone-beam artifact are visible in the off-midplanes such as top and bottom areas. Here, FDK-CNN refers to the image domain learning that learns the mapping between FDK images and the ground-truth images. Although TV and FDK-CNN methods can improve the image quality, there are still remaining conebeam artifacts, and the reconstruction does not preserve the detailed structures and textures. On the other hand, our method clearly removes the artifact, maintains sophisticated structures and textures.

Table I shows the quantitative comparisons of average PSNR, NSME, and SSIM and their computation times. Our method produces the best quantitative values in all metrics.

TABLE I
QUANTITATIVE COMPARISON FOR VARIOUS ALGORITHMS.

Metric	FDK	TV	FDK-CNN	Ours
PSNR [dB]	32.05	34.39	39.99	**42.49**
NMSE ($\times 10^{-3}$)	35.73	7.32	1.80	**0.94**
SSIM	0.78	0.78	0.91	**0.97**
Time [sec/slice]	-	14.06	0.10	0.12

V. CONCLUSIONS

In this paper, we provide a systematic study of DBP domain deep learning for CT reconstruction. Inspired by the BPF algorithm and the factorization approach that converts 3-D problem into a lower dimensional deconvolution problems, our neural network is designed as a 2-D neural network and was trained to learn the mapping between DBP data and the artifact-free images. Experimental results with interior tomography and conebeam artifact removal showed that the our method generalizes much better than the image domain CNN and is quantitatively and qualitatively superior to the existing methods.

VI. ACKNOWLEDGEMENTS

The authors would like to thanks Dr. Cynthia MaCollough, the Mayo Clinic, the American Association of Physicists in Medicine (AAPM), and grant EB01705 and EB01785 from the National Institute of Biomedical Imaging and Bioengineering for providing the Low-Dose CT Grand Challenge data set.

REFERENCES

[1] A. Katsevich, "Theoretically exact filtered backprojection-type inversion algorithm for spiral CT," *SIAM Journal on Applied Mathematics*, vol. 62, no. 6, pp. 2012–2026, 2002.
[2] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction," *Medical physics*, vol. 44, no. 10, 2017.
[3] E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep convolutional framelet denosing for low-dose CT via wavelet residual network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1358–1369, 2018.
[4] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.
[5] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.
[6] Y. Han, J. Gu, and J. C. Ye, "Deep learning interior tomography for region-of-interest reconstruction," in *Proceedings of The Fifth International Conference on Image Formation in X-Ray Computed Tomography*, 2018.
[7] Y. Han and J. C. Ye, "One network to solve all ROIs: Deep learning ct for any ROI using differentiated backprojection," *Medical Physics (in press), Also available as arXiv preprint arXiv:1810.00500*, 2018.
[8] J. C. Ye and W. K. Sung, "Understanding geometry of encoder-decoder CNNs," in *International Conference on Machine Learning*, 2019, pp. 7064–7073.
[9] Y. Han and J. C. Ye, "k-Space Deep Learning for Accelerated MRI," *IEEE Transactions on Medical Imaging (in press), also available as arXiv preprint arXiv:1805.03779*, 2019.
[10] J. D. Pack and F. Noo, "Cone-beam reconstruction using 1D filtering along the projection of M-lines," *Inverse Problems*, vol. 21, no. 3, p. 1105, 2005.
[11] Y. Zou and X. Pan, "Exact image reconstruction on PI-lines from minimum data in helical cone-beam CT," *Physics in Medicine and Biology*, vol. 49, no. 6, p. 941, 2004.
[12] ——, "Image reconstruction on pi-lines by use of filtered backprojection in helical cone-beam ct," *Physics in Medicine & Biology*, vol. 49, no. 12, p. 2717, 2004.
[13] M. Lee, Y. Han, J. P. Ward, M. Unser, and J. C. Ye, "Interior tomography using 1d generalized total variation. part ii: Multiscale implementation," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2452–2486, 2015.
[14] F. W. King, *Hilbert transforms*. Cambridge University Press Cambridge, UK, 2009, vol. 2.
[15] F. Dennerlein, F. Noo, H. Schondube, G. Lauritsch, and J. Hornegger, "A factorization approach for cone-beam reconstruction on a circular short-scan," *IEEE Transactions on Medical Imaging*, vol. 27, no. 7, pp. 887–896, 2008.
[16] L. Yu, Y. Zou, E. Y. Sidky, C. A. Pelizzari, P. Munro, and X. Pan, "Region of interest reconstruction from truncated data in circular cone-beam ct," *IEEE transactions on medical imaging*, vol. 25, no. 7, pp. 869–881, 2006.
[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*. Springer, 2015, pp. 234–241.

A novel 4D-CBCT reconstruction Framework via a Joint of Spatiotemporal Dictionary Learning and Prior Constraint

Shaohua Zhi, Marc Kachelrieß, and Xuanqin Mou

***Abstract*—Four-dimensional cone-beam computed tomography (4D-CBCT) is developed to provide a sequence of phase-resolved reconstructions for image-guided radiation therapy. However, 4D-CBCT images are degraded by severe streaking artifacts because it is an extreme sparse-view CT reconstruction process wherein only under-sampled projections are used for each phase. Considering the advantage of dictionary learning on representing structural features and correlation between neighboring pixels effectively, we develop a novel dictionary learning method for the scenario of 4D-CBCT reconstruction. In this study, we trained a spatiotemporal dictionary by exploring the high correlation among each phase. Furthermore, the spatiotemporal dictionary equipped with a prior knowledge is incorporated into a regularized 4D-CBCT iterative reconstruction framework. The proposed method could reconstruct a set of 4D-CBCT images achieving both high spatiotemporal resolution and detailed features preservation. A simulation phantom and a real patient dataset were tested to verify the advantage of our method on producing better 4D-CBCT images with less streaking artifacts and more detailed structural features, compared to the existing algorithms.**

***Index Terms*—Four-dimensional cone-beam computed tomography (4D-CBCT), Spatiotemporal dictionary learning, Prior knowledge.**

I. Introduction

Image-guided radiation therapy (IGRT) has been widely used in radiotherapy clinics[1]. In the IGRT system, an on-board kV Cone-Beam Computed Tomography (CBCT) unit is mounted on an MV linear accelerator (LINAC), which is further used to verify the patient's position, localize the target of treatment and offer possibilities for an update of the treatment protocol[2]. However, the respiratory motion of patients may lead to significant blurring artifacts due to the limited CT gantry rotation speed, which would affect the image quality of the reconstructions, as well as the accuracy of dose planning and delivery. Four-dimensional CBCT technique (4D-CBCT)[3] has been developed to overcome the shortcomings by first assigning the X-ray projections into several respiratory phases based on a breathing signal and then reconstructing a sequence of volumetric images. Nevertheless, severe streaking artifacts and noise are observed in each of them due to relatively fewer projections and large angular spacing in each phase.

To improve overall image quality, many efforts have been devoted to 4D-CBCT. Iterative reconstruction method equipped with a variety of regularizers or prior knowledge[4]–[7] are usually researched by taking the high temporal correlation of the phase-resolved images into account. In contrast to conventional sparse-view CT reconstruction methods that different phases of 4D-CBCT are individually considered, the 4D-CBCT iterative reconstruction frameworks are coupled with spatiotemporal constraints, such as temporal non-local means[5], and spatiotemporal TV[6]. Moreover, some works have attempted to incorporate prior knowledge into the reconstruction framework, such as the prior image constrained compressed sensing method (PICCS)[7].

Recently, deep learning techniques have achieved great success in CT image restoration and reconstruction[8]. However, to our knowledge, training a network needs high cost of labeling CT data and the network is modeled case by case without a robust generalization ability to reconstruct various CT image tasks. Especially in 4D-CBCT scenario labeling 4D-CT images is a difficult labor. Hence, it still needs to raise conventional regularization technologies to overcome these shortcomings.

It is shown that dictionary learning-based methods have gained promising results in the CT reconstruction field. Xu et al.[9] incorporated a dictionary learning-based sparse constraint into the low-dose CT iterative reconstruction framework for the first time. Focusing on the reconstruction task of the volumetric CT, a 3D dictionary learning method[10] was incorporated into the regularized iterative reconstruction framework with a higher representation efficiency of the 3D anatomic structures compared to 2D dictionary. In addition, a tensor-based dictionary learning method[11] was proposed for spectral CT reconstruction and material decomposition. Currently, most of the dictionary learning-based techniques are for spatial (2D/3D) cases and are not further investigated in the sequence of phase-resolved images for 4D-CBCT. It is expected that a trained spatiotemporal dictionary can represent the structures more efficiently by exploring the high correlation between each respiratory phase.

Motivated by this, in our study, we propose a 4D-CBCT regularized reconstruction framework by associating the trained spatiotemporal dictionary with the inherent prior property of 4D-CBCT. We trained the spatiotemporal dictionary using a sequence of high-quality 4D-CT images. In the iterative framework, the prior knowledge that regularize the

This work is supported in part by the National Natural Science Foundation of China (NSFC, No. 61571359) and the National Key Research and Development Program of China (No.2016YFA0202003). (Corresponding author: Xuanqin Mou.)

Shaohua Zhi and Xuanqin Mou are with the Institute of Image processing and Pattern recognition, Xi'an Jiaotong University, Xi'an 710049, China. (email: zhishaohua.1203@stuxjtu.edu.cn; xqmou@mail.xjtu.edu.cn)

Marc Kachelrieß is with German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120, Heidelberg, Germany. (email: marc.kachelriess@dkfz.de)

reconstructed sequence getting close to a prior reconstruction from the fully-sampled projections of the same dataset. Experiments validated the feasibility of the proposed method, especially in the streaking artifacts reduction, spatiotemporal resolution, and detailed structural features preservation.

The rest of this article is organized as follows. Section II describes the theory of the proposed method. Section III demonstrates the experimental design. Section IV analyzes the experimental results of a phantom data and real patient data. In Section V, we conclude the paper.

II. METHODS

A. Formulation

We will firstly formulate the 4D-CBCT iterative reconstruction, which consists of a projection binning process and an independent reconstruction process, as addressed in section I. In this procedure of the projection binning process, the total views of measured projections $\boldsymbol{P}$ are grouped into T subsets. Each projection subset at a specific phase is labeled as $\vec{\boldsymbol{p}}_t(t=1,2,...,T), t\in[1,T]$, representing one of the respiratory states in a respiratory cycle. Then, a set of phase-resolved images $\{\vec{\boldsymbol{X}}_t \mid t=1,2,...T\}$ were reconstructed independently via the regularized iterative reconstruction method, which is formulated as follows:

$$\vec{\boldsymbol{x}}_t^{opt} = \arg\min_{\vec{\boldsymbol{x}}_t} \left\| \boldsymbol{A}_t \vec{\boldsymbol{x}}_t - \vec{\boldsymbol{p}}_t \right\|_2^2 + \beta R(\vec{\boldsymbol{x}}_t) \tag{1}$$

where the first term denotes a data fidelity by minimizing a l_2 norm distance between the forward projections of the reconstructed image $\boldsymbol{A}_t\vec{\boldsymbol{x}}_t$ and the corresponding measured projection $\vec{\boldsymbol{p}}_t$. $R(\vec{\boldsymbol{x}}_t)$ is the regularization term by adding physical constraints. The regularization parameter β controls the trade-off between the two terms. Moreover, to make use of the high correlation property of each 4D-CBCT image, these images will be reconstructed together by regarding them as a spatiotemporal image matrix $\boldsymbol{X}=(\vec{\boldsymbol{x}}_1,\vec{\boldsymbol{x}}_2,...,\vec{\boldsymbol{x}}_T)$, in which each column represents an image vector at the given phase. Base on this, the forward projection model is formulated as:

$$\left(\vec{\boldsymbol{p}}_1 \quad \vec{\boldsymbol{p}}_2 \quad \cdots \quad \vec{\boldsymbol{p}}_T\right)^T = \left(\boldsymbol{A}_1\vec{\boldsymbol{x}}_1 \quad \boldsymbol{A}_2\vec{\boldsymbol{x}}_2 \quad \cdots \quad \boldsymbol{A}_T\vec{\boldsymbol{x}}_T\right)^T \tag{2}$$

Also, a compact form could be written as:

$$\boldsymbol{AX}=\boldsymbol{P} \tag{3}$$

Then, we modify Eq.(1) to reconstruct the spatiotemporal image sequence by the following objective function:

$$\boldsymbol{X}^{opt} = \arg\min_{\boldsymbol{X}} \left\| \boldsymbol{AX} - \boldsymbol{P} \right\|_2^2 + \beta R(\boldsymbol{X}) \tag{4}$$

B. Spatiotemporal Dictionary Learning

Before introducing the proposed dictionary learning method, we review the formulation of the conventional dictionary-based low-dose CT reconstruction method. Here, $\vec{\boldsymbol{x}}_{LD}$ denotes the target low-dose CT image and $\vec{\boldsymbol{p}}_{LD}$ is the corresponding projection data. The low-dose CT image is reconstructed by solving the constrained optimization function:

$$\begin{aligned} &\min_{\vec{\boldsymbol{x}}_{LD},\boldsymbol{\alpha}_s} \left\| \boldsymbol{A}\vec{\boldsymbol{x}}_{LD} - \vec{\boldsymbol{p}}_{LD} \right\|_2^2 + \beta \sum_s \left\| \boldsymbol{\alpha}_s \right\|_0 \\ &s.t. \ \left\| E_s\vec{\boldsymbol{x}}_{LD} - \boldsymbol{D}\boldsymbol{\alpha}_s \right\| \le \varepsilon \quad \forall s \end{aligned} \tag{5}$$

where E_s expresses the extraction operator for the s[th] data patch, which can be sparsely represented by a trained spatial dictionary $\boldsymbol{D}$ with a coefficient matrix $\boldsymbol{\alpha}_s$.

Regarding the reconstruction task of a sequence of 4D-CBCT images, the high correlation between different phases is considered in our proposed method. Different from 2D atoms of size $n_x \times n_y$ in the conventional dictionary, we develop a spatiotemporal patch of size $n_x \times n_y \times n_T$, where $n_x \times n_y$ is the spatial size in the space dimension and n_T is the number of phases in the time dimension. Normally, n_T is set 10 to stay consistent with the number of respiratory phases for the phase-resolved images. Each patch could reflect the motion trajectory if the patch is extracted in the moving region. Hence, a spatiotemporal dictionary $\boldsymbol{D}_{ST} \in \mathbb{R}^{K\times N}$ is trained with the sparse modeling software (SPAMS)[12]. Each column in $\boldsymbol{D}_{ST} \in \mathbb{R}^{K\times N}$ is a spatiotemporal atom of size and N denotes the number of spatiotemporal atoms. Hence, modifying Eq.(4) and Eq.(5) together by using the spatiotemporal dictionary learning (STDL), the 4D-CBCT iterative reconstruction optimization function can be written as:

$$\begin{aligned} &\boldsymbol{X}^{opt} = \arg\min_{\boldsymbol{X},\boldsymbol{\alpha}_s} \left\| \boldsymbol{AX} - \boldsymbol{P} \right\|_2^2 + \beta \sum_s \left\| \boldsymbol{\alpha}_s \right\|_0 \\ &s.t. \ \left\| E_s\boldsymbol{X} - \boldsymbol{D}_{ST}\boldsymbol{\alpha}_s \right\| \le \varepsilon \quad \forall s \end{aligned} \tag{6}$$

C. 4D-CBCT Reconstruction via a joint of Prior Constraint and Spatiotemporal Dictionary Learning

The DL based method is not always effective in recovering some finer structures, especially in the images those are deteriorated by structural artifacts due to the under-sampled projection views. We found that in the workflow of 4D-CBCT reconstruction, a prior image could be reconstructed by using the fully-sampled projections without phase binning procedure. The prior image features a nice quality in the static region, although it gets blurred in the small region of the diaphragm. We assume that the reconstructed images are expected to be similar to the prior image. Taking this property into the reconstruction framework, the proposed method associates the prior knowledge with the spatiotemporal dictionary learning method in order to keep the balance of spatiotemporal resolution and detailed structures preservation. Our proposed optimization problem could be summarized as:

$$\begin{aligned} &\boldsymbol{X}^{opt} = \arg\min_{\boldsymbol{X},\boldsymbol{\alpha}_s} \left\| \boldsymbol{AX} - \boldsymbol{P} \right\|_2^2 + \beta_1 \left\| \boldsymbol{X} - \boldsymbol{X}_{prior} \right\|_{TV} + \beta_2 \sum_s \left\| \boldsymbol{\alpha}_s \right\|_0 \\ &s.t. \ \left\| E_s\boldsymbol{X} - \boldsymbol{D}_{ST}\boldsymbol{\alpha}_s \right\| \le \varepsilon \quad \forall s \end{aligned} \tag{7}$$

where $\boldsymbol{X}_{prior}$ denotes the prior image and $\|\cdot\|_{TV}$ calculates the total variation of the subtraction of a reconstructed sequence and a temporal extension of the prior image. In Eq.(7), dealing with the prior constraint, a classical TV minimization method[13] is adopted. To optimize the two variables $\boldsymbol{X}$ and $\boldsymbol{\alpha}_s$ in spatiotemporal dictionary learning, an alternative minimization scheme is used to split it into two sub-problems, including the update of the reconstruction image with a fixed spars representation and a better sparse representation by OMP algorithm[14]. Please refer to [10] for detailed algorithm

implementation.

III. Experiments

A. Dictionary Learning

In this study, a global ST-dictionary was trained by a set of high-quality 4D-CT images, which is obtained from the 2019 AAPM 4D-CBCT challenge[15]. The images were acquired at different respiratory phases within a breathing cycle, showing obvious motion movements trajectory at the thorax region. Then, the spatiotemporal patches were extracted from the 4D-CT images with a size of $8\times8\times10$ and the direct current component was removed from each patch. Finally, the ST-dictionary of 500 atoms was constructed, with a fixed sparse coding error of 10^{-4}. The sequence of the 4D-CT images and the trained ST-dictionary are shown in Fig. 1.

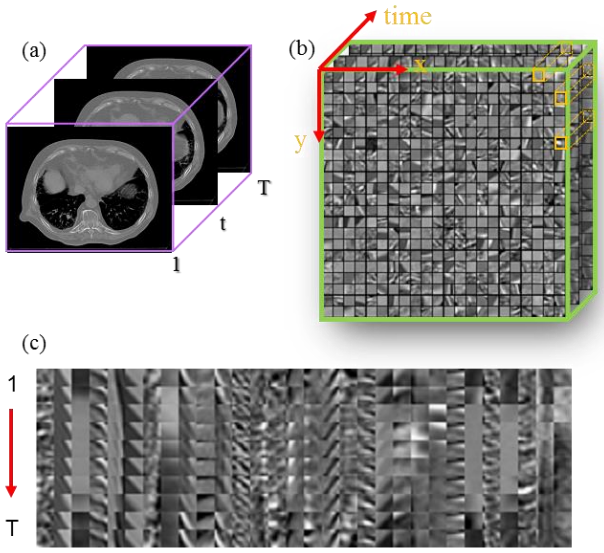


Fig. 1 Construction of a spatiotemporal dictionary. (a). A sequence of high-quality 4D-CT images is used to extract the training patches. (b) The trained global spatiotemporal dictionary with 500 atoms. The yellow cuboid is selected as one spatiotemporal atom. (c) Twenty representative spatiotemporal atoms are enlarged from (b), in which each column is an atom. The display window is [-0.15 0.15] in attenuation value.

B. Experimental Data

TABLE I
SCAN AND RECONSTRUCTION PARAMETERS OF XCAT AND PATIENT DATASET

Configuration	XCAT Phantom	Patient (On-Board imager™)
Source-to- Isocenter Distance	1000 mm	1000 mm
Source-to-Detector Distance	1500 mm	1500 mm
Detector Dimension	768	1024
Detector Voxel Size	1 mm	0.388 mm
Reconstruction Size	256*256*200	512*512*210
Projection Number	600	837

In this study, the XCAT phantom and patient dataset were used to evaluate the performance of the proposed method. There are ten phase-resolved images obtained with a breathing cycle of 5 seconds. To implement the algorithm, we only consider a two-dimensional image in each phase, and the sequence is three dimensional. The detailed parameters are listed in TABLE I.

C. Evaluation Metrics

To show the ability of the proposed method fairly, we compared our algorithm with existing methods, including the sparse constraint in terms of total variation (TV) minimization based method, PICCS method. The quantitative evaluation metrics in this study are the root mean square error (RMSE) and structural similarity index measure (SSIM), respectively.

IV. Results

A. Simulation Data

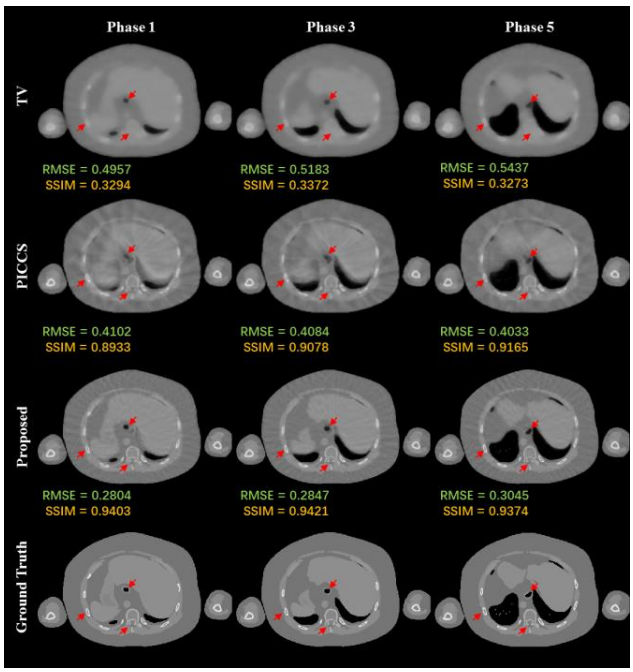


Fig. 2 The 4D-CBCT reconstructions of XCAT phantom. From top to bottom are TV minimization method, PICCS method, the proposed method, and the ground truth image. Three phase-resolved images at phase 1, phase 3, and phase 5, respectively. All the images are displayed at a display window of [0.005 0.025] in attenuation value (/mm).

Fig.2 compares the results at three respiratory phases by the TV regularized method, PICCS method, and our method, respectively. It can be seen that all the methods reduce the streaking artifacts to some extent. In detail, TV results get blocky and lose some detailed structures due to the limitation of TV. The result of PICCS could recover more detailed features but it still remains some residual streaks. With a joint of STDL and a prior constraint, the proposed method shows a better balance of achieving high spatiotemporal resolution and preserving distinct features simultaneously, as marked by the red arrows at the bone and the spine position.

TABLE II
AVERAGED RMSE AND SSIM OF DIFFERENT ALGORITHMS

Metric	TV	STDL	Proposed Method
Averaged RMSE	0.5281	0.4091	0.2941
Averaged SSIM	0.3278	0.9230	0.9397

Additionally, Fig.2 also reports the RMSE and SSIM value of three phases by various algorithms, respectively. The averaged RMSE and the averaged SSIM value of 10 phases are displayed in TABLE II. As shown in TABLE II, The result of TV method has the lowest SSIM and biggest RMSE. In contrast, the proposed method improves SSIM value by ~2.87 over the TV results. The proposed method also reduces ~55% compared with the TV method in terms of the RMSE value.

B. Patient Data

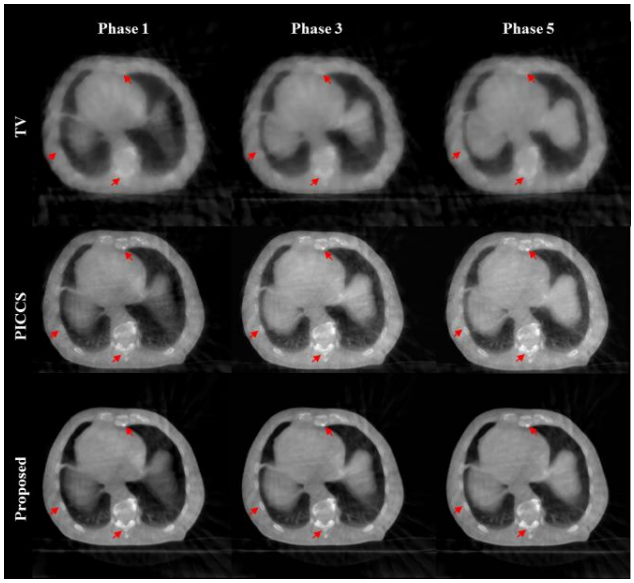


Fig. 3 The 4D-CBCT reconstructions of real patient data from VARIAN system. From top to bottom are TV minimization method, PICCS method, and the proposed method. Three phase-resolved images at phase 1, phase 3, and phase 5, respectively. All the images are displayed at the display window of [0.005 0.035] in attenuation value (/mm).

The real patient reconstructions are shown in Fig. 3. It is observed that the streaking artifacts are suppressed to effectively by the proposed method. At the same time, better image resolution is achieved at the red arrows marked region, which contains distinct features. Compared with the PICCS, the proposed method outperforms in the regions marked by red arrows in the reconstruction.

V. Conclusion

In this study, we proposed a novel reconstruction framework for 4D-CBCT. To fully take advantage of the high correlation between each phase-resolved image, a joint of spatiotemporal dictionary learning and a prior constraint is incorporated into the 4D-CBCT iterative framework. Experiments demonstrated that the proposed method could achieve both high spatiotemporal resolution and maintain detailed features.

References

[1] L. Xing *et al.*, "Overview of image-guided radiation therapy," *MedicalDosimetry*, vol. 31, no. 2, pp. 91–112, Jun. 2006, doi: 10.1016/j.meddos.2005.12.004.

[2] H. Asif, Hugo Geoffrey D., Kestin Larry L., Grills Inga S., and Y. Di, "Dosimetric advantages of four-dimensional adaptive image-guided radiotherapy for lung tumors using online cone-beam computed tomography," *International Journal of Radiation Oncology, Biology, Physics*, vol. 70, no. 2, pp. 582–589, 2008.

[3] J.-J. Sonke, L. Zijp, P. Remeijer, and M. van Herk, "Respiratory correlated cone beam CT: Respiratory correlated cone beam CT," *Medical Physics*, vol. 32, no. 4, pp. 1176–1186, Mar. 2005, doi: 10.1118/1.1869074.

[4] L. Ritschl, S. Sawall, M. Knaup, A. Hess, and M. Kachelrieß, "Iterative 4D cardiac micro-CT image reconstruction using an adaptive spatio-temporal sparsity prior," *Physics in Medicine and Biology*, vol. 57, no. 6, pp. 1517–1525, Mar. 2012, doi: 10.1088/0031-9155/57/6/1517.

[5] X. Jia, Z. Tian, Y. Lou, J.-J. Sonke, and S. B. Jiang, "Four-dimensional cone beam CT reconstruction and enhancement using a temporal nonlocal means method: Temporal nonlocal means for 4D-CBCT reconstruction," *Medical Physics*, vol. 39, no. 9, pp. 5592–5602, Aug. 2012, doi: 10.1118/1.4745559.

[6] H. Zhang, J. Ma, Z. Bian, D. Zeng, Q. Feng, and W. Chen, "High quality 4D cone-beam CT reconstruction using motion-compensated total variation regularization," *Phys. Med. Biol.*, vol. 62, no. 8, pp. 3313–3329, Mar. 2017, doi: 10.1088/1361-6560/aa6128.

[7] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets: Prior image constrained compressed sensing (PICCS)," *Medical Physics*, vol. 35, no. 2, pp. 660–663, Jan. 2008, doi: 10.1118/1.2836423.

[8] "Machine Learning for Tomographic Imaging - Book - IOPscience." https://iopscience.iop.org/book/978-0-7503-2216-4 (accessed Jan. 19, 2020).

[9] Qiong Xu, Hengyong Yu, Xuanqin Mou, Lei Zhang, Jiang Hsieh, and Ge Wang, "Low-Dose X-ray CT Reconstruction via Dictionary Learning," *IEEE Trans. Med. Imaging*, vol. 31, no. 9, pp. 1682–1697, Sep. 2012, doi: 10.1109/TMI.2012.2195669.

[10] T. Bai, H. Yan, X. Jia, S. Jiang, G. Wang, and X. Mou, "Z-Index Parameterization for Volumetric CT Image Reconstruction via 3-D Dictionary Learning," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2466–2478, Dec. 2017, doi: 10.1109/TMI.2017.2759819.

[11] Y. Zhang, X. Mou, G. Wang, and H. Yu, "Tensor-Based Dictionary Learning for Spectral CT Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 142–154, Jan. 2017, doi: 10.1109/TMI.2016.2600249.

[12] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online Dictionary Learning for Sparse Coding," in *Proceedings of the 26th Annual International Conference on Machine Learning*, New York, NY, USA, 2009, pp. 689–696, doi: 10.1145/1553374.1553463.

[13] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine & Biology*, vol. 53, no. 17, p. 4777, 2008.

[14] R. Rubinstein, M. Zibulevsky, and M. Elad, *Efficient Implementation of the K-SVD Algorithm using Batch Orthogonal Matching Pursuit*. .

[15] C. C. Shieh *et al.*, "SPARE: Sparse-view reconstruction challenge for 4D cone-beam CT from a 1-min scan.," *Med Phys*, Jun. 2019, doi: 10.1002/mp.13687.

Overcoming truncation artifacts caused by the patient table in polyenergetic statistical reconstruction on clinical C-arm CT data

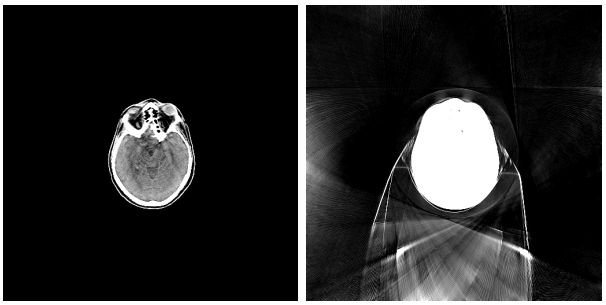
Richard N.K. Bismark, Oliver Beuing, and Georg Rose,

***Abstract*—In previous work, we applied the polyenergetic statistical reconstruction (PSR) technique by A. Elbakri and J. A. Fessler [1] in order to reduce beam hardening artifacts in C-arm CT data. As we reported in 2019 [2], the PSR suffers from the patient table that is partially truncated in the projection space resulting in severe artifacts. Such truncation artifacts are not observed when using other reconstruction methods. To overcome this problem, we propose to reconstruct the outer volume (OV) by an under sampled PSR with rougly a few iterations and a fraction of the measured projection data. Afterwards, we use the resulting OV within the polychromatic forward projection step of the PSR to account for the truncated objects outside of the area of the reconstruction. We demonstrate the feasibility of our proposed adaptation on three clinical data sets of a C-arm CT system and compare those results with a maximum likelihood reconstruction that is using a different beam hardening correction. We can state, that the additional computational overhead is not dramatically increasing the reconstruction time and the PSR becomes a viable tool for clinical diagnosis as it can overcome the truncation artifacts that are caused by the patient table and reduce beam hardening induced cupping and spill over artifacts as well.**

***Index Terms*—Beam hardening, C-arm CT, Cone beam, Polychromatic statistical reconstruction, Truncation, Patien table.**

I. Introduction

COMPUTED tomography is a widely used technique for patient scans in the clinical environment. Typically the patient lies on a table that is designed to not cause artifacts in the reconstruction. The table is truncated in the projection space. Thus, it is not fully visible in every X-ray image. As Bismark et al. [2], [3] reported, the polychromatic statistical reconstruction (PSR) technique suffers heavily when introduced to this kind of truncation setting. The PSR was first described by Elbakri and Fessler [1], [4] in the early 2000's and can be used to compensate beam hardening and scatter artifacts in CT images. Common methods to reduce beam hardening artifacts are based either on pre-processing of the measured data, like a water correction [5], [6] or a post image processing [7] of the reconstructed image. The PSR is based on the idea that a more accurate model of the X-ray absorption process – taking into account the energy dependend attenuation – can help to overcome beam hardening. Similar methods which also includes a more sophisticated forward projection models can be found in [10] or [11]. The characteristic beam hardening artifacts such as spill over, streaks and cupping are introduced to the common reconstructions due to their simplifying assumptions of the forward projection model of measured X-ray image. However, in order to make the PSR a viable tool for clinical diagnostic, it needs to be reliable in a typical clinical setting such as a truncated patient table of a C-arm system. In this paper, we propose a solution to this problem by tweaking the forward projection step to incorporate an outer volume (OV) that is an appropriate estimation of the structures that are truncated in the projection space. We show the viability of this method on three clinical head scan data sets recorded on a flat panel C-arm CT with respect to the patient table truncation.



(a) (b)

Fig. 1. One slice of a PSR image (512×512 with $2.0 \times 2.0\,\mathrm{mm}^2$) of a typical head scan with different window levels. The regularization was turned off for the sake of clarity. The window level (a) lies in interval of 1.01 to $1.15\ \mathrm{g/cm^3}$, (b) of 0.0 to $0.3\ \mathrm{g/cm^3}$. The pixel values of the outer volume that are drastically above the $0.0\ \mathrm{g/cm^3}$ correspond to the patient table. A low resolution PSR with a high smoothing (e.g. by increasing the regularization parameter) of such a an image (b) is used as prior knowledge in the proposed PSRov method.

II. Methods

The polychromatic statistical reconstruction technique (PSR) is maximizing a *Poisson* likelihood function that includes a more accurate forward projection model. For the sake of brevity, we refer to Elbakri and Fessler [1], [4] for the theoretical background and Bismark et al.[3] for in-depth information of the implementation. The iterations revolve around the reconstruction volume that represents the patient. Common CT reconstruction techniques represent the volume in terms of the linear attenuation coefficient $\mu(E, x, y, z)$. The PSR separates the energy dependence from the spacial domain

R. Bismark and G. Rose are with the Institute for Medical Engineering and Research Campus *STIMULATE*, University of Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany, richard.bismark@ovgu.de

Oliver Beuing was with the University Hospital Magdeburg, Institute for Neuroradiology, Germany

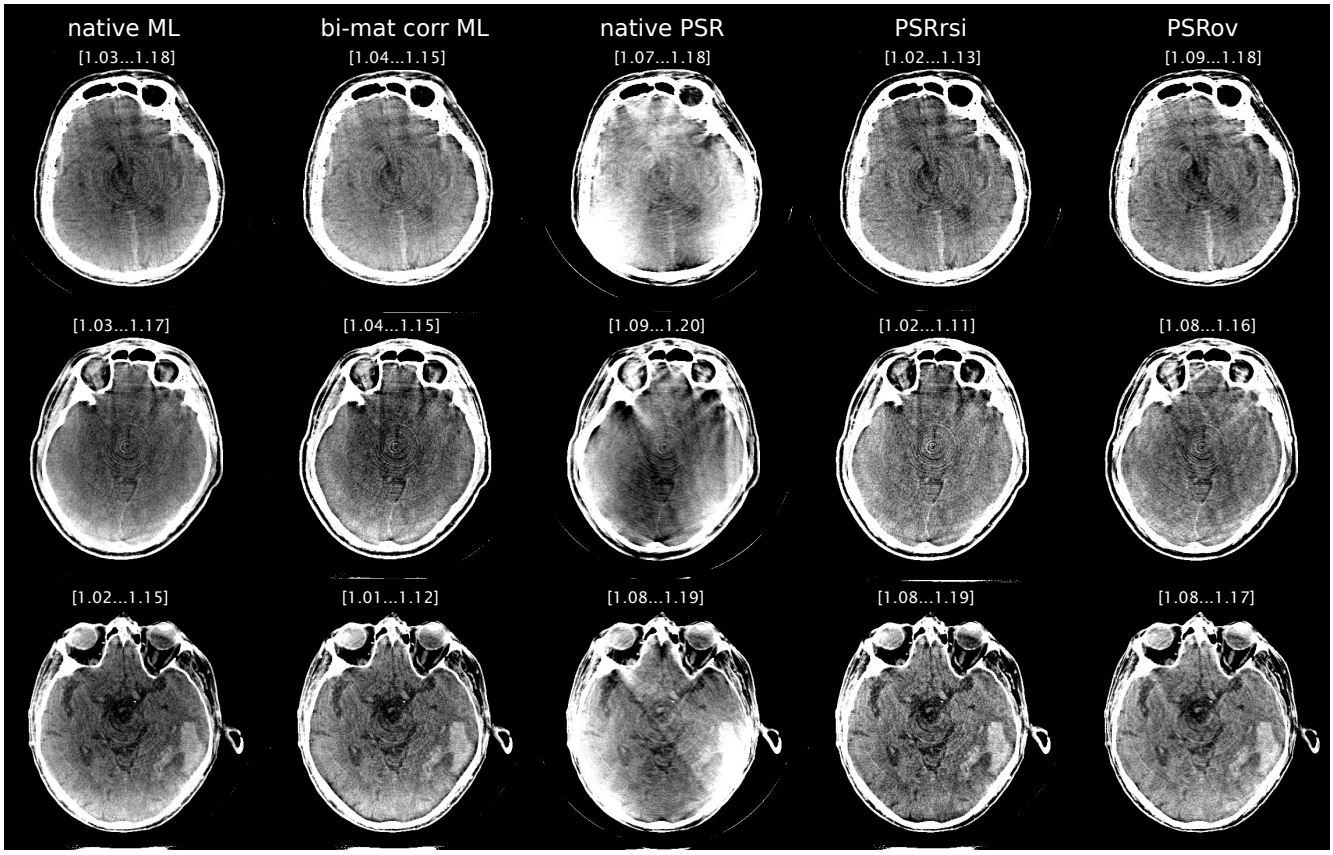


Fig. 2. Slices of reconstructions (464×464 pixels with $0.5 \times 0.5\,\mathrm{mm}^2$) of three different clinical datasets. The native ML suffers from beam hardening induced cupping and spill over from the bones into the tissue area. The bi-material corrected projections help to reduce the spill over but cannot fully reduce cupping. The native PSR is corrupted by the typical table truncation artifact. The PSRrsi yields a better reduction of beam hardening artifacts then the bi-material corrected ML. The proposed solution of PSRov gives a similar result. Note, there is no correction for the C-arm typical ring artifacts. All images show density values in $\mathrm{g/cm}^3$.

by introducing the product $\mu(E,x,y,z) = m(E)\rho(x,y,z)$. Thus, the iterative update step changes the density ρ in order to maximize the likelihood function. In this work we add a relaxation parameter $R \leq 1$ to the update step:

$$\rho_j^{n+1} = \left[\rho_j^n - R\left(\mathrm{D}^{-1}\nabla Q\left(\rho_j^n\right)\right)\right]_+ \tag{1}$$

$$d_j \propto \sum_i a_{ij}\gamma_i Y_i. \tag{2}$$

Here, $\nabla Q\left(\rho_j^n\right)$ is the gradient of a surrogate function Q of a Poisson likelihood cost function, $\mathrm{D} \propto \mathrm{diag}\{d_j\}$ is a weighting matrix, a_{ij} are the geometry factors, γ_i is the length of i-th ray through the volume, and Y_i is the intensity value measured by the i-th detector pixel. Q contains a polychromatic forward projection of the volume represented by the current iterate ρ_j^n. Note that the index j represents the voxel elements. Earlier studies [2] suggested that the patient table in the used head scan data corrupts the PSR results while comparable iterative ML [8] reconstructions of the same data sets are not affected in such a manner. Figure 1 is depicting the reconstruction of the outer volume explicitly. The reason for such a weakness of the PSR lies in the specific choice of the weighting matrix D. The iterative ML reconstruction is not compensating for beam hardening by its definition but uses state-of-the-art bi-material beam hardening corrected projections [9]. Both projection data sets, the uncorrected[1] projections and the bi-material corrected projections, are used with the ML method in order to have a good comparison to the different PSR adaptations presented in this paper.

III. Materials

We compare the PSR with an iterative ML reconstruction that is neglecting the polyenergetic nature of the X-ray propagation and measurements which lead to beam hardening artifacts. If we feed beam hardening corrected projections [9] into this ML routine, we obtain the bi-material corrected ML. Initializing the PSR [4] with zeros and iterate eq. (1) 40 times with $R = 1$ results in the "native PSR". In fig. 2 the table truncation artifact can be observed as the native PSR is suffering from the characteristic artifacts described in [2].

To reduce this artifact within the frame work of the PSR, we account for the truncation directly in the projection space. The outer volume (OV) contains the table as we basically doubled the physical size of the reconstruction volume and perform the native PSR algorithm. Fig. 1 visualizes that idea. This information gain can then be utilized to reduce the

[1]This means that those projections are not even water corrected.

inconsistencies between the forward projection of the current iterate and the measured projections.

In the first step, our proposed method under samples the OV by 8 iterations of the PSR (initialized with zeros) that is using larger but less voxels. Also we use only a fraction of the provided projection data. Here we used $128 \times 128 \times 178$ voxels with a spacial resolution of $4.0 \times 4.0 \times 1.0\,\mathrm{mm}^3$ instead of $464 \times 464 \times 178$ voxels with a spacial resolution of $0.5 \times 0.5 \times 1.0\,\mathrm{mm}^3$ and 248 views of 1240×960 detector pixels instead of 496 views while keeping the detector size[2]. To make sure the under sampled reconstruction of the OV does not suffer from streak artifacts due to lower angular sampling, we increase the regularization parameter by a factor of 100.

In the second step, we start the PSR routine as intended but give the forward projector (a polychromatic ray caster) the under sampled but larger volume in addition to the current iterate ρ_j^n. This means in particular that the OV is constant and cannot be updated by the iteration represented by equation (1). Since we did not change the very nature of the original algorithm we call this routine PSRov.

Note that, the down sampling of the OV leads to a smoother approximation and reduces reconstruction time for the OV and the PSRov as well. The forward projection of the additional OV in the PSRov update step increases the forward projection time just moderately due to the interpolation feature of the GPU and the down sampling[3] itself. This method is an extension of the work of Ziegler et al. [12].

Another important fact about our projection data is that we were cutting out the outer 30 pixels[4] by ignoring the data points. This is an easy way to reduce artifacts that are caused by the collimator.

Additionally, we performed the work around proposed in [2] of regularization by "early stopping". We initialize the PSR with the native ML and perform only one iteration of equation (1) with $R = 0.05$ including all X-ray images once.

IV. Results

Figure 2 shows different reconstructions of three clinical data sets including truncated projections of the patient table. The uncorrected ML suffers from the beam hardening induced cupping and the typical spill over effect which blurs the edge of the dense bone of the skull into the brain area. The bi-material corrected ML reduces that spill over and but cannot fully suppress the cupping. The table truncation artifact overshadows the image produced by the native PSR. In [2] we found by simulations that the truncation of the patient table is causing such an impairment. The two solutions to that problem (PSRrsi and PSRov) overcome that erroneous induction and yield similar beam hardening reduction as the bi-material corrected ML.

[2]those projections where not water corrected but only normalized and pre-processed by a scatter correction routine provided by the C-arm device: Artis Q biplane angiography system, *Siemens* Healthcare GmbH, Erlangen, Germany.

[3]The voxel edge length increases by a factor of 8.

[4]We go down from 1240×960 detector pixels to 1180×900 detector pixels and changing the projection matrices accordingly to this new geometry.

V. Conclusion

The idea of Elbakri and Fessler [4] to overcome the problem of beam hardening by simply modelling the non-linearity of the X-ray absorption process and its detection by integrating detectors into the mathematical framework of iterative statistical reconstruction can be observed in figure 2. For the first time, the PSRov is reducing beam hardening induced artifacts on different clinical data sets that include truncation of the patient table while beeing independend of the initialization. The PSRov does not require any new CT hardware or certain scanning trajectories (this applies to all iterative reconstruction techniques). Thus, it can potentially enhance clinical diagnostics on ordinary CT and C-arm systems that are used in hospitals all around the globe.

Since clinical CT scans include additional objects like a patient table, those objects are designed to cause the least additional extinction values to projections. Typical reconstruction techniques tend to increase voxel values around the reconstruction object in order to minimize the inconsistency between the forward projection while iterating. This behavior can be observed in the first column of fig. 2. Whereas the reconstructions have increased pixel values at the bottom of the images or a thin "ring" structure around the head. The best visible example of such a thin "ring" structure can be observed in the native PSR. Since such artifacts are caused by the truncation of the table in the projections and not by the collimator, this observation is consistent with the fact, that the thin "ring" artifact around the head and the increased pixel values at the bottom are vanishing in the PSRov which is taking the truncation into its forward projection model.

We observed, that the reconstruction time is not increased dramatically by the additional forward projecting in each iteration step. A detailed performance evaluation of our C++ and openCL multi-GPU implementation[5] needs to be done and is considered to be future work.

In order to have an evaluation study about the diagnostic value, additional data sets have to be reconstructed by different state-of-the-art methods and compared by physicians, which is considered to be future work as well.

Acknowledgment

The work of this paper is partly funded by the German Ministry of Education and Research (BMBF) within the Forschungscampus *STIMULATE* under grant number '13GW0095A'.

References

[1] I. Elbakri and J. Fessler, *Statistical image reconstruction for polyenergetic X-Ray computed tomography*, IEEE Transactions on Medical Imaging, 89–99 (2002).

[2] R. Bismark, O. Beuing and G. Rose, *Truncation artifacts caused by the patient table in polyenergetic statistical reconstruction on real C-arm CT data*, in The 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine 2019.

[5]We were using 1 to 4 AMD FirePro S10000 graphics cards each with two 'Tahiti' chips (bought 2013).

[3] R. Bismark, Robert Frysch, Shiras Abdurahman, Oliver Beuing, Manuel Blessing and Georg Rose, *Reduction of beam hardening artifacts on real C-arm CT data using polychromatic statistical image reconstruction*, in Zeitschrift für Medizinische Physik, 0939-3889 , 2019, URL: https://doi.org/10.1016/j.zemedi.2019.10.002

[4] I. Elbakri and J. Fessler, *Segmentation-free statistical image reconstruction for polyen- ergetic X-Ray computed tomography with experimental validation*, in Physics in Medicine and Biology 48 (15), 2453 (2003)

[5] O. Nalcioglu and R. Lou, *Post-reconstruction method for beam hardening in computerised tomography*, in Physics in Medicine and Biology 2 (24), p. 330, 1979

[6] Peter. Joseph and R. Spital, *A Method for Correcting Bone Induced Artifacts in Computed Tomography Scanners*, in Journal of Computer Assisted Tomography 2 (1), .p. 100-108, 1978

[7] M. Zellerhoff, B. Scholz, E. Ruehrnschopf and T. Brunner, *Low contrast 3D reconstruction from C-arm data*, in Proc. SPIE 5745, p. 646-655, 2005

[8] R. Frysch, T. Pfeiffer, S. Bannasch, et al., *C-arm perfusion imaging with a fast penalized maximum-likelihood approach*, 2014

[9] S. Abdurahman, R. Frysch, R. Bismark, M. Friebe and G. Rose, *Calibration free beam hardening correction using grangeat-based consistency measure*, in IEEE NSS/MIC 2016, 2016.

[10] S. Xu, A. Uneri, A. Jay Khanna, J. Siewerdsen and J. Stayman, *Polyenergetic known-component CT reconstruction with unknown material compositions and unknown x-ray spectra*, in Physics in Medicine and Biology 62 (8), 3352-3374, (2017)

[11] T. Humphries, J. Winn and A. Faridani, *Superiorized algorithm for reconstruction of CT images from sparse-view and limited-angle polyenergetic data*, in Physics in Medicine and Biology 62 (16),6762-6783, 2017.

[12] A. Ziegler, T. Nielsen M. Grass, *Iterative reconstruction of a region of interest for transmission tomography* , in Medical Physics 35 ,1317-1327, 2008.

Fused analytical iterative reconstruction for sparse view dual energy computed tomography

Mengfei Li, Hao Gao, and Yunsong Zhao*

***Abstract*—in this abstract, we study image reconstructions for sparse view DECT (Dual Energy Computed Tomography). Specifically, we extend our FAIR (Fused Analytical Iterative Reconstruction) imaging framework, which was originally proposed for traditional CT, to DECT. The extension is achieved by replacing the FBP operator used in the original FAIR framework with DE-FBP (Dual Energy-FBP) operator, which was proposed by us for DECT image reconstruction. In addition, regularizers are applied on both the basis material images. Numerical experiments verify the effectiveness of the proposed framework. It is shown that, the proposed framework can obtain satisfactory results from 60 equally sampled projections.**

***Index Terms*—dual energy CT, fused analytical iterative reconstruction method, iterative reconstruction method, material decomposition.**

I. Introduction

Since the invention of x-ray CT (Computed Tomography), it has been extensively used in both medical area and industrial area. But it is well known that the ionizing x-ray radiation could increase cancer risks. Therefore, reduction of the radiation dose is gaining more and more attention in the medical area. There are two main ways to reduce radiation dose. One is reducing the x-ray flux for each projection while maintaining the number of views and the other is reducing the view number while keeping the SNR (Signal to noise ratio) of each projection. In this abstract, we focus on the latter way which also has potential applications in a stationary CT architecture. In addition, the spectral information of the x-ray is also considered, i.e., in this abstract, we study image reconstruction from sparsely sampled projections of a DECT (Dual Energy CT).

In DECT, the object is scanned with two different x-ray spectra. So compared with conventional CT, more information of the scanned object is obtained, which can be used to reconstruct material-selective images or energy-selective images [1,2] and can further be used to material identification.

Generally, image reconstruction from sparsely sampled projections is changing because of the lack of sufficient data, which also makes the problem ill-posed. To overcome this problem, a common approach is to add prior information of the scanned object into the imaging model. After years of study, researchers have proposed several prior models, among which, we think, the sparsity of the image in image domain or some transformation domain should be the most popular and successful one. According to this assumption, different specific models are proposed, including TV (Total Variation) model [3], tight-frame model [4], dictionary learning model [5], to name a few. In this abstract, we take TV model as an example to illustrate our imaging framework.

The imaging framework proposed in this abstract is an extension of our FAIR (Fused Analytical Iterative Reconstruction) imaging framework [6], which was originally propose for traditional CT. The FAIR framework is an iterative framework for solving an optimization problem which includes two terms in its objection function, a data fidelity term and a regularization term. The FAIR solves the problem by making use of the data fidelity term and the regularization term in an alternating manner to update the estimated images. In the FAIR, the estimated image is updated according to the data fidelity term by adding the residual image reconstructed from the residual projections with the FBP operator. To extend FAIR to DECT, we replace the FBP operator with our DE-FBP (Dual Energy-FBP) operator [7], which updates the basis material images according to the two sets of polychromatic projections. In addition, regularizers are applied on both the basis material images. Due to the lack of convex property of the objective function for DECT imaging, we cannot prove the convergence of the proposed framework theoretically. But numerical experiments have verified its effectiveness. It is shown that, the proposed framework can obtain satisfactory images from equally sampled 60 projections.

II. Materials and Methods

A. CT imaging model

Generally, image reconstruction problem can be modeled as an operator equation:

$$Ax = y \tag{1}$$

For traditional CT, A is the Radon transform, y is the projections or line integrals of the image x. While for DECT, A is the polychromatic projection operator corresponding to

This work was supported in part by the Beijing Education Committee, and the National Natural Science Foundation of China (grant numbers 61771324, 61901127, 61827809, and 61871275). (Corresponding author: Yunsong Zhao.)

M. Li is with the Division of Ionizing Radiation, National Institute of Metrology, Beijing 100029, China and with Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, 100048, China (e-mail: lmfly@cnu.edu.cn).

H. Gao is with Department of Radiation Oncology, Emory University, Atlanta, GA 30322 USA (e-mail: hgao24@emory.edu).

*Y. Zhao is with the school of Mathematical Sciences, Capital Normal University, Beijing, 100048, China, and with Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, 100048, China, and with Department of Radiation Oncology, Emory University, Atlanta, GA 30322 USA (e-mail: zhao_yunsong@cnu.edu.cn).

two different x-ray spectra and y is the obtained polychromatic projections, $x = (f; g)$, where f and g are two basis material images to be reconstructed. Specifically, along a given x-ray path L, the polychromatic projection for the normalized x-ray spectrum $S(E)$ is

$$p(L) = -\ln \int_E S(E) e^{-\phi(E)F - \theta(E)G} dE$$

where

$$F = \int_L f dl \,, G = \int_L g dl$$

are the line integrals of the basis material images f and g along the x-ray path L; $\phi(E)$ and $\theta(E)$ are the mass attenuation coefficient of the two basis materials.

When we have sufficient projections, we can solve Eq. (1) directly with classic reconstruction methods to get satisfactory reconstruction results x, such as FBP or ART for traditional CT, or their variants for DECT, such as DE-FBP [7] and EART (Extended ART) [8]. But when the projections are not sufficient, e.g., in sparse view imaging, we usually convert the image reconstruction problem into an optimization problem and add some extra regularizer to the objective function. The general form of the problem is then as follows,

$$x^* = \arg\min_x \frac{1}{2} \|Ax - y\|_2^2 + \lambda R(x)$$

where $\|Ax - y\|_2^2$ is the data fidelity term and $R(x)$ is the regularizer, λ is a regularization parameter or the model parameter to balance data fidelity and image regularization. There are several regularizers that are proposed in literatures such as TV regularizer, image gradient L0 norm regularizer, wavelet domain regularizer, to name a few. In the following of this abstract, we will take the anisotropic TV regularizer as an example for the discussion and the experimentation, there should be no difficult to replace it with other regularizer. At this time, we have

$$R(x) = \|\nabla x\|_1 = \sum_{ij} \left|x_{i+1,j} - x_{i,j}\right| + \left|x_{i,j+1} - x_{i,j}\right|$$

for traditional CT and

$$R(x) = \|\nabla x\|_1 = \sum_{ij} \left|f_{i+1,j} - f_{i,j}\right| + \left|f_{i,j+1} - f_{i,j}\right| + \left|g_{i+1,j} - g_{i,j}\right| + \left|g_{i,j+1} - g_{i,j}\right|$$

for DECT.

B. Fused analytical and iterative reconstruction

Fused analytical and iterative reconstruction is a derived from the so-called proximal gradient method. For problem (1), the corresponding iteration scheme is as follows:

$$\begin{aligned} x^{n+\frac{1}{2}} &= x^n - s^n A^\top (Ax^n - y) \\ x^{n+1} &= \arg\min_x \frac{1}{2} \left\| x - x^{n+\frac{1}{2}} \right\|_2^2 + s^n \lambda \|\nabla x\|_1 \end{aligned}$$

The first step of the above scheme is gradient descend for data fidelity term, where $A^\top$ is the transpose of A. The second step of the scheme is to smooth the estimated image $x^{n+1/2}$ obtained from the first step with the regularization term. The term $\frac{1}{2} \left\| x - x^{n+\frac{1}{2}} \right\|_2^2$ can also be regarded as a regularizer which makes the objective function to have a better property, such as convexity. When $\frac{1}{2}\|Ax - y\|_2^2$ is convex, it can be shown that the proximal gradient method is converge under some assumptions on the relaxation parameter s^n.

We know that, for CT imaging, the convergence speed of this scheme is slow due to the inefficiency of the gradient descent method in solving the data fidelity term. So, to improve the convergence speed, we replace $A^\top$ with the filtered back-projection operator F and obtains the FAIR iteration scheme:

$$\begin{aligned} x^{n+\frac{1}{2}} &= x^n - s^n F(Ax^n - y) \\ x^{n+1} &= \arg\min_x \frac{1}{2} \left\| x - x^{n+\frac{1}{2}} \right\|_2^2 + s^n \lambda \|\nabla x\|_1 \end{aligned} \quad (2)$$

In this case, the formula can be interpreted as we add the residual image reconstructed from residual projections $Ax^n - y$ onto current estimation to obtain a new estimation. In Ref. [6], we have also shown that the scheme still converges when we replace $A^\top$ with F.

C. Fused analytical and iterative reconstruction for DECT

The extension of FAIR method to DECT is based on our previous work [7], which is proposed for DECT imaging. In this abstract, we call it DE-FBP, for the emphasize that it is used for DECT imaging. The iteration scheme of DE-FBP is

$$\begin{aligned} f^{n+1} &= f^n + \Delta f_q^n + \Delta f_q^n \\ g^{n+1} &= g^n + \Delta g_q^n + \Delta g_q^n \end{aligned}$$

where $\Delta f_q^n, \Delta f_q^n, \Delta g_q^n, \Delta g_q^n$ are residual images reconstructed from weighted residual projections, the subscript p, q indicates two different x-ray spectra. It is shown that the DE-FBP method is very efficient, 2 or 3 iterations are enough to get satisfactory reconstructions. For detail information, including how to weight the residual projections, please refer to [7].

Considering that the first step of AIR is just to update estimated images with projections and that the DE-FBP method for DECT is an efficient method for images update from dual energy projections, we replace the F in Eq. (2) with the DE-FBP operator obtains the DE-FAIR iteration scheme for DECT

$$\begin{aligned} f^{n+\frac{1}{2}} &= f^n + \Delta f_q^n + \Delta f_q^n \\ g^{n+\frac{1}{2}} &= g^n + \Delta g_q^n + \Delta g_q^n \\ f^{n+1} &= \arg\min_f \frac{1}{2} \left\| f - f^{n+\frac{1}{2}} \right\|_2^2 + s^n \lambda \|\nabla f\|_1 \\ g^{n+1} &= \arg\min_g \frac{1}{2} \left\| g - g^{n+\frac{1}{2}} \right\|_2^2 + s^n \lambda \|\nabla g\|_1 \end{aligned}$$

III. EXPERIMENTS

In this section, we demonstrative some preliminary experimental results of the proposed DE-FAIR for DECT. For simplicity, all the experiments are restricted to fan beam CT. There should be no difficulty to extend the proposed method to cone beam CT.

A. *Experiment Setups*

A slice of the 3D FORBILD thorax phantom is used in the simulation. Fig. 1 shows the image of the phantom with resolution 512×512. For detailed information about the phantom please refer to http://www.imp.uni-erlangen.de/phantoms/index.htm. The x-ray spectra used in the experiments are simulated with the free software SpectrumGUI (see https://sourceforge.net/projects/spectrumgui/). The voltages are 80 kV and 140 kV, respectively. Fig. 2 shows the normalized x-ray spectra. In the simulation, the phantom is treated as a composition of water material and bone material with different densities. Water and bone are also used as the basis materials in image reconstruction. The mass attenuation coefficients of the basis materials, i.e., water and bone, are obtained from NIST (see: https://physics.nist.gov/PhysRefData/XrayMassCoef/tab4.html).

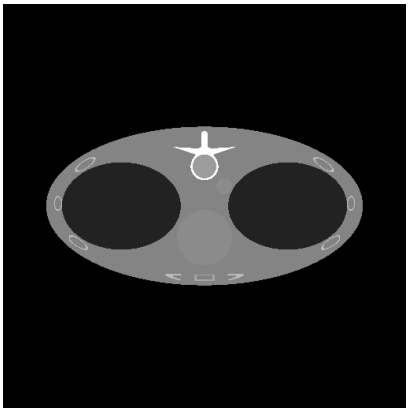


Fig. 1. The slice of the thorax phantom used in the numerical simulation

The parameters of the scanning configuration are set as follows: the focus-to-object distance (FOD) is 1000 mm; the focus-to-detector distance (FDD) is 1300 mm; the line detector includes 512 units with 0.3 mm detector cell size. The projections corresponding to the two x-ray spectra are geometrically consistent. For each x-ray spectrum, 60 projections are simulated over 2π.

To test the noise performance of the proposed method, noisy projections are also simulated by adding Poisson noises to the data. The x-ray intensity for each path is 10^6.

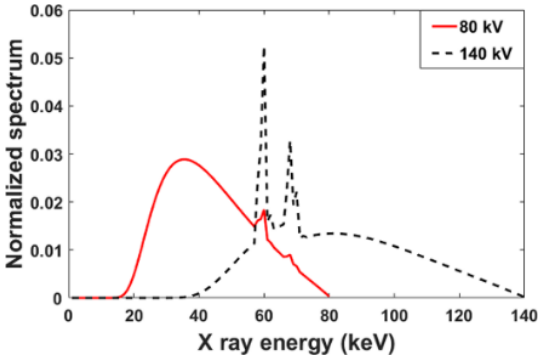


Fig. 2. The x-ray spectra used in the numerical simulation.

B. *Results*

Fig. 3 and Fig. 4 show the images reconstructed with the DE-FBP method and the proposed method (DE-FAIR) from noise-free data and noisy data. In each figure, the first row is the images obtained by DE-FBP, while the second row is the images obtained by the proposed method. The first and the second column are the images of water material and bone material, while the third column is the 70 keV monochromatic images (μ-image).

The number of iterations for all the experiments are 30. It should be noted that in the implementation of DE-FBP, a median filtering with a 3×3 kernel size is performed on the basis material images in each iteration. In DE-FAIR, the anisotropic TV model [9] is used in the image smoothing step of the proposed method and the regularizer parameter in the iterations has been tuned to try to minimize the noise and artifacts in the basis material images.

We can see from Fig.3 and Fig.4 that, with the median filter, the DE-FBP method has some effects on noise suppression. But the sparse view artifacts are obvious on the water material images and the monochromatic images. While in the images reconstructed with the proposed DE-FAIR method, both noises and sparse view artifacts are removed effectively.

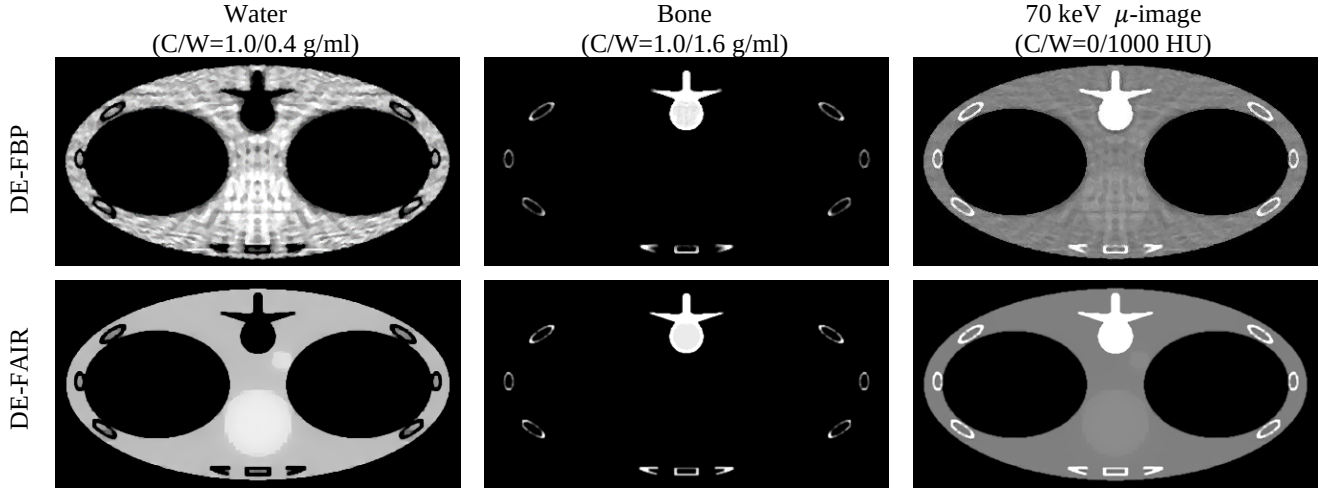


Fig. 3. The images reconstructed with the DE-FBP method and the proposed method from noise-free data.

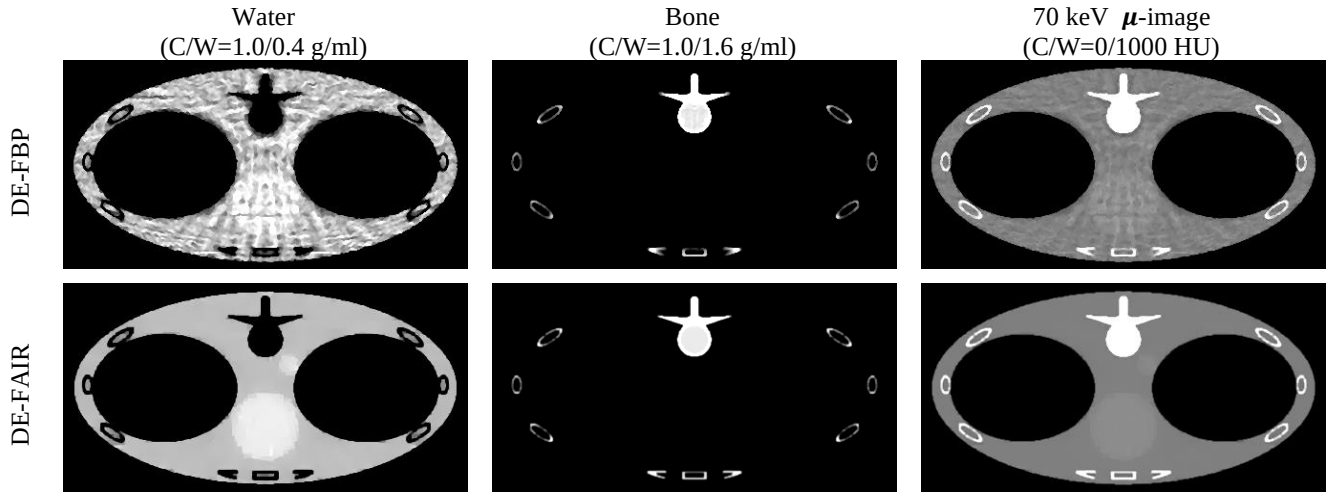


Fig. 4. The images reconstructed with the DE-FBP method and the proposed method from noisy data with intensity 10^6.

IV. Conclusion and Discussion

In this abstract, we propose an image reconstruction framework for sparse view DECT, which incorporate the DE-FBP operator and the FAIR imaging framework. Preliminary experiments verify its effectiveness. Satisfactory reconstruction results are obtained with the proposed method from equally sampled 60 projections.

In the experiments, we have used the anisotropic TV norm of the images as the regularizer for its simplicity. More sophisticated regularizers will be tested to further improve the image qualities in the future.

References

[1] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerized tomography," *Physics in medicine and biology*, vol. 21, no. 5, pp. 733–744, 1976.

[2] J. Fornaro *et al.*, "Dual- and multi-energy CT: approach to functional imaging," *Insights into imaging*, vol. 2, no. 2, pp. 149–159, 2011.

[3] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in medicine and biology*, vol. 53, no. 17, pp. 4777–4807, 2008.

[4] B. Zhao, H. Gao, H. Ding, and S. Molloi, "Tight-frame based iterative image reconstruction for spectral breast CT," *Medical physics*, vol. 40, no. 3, p. 31905, 2013.

[5] S. Tan *et al.*, "Tensor-based Dictionary Learning for Dynamic Tomographic Reconstruction," *Physics in medicine and biology*, vol. 60, no. 7, pp. 2803–2818, 2015.

[6] H. Gao, "Fused analytical and iterative reconstruction (AIR) via modified proximal forward-backward splitting: a FDK-based iterative image reconstruction example for CBCT," *Physics in medicine and biology*, vol. 61, no. 19, pp. 7187–7204, 2016.

[7] M. Li, Y. Zhao, and P. Zhang, "Accurate Iterative FBP Reconstruction Method for Material Decomposition of Dual Energy CT," *IEEE transactions on medical imaging*, vol. 38, no. 3, pp. 802–812, 2019.

[8] Y. Zhao, X. Zhao, and P. Zhang, "An extended algebraic reconstruction technique (E-ART) for dual spectral CT," *IEEE transactions on medical imaging*, vol. 34, no. 3, pp. 761–768, 2015.

[9] T. Goldstein and S. Osher, "The Split Bregman Method for L1-Regularized Problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 323–343, 2009.

Dental panoramic image reconstruction from cone-beam CT projection data

Taejin Kwon, Da-in Choi, Jaehong Hwang, Jaesung Lee, Jihwan Lim, Inje Lee, and Seungryong Cho

Abstract— **Panoramic dental x-ray imaging is widely performed by dentists in everyday practice, and is useful for diagnosis and planning of various dental treatments. Dental cone-beam CT, which provides rich image information of 3D dental anatomy and thus is considered appropriate for more demanding cases, is also a popular and important imaging modality in the platform of digital dentistry. It is often the case that dentists need a panoramic 2D image in addition to a 3D CT image. In an attempt to remove such an additional 2D panoramic scan, extracting a 2D panoramic image of the 3D volume image has been investigated. In this work, we propose a new imaging approach that reconstruct 2D panoramic image out of CT projection data instead of the CT volume image. The proposed method retains the shift-and-add reconstruction scheme of a typical panoramic dental imaging, and it has successfully produced a 2D panoramic image with a quite comparable image quality to the real panoramic radiography.**

Index Terms— **Dental panoramic radiography, Cone-beam computed tomography, image reconstruction**

I. Introduction

DIGITAL dentistry is referred to as a new paradigm of dental procedures that make use of advanced digital technologies including 2D/3D x-ray imaging, optical imaging, 3D printing, and so on[1]-[6]. Particularly, x-ray 2D panoramic dental radiography and 3D cone-beam CT are becoming dominant modalities of dental diagnosis and treatments. 2D panoramic radiography is indeed widely used in everyday practices for various dental disease diagnosis and associated treatments [7][8]. Dental cone-beam CT provides rich image information of 3D dental anatomy and thus is considered appropriate for more demanding cases [1].

Panoramic radiography system is composed of an x-ray source and a vertical line detector array mounted on a gantry. The arm connecting the source and the detector has usually a movable pivoting point around which the arm is rotating [4][9]. By design, such an imaging system acquires a number of projections along the scanning path. Stitching the projections would form a 2D radiographic image that covers entire dental anatomy including maxilla, mandible, and temporomandibular jaw (TMJ). By purposely exposing x-ray in a way that the neighboring projections would have substantial overlap, one can get depth-resolved slice images through a tomosynthetic image reconstruction process [4][9]-[12]. Out of such a set of depth-resolved slices, a sharpness detection algorithm can be utilized to pick up the sharpest slice image. This technique is often called auto-focusing [12]. As panoramic radiography is widely performed and is considered a primary imaging modality, dentists are familiar with the panoramic radiographs and their image interpretation is rather straightforward and reliable.

3D dental cone-beam CT imaging is increasingly used today for various treatments particularly including dental implant surgery [2][13]. However, as 3D CT is considered a rather recent addition to dentistry, it is often the case that dentists need a panoramic radiograph in addition to a 3D CT image. Vendors indeed provide a two-in-one commercial CT scanner that combines CT imaging with panoramic imaging. The system often comes in a versatile shape with two separate detectors mounted together: one for CT imaging and the other for panoramic radiography. This is not cost-effective and also not desirable from the patient dose perspective. Therefore, synthesizing dental panoramic radiographs out of the volumetric CT image has been investigated. Although this approach fully enjoys the 3D nature of CT image, e.g., arbitrary slice-cut can be drawn for demonstrating tomographic information in need, its image appearance may be somewhat different from conventional panoramic radiograph [7][8].

In this work, we propose a new reconstruction method of dental panoramic radiograph that directly uses the CT projection data. The key idea lies in that a virtual panoramic scanning path is introduced and that virtual panoramic projection data are synthesized in accordance. Once the virtual data are prepared, the remaining image reconstruction part is

This work was supported in part by Korean National Research Foundation [NRF-2016M3A9E9941837, NRF-2018M3A9E9024949, and NRF-2019M2A2A4A05031487] and the institute of Civil Military Technology Cooperation funded by the Defense Acquisition Program Administration and Ministry of Trade, Industry and Energy of Korean government under grant [UM19207RD2].

Taejin Kwon is with the Korea Advanced Institute of Science and Technology, Daejeon, 34141 Rep. of KOREA (telephone: +82)042-350-3868, e-mail: ktj6584@kaist.ac.kr).

Da-in Choi is with the Korea Advanced Institute of Science and Technology, Daejeon, 34141 Rep. of KOREA (telephone: +82)042-350-3868, e-mail: okpc426@kaist.ac.kr).

Jaehong Hwang is with the Korea Advanced Institute of Science and Technology, Daejeon, 34141 Rep. of KOREA (telephone: +82)042-350-3868, e-mail: okpc426@kaist.ac.kr).

Jaesung Lee is with the Dentium Co., Ltd., Yongin-si, 16803 Rep. of KOREA (telephone: +82)070-7098-7577, e-mail: rndljs@dentium.com).

Jihwan Lim is with the Dentium Co., Ltd., Yongin-si, 16803 Rep. of KOREA (telephone: +82)070-7098-7577, e-mail: jhlim@dentium.com).

Inje Lee is with the Dentium Co., Ltd., Yongin-si, 16803 Rep. of KOREA (telephone: +82)070-7098-7577, e-mail: ijlee@dentium.com).

Seungryong Cho is with the Korea Advanced Institute of Science and Technology, Daejeon, 34141 Rep. of KOREA (telephone: +82)042-350-3828, e-mail: scho@kaist.ac.kr).

pretty much the same with existing panoramic image reconstruction. To enable auto-focusing, we have prepared a number of deformed projections per a given projection. Here, deformation refers to either compression or expansion along the detector width direction. The shift-and-add method has then been used for image reconstruction resulting in a number of depth-resolved panoramic slices. Finally, auto-focusing has been applied to come up with the sharpest image. The proposed method produced quite comparable images to the real panoramic radiographs.

II. Method

The key idea for this research is to synthesize virtual panoramic projection data along a virtual panoramic scan path. After acquiring the virtual panoramic projection data, the rest of the image reconstruction is performed in much the same way as the existing panoramic image reconstruction. The overall process is shown in the schematic below.

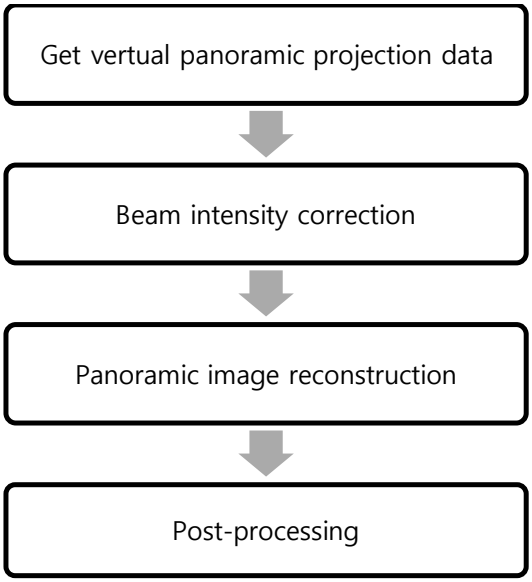


Fig. 1. Pipeline of proposed panoramic image reconstruction using virtual panoramic projection data acquired from CT scan data.

First, virtual panoramic projection data is synthesized from the CT data along the virtual panoramic scan trajectory, and then beam intensity is corrected through weighting. Then, as in the existing panorama image reconstruction method, a shift-and-add method is applied to obtain panorama images of several slices. Finally, auto-focusing has been applied to come up with the sharpest image, and the final image is obtained through additional post-processing such as filtering.

A. *CT scan system and Panoramic scan system*

The data set used in the study was obtained using the dental CBCT imaging system rainbow CT (Dentium Co., Ltd.). Since both panoramic scan data and CT scan data can be obtained from the dental CBCT system, a panoramic radiograph obtained from the system was used as a comparison group. The virtual panoramic scan trajectory was also referred to as the trajectory of the existing system. The conventional panoramic scan system is a system in which the axis of rotation moves only in one axis (back head - face).

TABLE I
Scan System Parameters

S. N.	Parameter	CT scan	Panoramic scan
1	Distance of source to detector	609.07 mm	609.07 mm
2	Distance of source to isocenter	401.85 mm	401.85 mm
3	Detector array size (pixels × pixels)	654 × 664	84 × 1328
4	Image array size (pixels × pixels)	512 × 512	2430 × 1328
5	Rotation center	Movable	Fixed
6	Scan angle(degree)	360	230
7.	Number of views per rotation	600	2553

B. *Virtual panoramic projection data synthesis*

Since the CT scan system scans 360-degree projection data of an object using a wider detector than the panoramic scan system, the CT scan data includes the data of the panoramic scan system. Therefore, it is possible to synthesize panoramic projection data of desired trajectory from CT scan data through appropriate calculation.

In this study, the parallel beam rotation angle and detector position of virtual panoramic scan trajectory was calculated, and the virtual panoramic projection data was obtained by extracting the parallel beam part from the CT data [14].

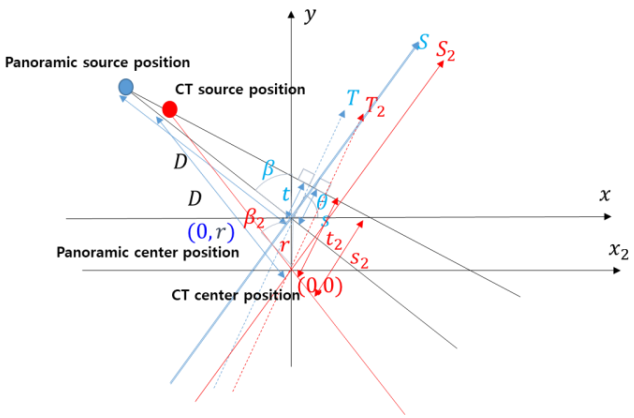


Fig. 2. The beam geometry after parallel rebinning along a virtual panoramic scan trajectory.

$$\theta = \beta + artcan\left(\frac{s}{D}\right) \text{: parallel beam source angle} \quad (1)$$

$$t = \frac{sD}{\sqrt{s^2+D^2}} \quad (2)$$

$$t_2 = t + rsin(\theta) = \frac{s_2 D}{\sqrt{s_2^2+D^2}} \text{: parallel beam detector position} \quad (3)$$

$$s_2 = \frac{t_2 D}{\sqrt{D^2-t^2}} \text{: fan beam imaginary detector position in CT} \quad (4)$$

$$\theta = \beta_2 + \arctan\left(\frac{s_2}{D}\right) \quad (5)$$

$$\beta_2 = \theta - \arctan\left(\frac{s_2}{D}\right) \text{: fan beam source angle in CT} \quad (6)$$

Using Fig. 2 and equations (1) ~ (6), virtual panoramic projection data can be obtained along a given virtual panoramic scan trajectory from CT data. First, the detector position (2) and rotation angle (1) of the parallel beam corresponding to the panoramic projection data are calculated according to the virtual panoramic scan system. Next, since the axis of rotation moves in the real panoramic scan, the parallel beam detector position is calculated in consideration of the movement of the axis of rotation (3). Finally, the position of the beam corresponding to the parallel beam is calculated from the CT data (4), (6). This calculation is repeated for each panoramic scan trajectory to extract 84x1328x2553 virtual panoramic projection data set from a 654x664x600 CT data set.

C. *Beam intensity correction*

Panoramic radiography has the advantage of showing a variety of information for dental diagnosis such as tooth structure, tooth roots, gums and nerves at once. Most of the information needed for this diagnosis is in the front of the head, including the jaw. In a panoramic scan, the x-ray passes through the back of the head and neck bones in addition to the front of the head, and the panoramic scan data includes all material information that the x-ray has passed. The shift-and-add method is used to obtain an image focused on a specific area, but material information outside the specific area is also added unfocused. Therefore, it can be seen that the image value of the reconstructed image part using the projection in the direction of passing more material on the path of the beam is significantly higher than that of other projection parts. To eliminate these artifacts, a real panoramic scan compensates for the number of photons attenuated by materials other than the region of interest by irradiating beams of different intensity depending on the projection. Thus, even if panoramic projection data is obtained along the same panoramic scan trajectory, the intensity values differ from the actual panoramic scan data and the data synthesized from the CT scan data. In this study, the beam intensity was corrected similarly to the actual panoramic system by varying the weighting for each synthesized projection.

D. *Panoramic image reconstruction scheme*

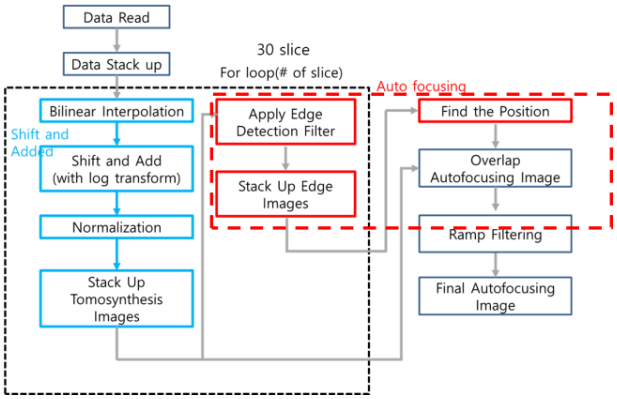


Fig. 3. Schematic of panoramic image reconstruction using panoramic scan data applying shift-and-add method.

Panorama image reconstruction using panorama scan data is performed according to the above figure.

First, create a panoramic radiograph of several slices from the panoramic scan data using the shift-and-add methods [11][12]. In the shift-and-add method, the position of the focused image layer changes with the amount of shift. Therefore, if we apply shift-and-add method with different shift amounts, we can get panoramic radiography of several slices with a different focus. After that, a well-focused panoramic radiograph is extracted using slice images with different focus depending on the part through the autofocusing method [12]. Each slice image is divided into small patches to find a slice with the best focus by comparing the sharpness of each part, and creates a well-focused panoramic radiograph by calling and combining a patch image from the panoramic slice images.

III. RESULTS

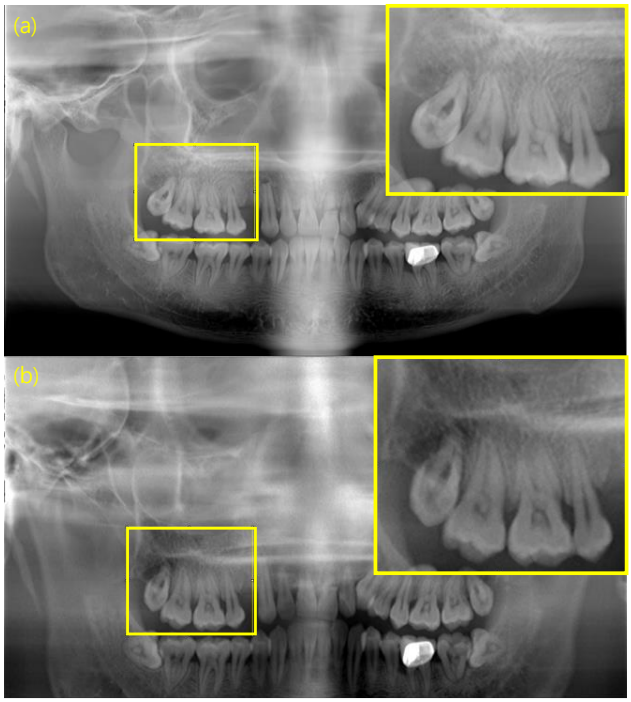


Fig. 4. (a): Panoramic image reconstructed using actual panoramic projection data, (b): Panoramic image reconstructed using virtual panoramic projection data synthesized from CBCT projection data.

Figure 4 shows a panoramic image reconstructed using panoramic scan data and CT scan data. Image (a) is a panoramic image obtained by applying the shift-and-add method and autofocusing to the panoramic projection data acquired by the existing panoramic scan system. The image of (b) is an image obtained by scanning the same phantom as used in (a) by CT scan system, synthesizing panoramic projection data from CT scan data, and obtaining a panoramic image by the same method as (a). Although the vertical magnifications of the (a) and (b) images are slightly different, it can be seen that the panoramic image of very similar quality to the actual panorama scan can be obtained using CT scan data. Since the conventional method of synthesizing a panoramic image from head CT data mainly uses a reconstructed CT image, a much lower resolution panoramic image is obtained according to the resolution of the

CT image, and many studies have been conducted to overcome this[15][16]. However, the method proposed in this study was able to reduce the resolution reduction by reconstructing the panoramic image using the projection data directly.

IV. Conclusion & Discussion

As a result, as shown in Fig. 4, an image similar to the existing panoramic radiograph can be generated from the dental CBCT image using the proposed method. Unlike the actual panoramic scanning system, the axis of rotation is fixed in the CT scanning system. Therefore, it can be seen that there is a slight magnification difference due to the positional difference of the rotation axis. However, if necessary, we can also compensate for this by calculating the magnification based on the position of the axis of rotation. The proposed method is to synthesize a set of panoramic projection data from CT data along the panoramic scan trajectory. Therefore, as shown in Fig. 5, various panorama images can be created according to how to set the panorama scan trajectory.

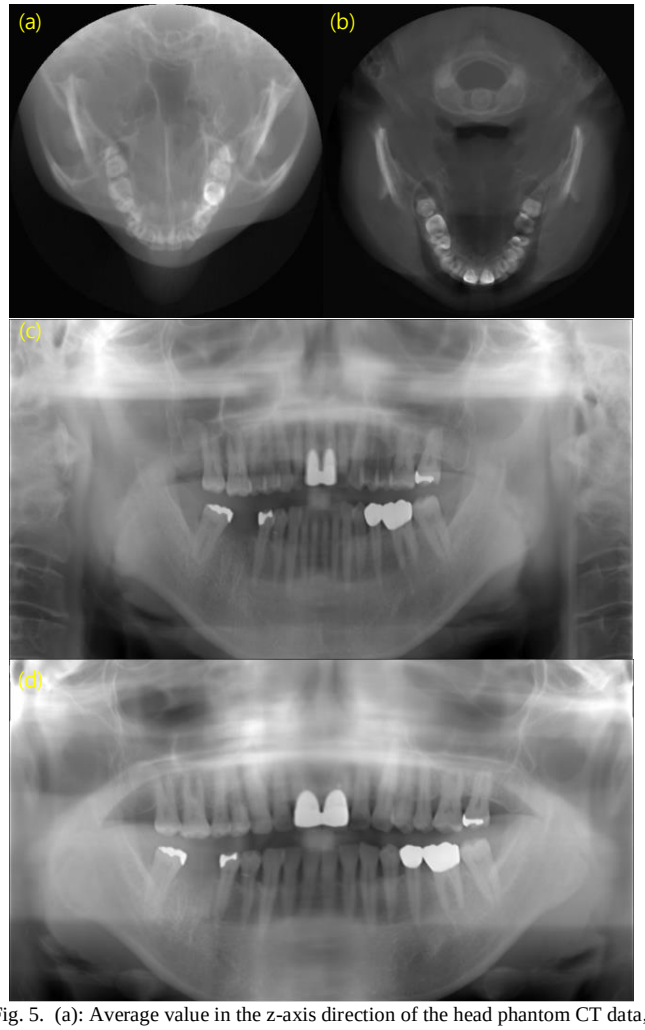


Fig. 5. (a): Average value in the z-axis direction of the head phantom CT data, (b): Average value in the z-axis direction of clinical CT data, (c): A panoramic image created by applying an existing panoramic scan trajectory to a clinical CT dataset, (d): A panoramic image created by applying a new panoramic scan trajectory to a clinical CT dataset.

In addition, panoramic scan datasets of various trajectories within the CT scan area can be synthesized without hardware limitations. Conventional panoramic scanning systems have a limited trajectory that can be set according to the hardware, and only one data set can be obtained at a time. Due to the nature of the panoramic radiograph using the shift-and-add method, an out-of-focus image is obtained when a tooth leaves the well-focused image layer. The proposed method, which can freely set the panoramic scan trajectory from the dental CBCT scan data, can freely change the shape and position of a well-focused image layer, thus overcoming the limitations of the existing panoramic scan system.

References

[1] Pauwels, Ruben, et al. "Technical aspects of dental CBCT: state of the art." Dentomaxillofacial Radiology 44.1 (2014): 20140224.

[2] Jacobs, Reinhilde, et al. "Cone beam computed tomography in implant dentistry: recommendations for clinical use." BMC oral health 18.1 (2018): 88.

[3] Kakuma, Hideo, Kohji Ohbayashi, and Yasuhiko Arakawa. "Optical imaging of hard and soft dental tissues using discretely swept optical frequency domain reflectometry optical coherence tomography at wavelengths from 1560 to 1600 nm." Journal of biomedical optics 13.1 (2008): 014012.

[4] Ogawa, Koichi, et al. "Development of a new dental panoramic radiographic system based on a tomosynthesis method." Dentomaxillofacial Radiology 39.1 (2010): 47-53.

[5] Tunchel, Samy, et al. "3D printing/additive manufacturing single titanium dental implants: a prospective multicenter study with 3 years of follow-up." International journal of dentistry 2016 (2016).

[6] Dawood, Andrew, et al. "3D printing in dentistry." British dental journal 219.11 (2015): 521.

[7] Bing, Han, et al. "An automatic method of synthesizing panoramic radiograph by unwrapping dental CT image." 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC). IEEE, 2011.

[8] Luo, Ting, et al. "Automatic synthesis of panoramic radiographs from dental cone beam computed tomography data." PloS one 11.6 (2016): e0156976.

[9] Katsumata, Akitoshi, et al. "Initial evaluation of linear and spatially oriented planar images from a new dental panoramic system based on tomosynthesis." Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and

[10] Dobbins III, James T., and Devon J. Godfrey. "Digital x-ray tomosynthesis: current state of the art and clinical potential." Physics in medicine & biology 48.19 (2003): R65.

[11] Noujeim, M., et al. "Pre-clinical evaluation of a new dental panoramic radiographic system based on tomosynthesis method." Dentomaxillofacial Radiology 40.1 (2011): 42-46.

[12] Lee, Taewon, Yeon Ju Lee, and Seungryong Cho. "Auto-focused panoramic dental tomosynthesis imaging with exponential polynomial based sharpness indicators." Medical Imaging 2017: Image Processing. Vol. 10133. International Society for Optics and Photonics, 2017.

[13] Monsour, P. A., and R. Dudhia. "Implant radiography and radiology." Australian dental journal 53 (2008): S11-S25.

[14] Kim, Changhwan, et al. "Data consistency-driven scatter kernel optimization for x-ray cone-beam CT." Physics in Medicine & Biology 60.15 (2015): 5971.

[15] Hofmann, Christian, Michael Knaup, and Marc Kachelrieß. "Reducing intra plane blurring in dental panoramas." 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC). IEEE.

[16] Hofmann, Christian, Michael Knaup, and Marc Kachelrieß. "Removing intra plane blurring in dental panoramas." Medical Imaging 2013: Physics of Medical Imaging. Vol. 8668. International Society for Optics and Photonics, 2013.

An intelligent alternating edge-preserving diffusion and smoothing neural network for limited-angle computed tomography reconstruction

Genwei Ma, Yinghui Zhang, Tong Wang, Xing Zhao and Hongwei Li *

***Abstract*—Limited-angle computed tomography is an challenging but important task in certain medical and industrial applications for non-destructive testing. To improve the reconstruction quality, this paper proposes an intelligent alternating edge-preserving diffusion and smoothing neural network (IAEDSNN). It provides intelligent directional regularization layers which encodes the visible edges as priors. Experiments demonstrate that IAEDSNN could produce promising results and outperforms significantly both the conventional regularization-based iterative methods and other popular deep learning architectures.**

***Index Terms*—Limited-angle computed tomography, directional regularization, visible edges, convolutional neural network**

I. Introduction

It is computed tomography (CT) that allows us to reconstruct the inner structures of the scanned objects in a quantitative and nondestructive way. In certain applications like the chest and breast imaging, long and thin plates imaging et.al., projections can only be gathered in a limited angular range, which leads to the severe ill-posed limited-angle reconstruction problem. If one applies the conventional reconstruction algorithms like filtered back projection (FBP) and simultaneous algebraic reconstruction technique (SART), blurring and streak artifacts would be introduced.

In the last two decades, various algorithms, trying to reconstruct better images from limited-angle data, have been proposed. The basic idea of these algorithms is to incorporate various priors into an iterative reconstruction procedure. Nevertheless, it has been reported that iterative reconstruction algorithms based on regularizations, e.g. total variation (TV) [1], could not work well on limited-angle data. To better preserve the edges and suppress the artifacts for limited-angle reconstruction, the L_0 norm [2] of image gradient based models and algorithms have been proposed, which recover edge and structure information more effectively. Especially, the alternating edge-preserving diffusion and smoothing (AEDS) model and algorithm proposed in [3] produces state-of-the-arts reconstructions, by fully utilizing the visible singularities as priors. One drawback of the AEDS algorithm, however, is that its efficiency is based on the essential assumption that the CT images can be approximated well by piecewise constant functions. This assumption is not always valid in practice. Another limitation of the AEDS algorithm is that it converges quite slow. For general problems, it might need about 2000 iterations to reconstruct satisfying images.

In recent years, deep learning (DL) has attracted a lot of attention because of its outstanding performance in image precessing, natural language processing et. al. So, some researchers began to apply deep learning for image reconstructions. It is used as the post-processing technique at the beginning, and then combined with the reconstruction procedure recently. The robustness of deep learning in practical applications is still a concern, considering that deep neural networks might be vulnerable to small perturbations, such as adversarial examples. By combining with the reconstruction procedure, the robustness can be improved [4]–[6]. Compared with iterative reconstruction algorithm based on regularizations, the DL-based regularizers are more Intelligent, and the corresponding algorithm could converge faster and achieve higher quality reconstructions. For limited-angle reconstructions, various DL models have also been proposed, which are usually constructed by unrolling some iterative reconstruction scheme into deep learning architectures [7], [8]. According to our knowledge, these newly proposed models did not explicitly take the visible edges as priors, i.e. the visible edge prior is not utilized to optimize the network architecture.

In this paper, we aim to build a DL-based neural network which takes fully consideration the visible edge prior. This is achieved by unrolling the AEDS reconstruction algorithm and then combining other ingredients into the network for better efficiency. One distinctive aspect of the proposed model is that only one-dimensional kernels are employed in the Convolutional Neural Networks (CNN), instead of the usual two-dimensional kernels widely adopted in the literature. The main contributions of our work can be summarized as follows:

- The proposed intelligent regularization terms could be adaptively learned from the training data, free of parameter tuning.
- The proposed network architecture explicitly incorporates the visible edge information as priors, which is, according to the best of our knowledge, brand new in the limited-angle reconstruction community.
- The proposed model builds into the attention mechanism, which is used to preserve visible feature and edge in-

This work is supported in part by the National Natural Science Foundation of China under Grant 61971292, 61827809 and 61771324 .

Asterisk indicates corresponding author.

Genwei Ma, Yinghui Zhang, Tong Wang, Xing Zhao and Hongwei Li are with school of Mathematical Sciences, Capital Normal University, Beijing, 100048, China, and with Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, 100048, China. e-mail: (MaGenwei@126.com)

E-mail: hongwei.li91@cnu.edu.cn.

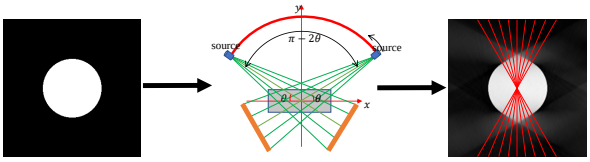


Fig. 1. A uniform-density disk under the scanning angular ranges from 120° data ($[\frac{\pi}{6}, \frac{5\pi}{6}]$). The red region display the invisible edge and the rest part is the visible edges.

formation better during the iterations, such that better efficiency follows.

The remainder of this paper is organized as follows. In section II, the proposed model, i.e. IAEDSNN, is described. Section III is dedicated to the experiments. We conclude our paper in Section IV.

II. Method

A. The IAEDSNN model and algorithm

Without loss of generality, let's fix the scanning configuration for limited-angle fan-beam CT as illustrated in Fig.(1). With the specified scanning configuration, the distribution pattern of visible and invisible edges has motivated the alternating edge-preserving diffusion and smoothing (AEDS) model for limited-angle reconstruction [3]. Let A and b denote the projection matrix and the acquired data, respectively, then the AEDS model reads

$$\min_u \left\{ \frac{1}{2}\|Au-b\|_W^2 + \lambda_1 \phi(u) + \lambda_2 \psi(u) \right\}, \qquad (1)$$

where u is the linear attenuation coefficient to be reconstructed and W is defined as a diagonal matrix [9]. In the AEDS algorithm, $\|\nabla_x u\|_0$ or $\|\nabla_x u\|_1$ is chosen for $\phi(\cdot)$, while $\|\nabla_y u\|_0$ or $\|\nabla_y u\|_1$ is chosen for $\psi(\cdot)$, where $\nabla_x u$ and $\nabla_y u$ denote the gradient operator along the $x-$direction and $x-$directionrow, respectively.

Even the AEDS model achieves very promising results compared to existing algorithms, it still suffers from some limitations: its effectiveness relies heavily on the piecewise-constant assumption for the ideal u. Besides, the parameters of the AEDS algorithm needs fine-tuning and convergence is also slow.

To overcome the above limitations of AEDS, an intelligent alternating edge-preserving diffusion and smoothing neural network (IAEDSNN) is proposed. After the roles of $\phi(u)$ and $\psi(u)$ in AEDS model are taken over by directional convolution neural network, the model of IAEDSNN comes into being

$$u = arg\min_u \frac{1}{2}\|Au-p\|_W^2 + \lambda_1 CNN_h(u) + \lambda_2 CNN_v(u), \qquad (2)$$

where CNN_h and CNN_v denote the horizontal x-direction transform and vertical y-direction transform operators implemented by directional CNN, respectively. CNN_h and CNN_v play different roles in (2), having more flexible effect of edge-preserving diffusion and smoothing, respectively, which contributes to loose the piecewise-constant assumption. The parameters λ_1 and λ_2 are weighting parameters to adjust the effect of CNN_h and CNN_v in (2). Fortunately, they can be intelligently learned by CNN, free of fine-tuning.

Algorithm 1 IAEDSNN

Input: projection data p, initialization $u^{(0)} = 0$, iteration number n and relaxation parameter ω for OS-SART.
Output: $u^{(n)}$
  Main Iteration Loop
  for $k = 0; k < n; k++$ **do**
    # setp 1: reconstuction image by OS-SART
    $u^{k+\frac{1}{3}} = \text{OS-SART}\ (u^{(k)}, p, \omega)$
    # step 2: horizontal x-direction CNN regularization
    $u^{k+\frac{2}{3}} = CNN_h(u^{k+\frac{1}{3}})$
    # step 3: vertical y-direction CNN regularization
    $u^{k+1} = CNN_v(u^{k+\frac{2}{3}})$
  end for

The IAEDSNN can be summarized as **Algorithm 1** described as above. First, just one sweep of OS-SART was used for reconstruction in the IAEDSNN workflow. Then, the horizontal x-direction regularization obtains edge-preserving diffusion. Next, vertical y-direction regularization achieves edge-preserving smoothing.

B. Edge-preserving accelerating extension

Considering the high computation cost and slow convergence speed of the iterative algorithms, we incorporate the attention mechanism and the pooling layer to accelerate the procedure of edge-preserving diffusion from visible features. The neural network architecture diagram is shown in Fig. 2.

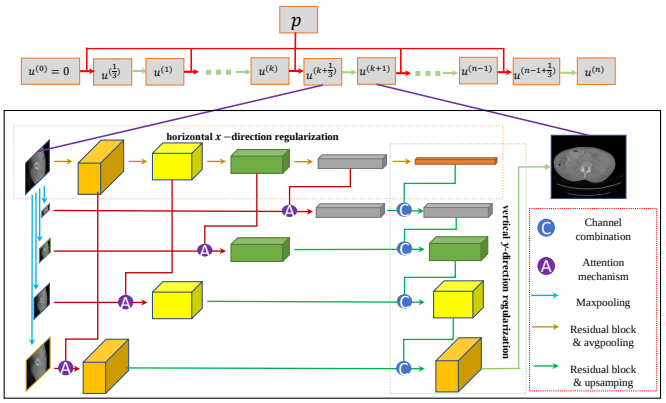


Fig. 2. A schematic workflow of the proposed IAEDSNN, which accelerates the procedure of learning invisible features using attention mechanism and pooling layer.

C. OS-SART and its back-propagation

Given the ordered-subset $\{i_1, i_2, \cdots, i_n\}$, corresponding system matrix subset $\{A_{i_1}, A_{i_2}, \cdots, A_{i_n}\}$ and projection data $\{p_{i_1}, p_{i_2}, \cdots, p_{i_n}\}$, the k-th iteration image, which is updated after each subset is considered, can be obtained by

$$\begin{aligned} u^{(k+\frac{1}{3n})} &= u^{(k)} + \omega A_{i_1}^T (p_{i_1} - A_{i_1} u^{(k)}) \\ u^{(k+\frac{2}{3n})} &= u^{(k+\frac{1}{3n})} + \omega A_{i_2}^T (p_{i_2} - A_{i_2} u^{(k+\frac{1}{3n})}) \\ &\vdots \\ u^{(k+\frac{1}{3})} &= u^{(k+\frac{n-1}{3n})} + \omega A_{i_n}^T (p_{i_n} - A_{i_n} u^{(k+\frac{n-1}{3n})}), \end{aligned} \qquad (3)$$

where u represents the desired objective image and ω is relaxation parameter for OS-SART.

To perform the back-propagation procedure for the proposed framework, the gradients in each layer need to be calculated. Most of the gradient computation is similar to other well-known convolution neural network (CNN) models except that some extra care is needed for the layer with OS-SART operations. In addition, the calculations are purely algebraic. Supposing the gradient of $u^{k+\frac{1}{3}}$ is $\partial u^{k+\frac{1}{3}}$, then the associated gradient $\frac{\partial u^{k+\frac{1}{3}}}{\partial u^k}$ can be obtained through the chain rule. The gradient of the OS-SART can be expressed as

$$\frac{\partial u^{k+\frac{1}{3}}}{\partial u^k} = (1-\omega A_{i_1}^T A_{i_1})(1-\omega A_{i_2}^T A_{i_2})\cdots(1-\omega A_{i_n}^T A_{i_n}). \tag{4}$$

III. Experimental results

In this section, we evaluate the performance of the proposed IAEDSNN by comparing with classical and learning based reconstruction methods. The geometrical scanning parameters of the simulated CT system are listed in Table I.

A. Experimental setup

We consider two different types of datasets to validate the performances of our proposed reconstruction method,

- Clinical dataset. The first dataset for the validation is downloaded form website (https://www.cancerimagingarchive.net/), which contains 2732 thickness full dose CT images from 13 patients. For the dataset used in our experiment, the number of training images is 2000 from 8 patients, and the number of testing images is 732 from the other 2 patients. All of them are from the abdomen, chest, and lung parts of body, with the size of 512×512.
- Printed circuit board (PCB) dataset. The PCB data are obtained by scanning 7 real flat objects. The complete projection data are acquired by full angular scanning. Then, the OS-SART algorithm is applied on these complete data to reconstruct the reference ones. The reconstructed images are also size of 512×512.

TABLE I
Geometrical parameters for clinical data and PCB data experiments

Parameter	Clinical data	PCB data
Number of detector units	600	1024
Scanning angle (degree)	$[20°, 160°]$	$[30°, 150°]$
X-ray source to the rotation center (mm)	570	350
X-ray source to detectors (mm)	1015	782
Scanning angular interval (degree)	0.5	0.5

The proposed IAEDSNN is implemented with python using pytorch frame and all the experiments are performed on a PC (Intel(R) Xeon(R) CPU E5-2643 v4 and 512GB RAM). The training stage is executed by using a graphic processing unit card GTX 1080ti with 11GB video memory.

B. Network training and parameters selection

To quickly training our proposed network, training processing should maximally reduce the forward and back-propagation procedure. Hence, we propose a segment-by-segment pre-training phase followed by end-to-end training phase for fine-tuning.

1) Pre-training: For clinical data and PCB data, we set the iteration times as 3 and 6, respectively. For the first iteration, we set the reconstruction images from OS-SART algorithm with zero-initial as the inputs and the reference images are regraded as label images. Thanks to OS-SART, learning parameters and the inputs gradient are not needed. For later iterations, we set reconstructed results of the output from previous layer by OS-SART as the inputs. The loss function is solved by adopting the Adam technique with the learning rate of 10^{-4} and the decay factor of 5×10^{-6}. 100 epochs are used as an empirical stopping criterion and training samples are randomly shuffled.

2) Fine-tuning: The second stage is an end-to-end training. The loss function is solved by adopting the standard stochastic gradient descent technique with the learning rate of 1×10^{-6}.

C. Experimental results

Fig. 3 and Fig. 4 show the images reconstructed from the test data from different part of human body by using different algorithms. The clinical data do not satisfy the assumption of piecewise constant or piecewise smooth. So, the iterative algorithms based on regularization obtain relatively worse results. By checking the results carefully, one can see that the results reconstructed by the IAEDSNN have more complete structures and less artifacts. As shown in Fig. 3 and 4 (c) - (e), some structures of the results constructed by OS-SART, L_0 and L_1 regularization are neither clear nor complete. The results of FBPConvNet is better than these three methods, but have more artifacts and less details than the proposed one's. The PSNR and SSIM values of the reconstructed images for the different methods is shown in Table II. As shown in Table II, the proposed algorithm achieves a better approximation to the reference image, possessing the highest values among compared methods.

TABLE II
PSNR and SSIM of different algorithms for clinical images

Algorithm	Result 1		Result 2	
	PSNR	SSIM	PSNR	SSIM
OS-SART	24.758	0.88395	24.139	0.89314
IR with L_0 regularization item	26.324	0.92014	25.833	0.92328
IR with TV regularization item	26.355	0.92075	25.748	0.9246
FBPConvNet	42.665	0.99346	28.190	0.96574
IAEDSNN	45.701	0.99654	29.015	0.96193

The difficulty for PCB imaging is to recover long invisible edge. Fig.5 and Fig.6 show the results of different algorithms. The AEDS algorithm performs better, when the long invisible edge are near horizontal as shown in Fig.5 (e). However, when the slope of horizontal edges becomes larger, it will fail to reconstruct the tilted edges of images well, as shown in Fig. 6 (e). The FBPConvNet shows poor robustness and can

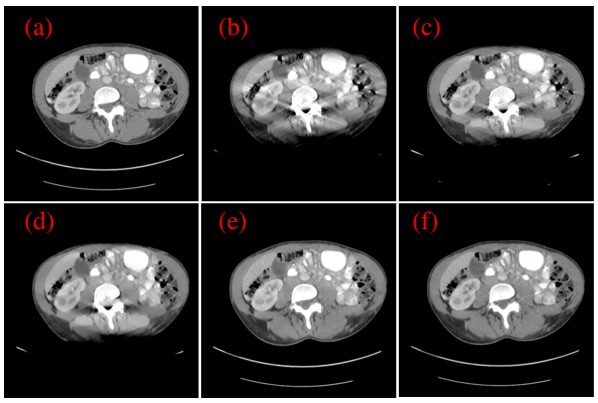


Fig. 3. Reconstructed results. (a) The reference image, (b) OS-SART, (c) L_0 regularization, (d) L_1 regularization, (e) FBPConvNet, (f) IAEDSNN. The display window is [-160,240] HU.

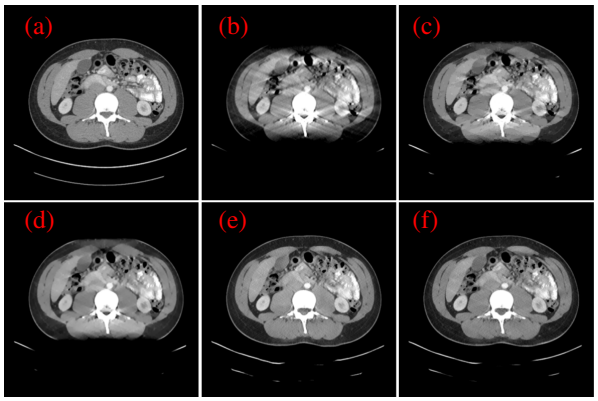


Fig. 4. Reconstructed results. (a) The reference image, (b) OS-SART, (c) L_0 regularization, (d) L_1 regularization, (e) FBPConvNet, (f) IAEDSNN. The display window is [-160,240] HU.

only recover disconnected boundaries, when data is slightly inconsistent. All of the above cases make little difference to the performance of the propose method, as shown in Fig. 5 (f) and Fig.6 (f).

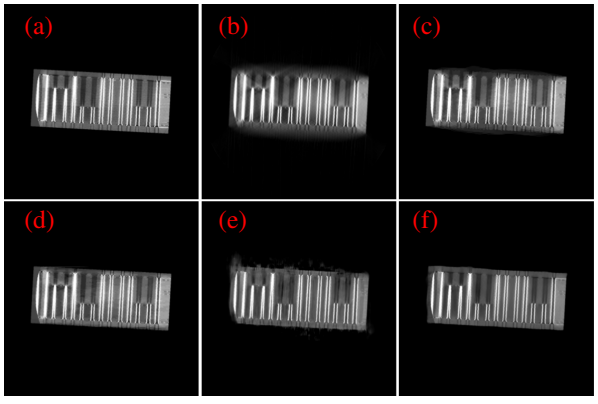


Fig. 5. Reconstructed results. (a) The reference image, (b) OS-SART, (c) L_0 regularization, (d) AEDS, (e) FBPConvNet, (f) IAEDSNN. The display window is [0,0.7].

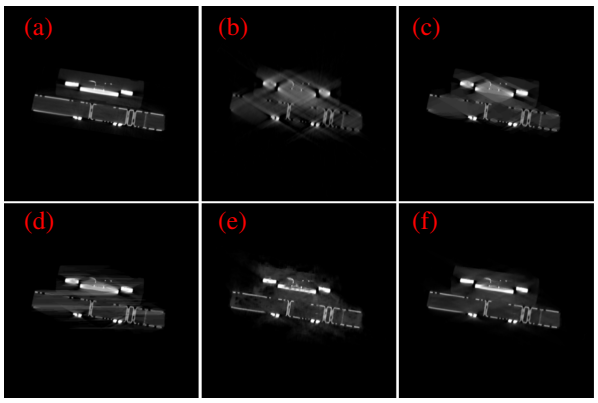


Fig. 6. Reconstructed results. (a) The reference image, (b) OS-SART, (c) L_0 regularization, (d) AEDS, (e) FBPConvNet, (f) IAEDSNN. The display window is [0,0.7].

IV. Conclusions

We have proposed an intelligent alternating edge-preserving diffusion and smoothing network for limited-angle data reconstruction. According to the specified scanning configuration, the edges are visible along the x-direction, while invisible along the y-direction, which is the key for designing two direction regularization terms. Experiments against existing popular methods show that the proposed IAEDSNN algorithm can, not only, break the restriction caused by the the piecewise constant assumption, but also reconstruct images from limited-angle data much better than state-of-the-art methods.

References

[1] E. Y. Sidky, C. M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct," *Journal of X-ray science and technology*, vol. 14, no. 2, pp. 119–139, 2009.

[2] W. Yu and L. Zeng, "ℓ_0 gradient minimization based image reconstruction for limited-angle computed tomography," *PLOS ONE*, vol. 10, no. 7, pp. 1–15, 07 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0130793

[3] J. Xu, Y. Zhao, H. Li, and P. Zhang, "An image reconstruction model regularized by edge-preserving diffusion and smoothing for limited-angle computed tomography," *Inverse Prob.*, 2018.

[4] H. Chen, Y. Zhang, Y. Chen, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "Learned experts' assessment-based reconstruction network ("learn") for sparse-data ct," *IEEE Transactions on Medical Imaging*, 07 2017.

[5] y. yang, J. Sun, H. Li, and Z. Xu, "Deep admm-net for compressive sensing mri," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Curran Associates, Inc., 2016, pp. 10–18. [Online]. Available: http://papers.nips.cc/paper/6406-deep-admm-net-for-compressive-sensing-mri.pdf

[6] j. he, Y. Yang, Y. Wang, D. Zeng, Z. Bian, H. Zhang, J. Sun, Z. Xu, and J. Ma, "Optimizing a parameterized plug-and-play admm for iterative low-dose ct reconstruction," *IEEE Transactions on Medical Imaging*, pp. 1–1, 08 2018.

[7] J. Wang, L. Zeng, C. Wang, and Y. Guo, "Admm-based deep reconstruction for limited-angle ct," *Physics in Medicine and Biology*, vol. 64, 04 2019.

[8] H. Zhang, B. Dong, and B. Liu, "Jsr-net: A deep network for joint spatial-radon domain ct reconstruction from incomplete data," 2018.

[9] M. Jiang and G. Wang, "Convergence studies on iterative algorithms for image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 22, no. 5, pp. 569–579, 2003.

A New Imaging Modality Based on The Flat-Panel Sources

Jiayu Duan, Yang Li, Jianmei Cai and Xuanqin Mou

***Abstract*—With the pursuit of searching for solutions to conventional X-ray sources, the field emitters based sources have been developed. It can equip more than thousands of X-ray sources in one chip based on cold cathodes of nanomaterials, differing from the conventional X-ray sources. With the merits of high spatial resolution and fast response, it enables people to design multiple novel imaging geometries based on irregular situations such as imaging in narrow space and for bulk workpieces. In this paper, we propose a dense flat-panel source to design a portable imaging system. It can be used for narrow space imaging and industrial imaging for bulky workpieces. Theoretically, each X-ray source in the device should lighten one by one to get a non-overlapped measurement. However, the current radiation dose of the whole device is quite low due to the high density of the flat-panel sources. The addressable lighting of the flat-panel source may not be able to provide enough number of photons for imaging. At the same time, if we light all the sources to guarantee the X-ray energy for imaging, the projection will be overlapped and hard to optimize. Hence, we investigate the imaging probability of the overlapped projection. We propose a novel imaging modality using the flat-panel sources and a coupled algorithm to estimate a feasible radioscopy image of the object. With an adaptive estimation process, the proposed method can recover the virtual parallel projection from the overlapped cone-beam projection. Quantitative and qualitative results show the efficiency of our proposed method, which shows the potential of the new imaging modality.**

Keywords— flat-panel sources, image recovery method

I. Introduction

X-ray computed tomography has shown great power in both clinical and industrial fields with the merit of non-destructive. Although this technology benefits our lives, the traditional X-ray source still has some drawbacks, which limit further development. In order to cover the object in the imaging region of interest, it demands a long stand-off distance to the object of interest. It leads to the bulk imaging system. Hence, it degrades the employment in the narrow space. Meanwhile, in the clinic, the pursuit of a low radiation dose is a principle in medical treatment for the X-ray. The high dose of X-ray may cause genetic changes. However, the long stand-off distance results in unnecessary radiation in the no-interest region. Moreover, due to the intrinsic characteristic of the conventional X-ray source, it has long response time and short lifetime and high power consumption. Theses limit suitability for situations outside a traditional hospital setting. [1]

Nowadays, with the urgent request for innovative solutions to the conventional X-ray source problems, the field emitters based source has been developed[2-4] to overcome these drawbacks. By replacing the thermionic cathodes with cold cathodes of nanomaterials, the new source can become more compact with high temporal and spatial resolution. The cold cathode source has fast response, low energy consumption, and addressable switching abilities, which benefit the design of many different imaging geometries. Up to now, many fields make an attempt using the new X-ray source. For example, digital breast tomosynthesis uses the carbon nanotube array sources (CNT) instead of the traditional source to acquire information to reconstruct, which avoids the long acquisition time and motion blur[5, 6]. Gao et al. [7] presented a prototype physiologically gated micro-computed tomography (micro-CT) scanner based on a carbon nanotube field emission micro-focus x-ray source. The novel x-ray source allows x-ray pulses and imaging sequences to be readily synchronized and gated to non-periodic physiological signals from small animals. In this way, it can eliminate the motion artifacts caused by the physiological movement of small animals.

Moreover, with the development of nanomaterial technology, a more compact version of the field emission source: the flat-panel source is proposed. Flat-panel X-ray source can pack thousands of X-ray cells in one chip [2]. Same as CNT array, the flat-panel source is also addressable. With these characteristics, it can significantly reduce the distance of source-to-object. It offers a chance to cut down the acquisition time and fewer mechanical constraints. Edwin et al. proposed a new flat panel source based on CNT that each X-ray cell is addressable and corresponds to a pixel in an X-ray image[8]. The proposed flat-panel source generates a parallel beam, which eliminates the distortion of the magnification of the imaging object. While employing the flat panel sources in practice is still on the way due to technology limitations. The voltage of Edwin's X-ray can reach to 30 kVp without electrical breakdown. But it still quite low for imaging. Currently, many companies have launched their cone-beam flat-panel sources, such as Adaptix, Stellarray, etc. Adaptix proposed an addressable 5X5 flat-panel source with a 1cm interval used in chest tomography and tomosynthesis. Stellarray designed a flat-panel source without addressable lighting for medical product sterilization, water and wastewater treatment, food sterilization, and materials processing.

In this paper, we proposed a denser packed flat-panel source with 48X48 cells in one chip. Each cell has a 1mm interval. Compared to the long interval and small size of sources, it can provide more angle information. Moreover, the lighting filed is more uniform, which alleviates the distortion. Currently, we are still on the way to solve the

JY Duan,Y Li, JM Cai, and XQ Mou are with the Institute of Image processing and Pattern recognition, Xi'an Jiaotong University, Xi an, Shaanxi 710049, China. (e-mail: duanjiayu@stu.xjtu.edu.cn yangl9394@163.com, , xqmou@mail.xjtu.edu.cn)

addressable lighting of the proposed flat-panel source. Because all the cells are packed in one small chip, the power of the device is quite low. Only lighten one cell at a time cannot provide enough number of X-ray photons to satisfy the imaging requirement. However, lighten all the cells up causes the projection measurement overlapped, which makes the problem hard to solve. In order to evaluate the imaging ability of the proposed source, we design a coupled algorithm that recovers a virtual parallel projection of the imaged object from the overlapped cone-beam measurement, which can be useful both in industrial and medical imaging.

II. METHODS

Mathematically, the physical process of the flat panel sources can be represented as:

$$I_p = \sum_{q=1}^{Q} s_{mn} v_{mn} I_0 e^{-g_{pq}} + \varepsilon , \tag{1}$$

where p is the detector index, q is the source index, I_0 is the initial photon intensity of each ray. v_{mn} and s_{mn} represent beam shape and detector reception efficiency, respectively. m,n is the coordinate on the detector plane, which indicates the absolute length between the $q-th$ source center and $p-th$ detector. g_{pq} represents the attenuation coefficient integral of the path through the $q-th$ source unit to the $p-th$ detector. ε denotes as the noise and scatter component during the measurement.

Assume that there exist virtual ray sources of an array, each is parallel to the real detection plane. And the number of virtual ray sources is the same as the number of detectors. Hence, we can obtain the analogous radioscopy measurement without aliasing, by rebinning the overlapped cone-beam measurement to a parallel beam measurement. Moreover, the projection data obtained by parallel beams have the same amplification factor at different locations. It is easy to reveal the inner structure of the object.

Suppose that the ray from q source to p detector is rebinned to the virtual ray source denoted as ii'. The line integral of ii' can be narrated as:

$$g_{i'} = w_{pq} g_{pq} \tag{2}$$

where w_{pq} is the rebin coefficient represented as:

$$w_{pq} = \frac{D}{\|p-q\|} \tag{3}$$

$\| p\text{-}q \|$ is the distance between q source to p detector. The detail is decipted in Fig.1.

Taking all sources into the calculation, the projection in the $p-th$ detector can be represented as:

$$g_{p'} = \frac{\sum_p \sum_q w_{pq} g_{pq}}{\sum_p \sum_q w_{pq}} \tag{4}$$

$g_{p'}$ is the weighted average of the projection value at detector p. The virtual parallel value may come from the rearranged ray of the same ray source to the adjacent detector (due to interpolation), or from different ray sources to different detectors.

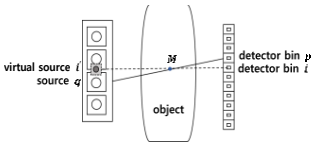
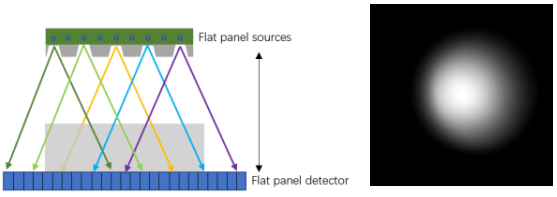


Fig. 1. The virtual parallel beam



(a) (b)

Fig. 2. Imaging system. (a)The imaging system, the gray wedge-shaped in front of the flat panel source indicates the collimator of the source. (b) The air scan measurement

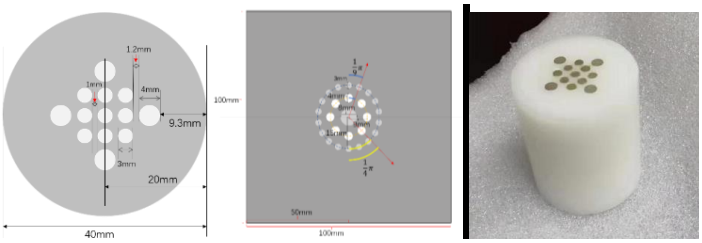
Hence, the transformation from cone-beam to the parallel beam can be denoted as:

$$g_{p_i q_i} = m_{p_i q_i}\left(g_{p'}\right) = \frac{g_{p'} \sum_p \sum_q w_{pq} - \sum_{i,i\neq p_i} \sum_{j,j\neq q_i} w_{ij} g_{ij}}{w_{p_i q_i}} \tag{5}$$

where $m_{p_i q_i}(.)$ is the transformation function for every source and detector. In this problem, we can use the least square method to optimize.

$$g'_p = \arg\min_{g'_p} \sum_{p=1}^{P} \left\| I_p - \sum_{q=1}^{Q} s_{mn} v_{mn} I_0 e^{-m_{pq}\left(g_{p'}\right)} \right\|_2^2 \tag{6}$$

However, Eq.6 is hard to optimize, because the unknown numbers in the optimization equation are much larger than the known numbers and have exponential terms in the equation. These make the optimization problem ill-posed and non-convex. Hence, we divide this problem into two situations. First, when the size of the measured object is smaller than that of the lighting range of the sources, $m_{p_i q_i}\left(g_{p'}\right)$ is solved by using the estimation of the shape and size of the object and considering the boundary conditions of the object. Second, when the size of the measured object is bigger than that of the lighting range of the flat panel sources, the weight is proportional to the length between source and detector.



(a) Phantom1 (b) Phantom2 (c)real phantom

Fig. 3. Two phantoms used in Simulation. (a) Two materials are simulated. The lighter dots denote the PMMA material and the background is composed of PE. (b) Three materials are generated, background denotes the PE, the gray dots represent PS in the phantom2, lighter dots are the PMMA materials. (c) real phantom based on the phantom1.

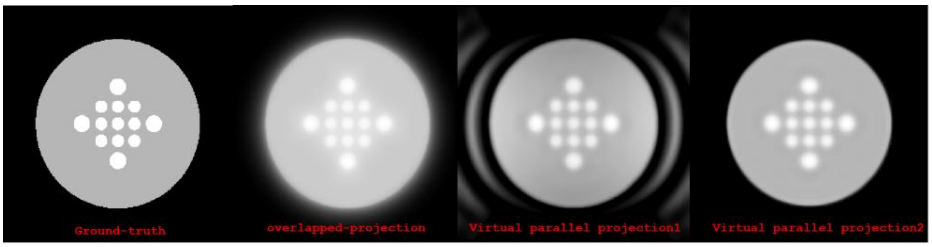


Fig. 4. The result of phantom1. (a) The ground truth image, (b) The overlapped projection (c) is the recover result without height information (d) is the recover result with height information. The display window set as [1 1.5].

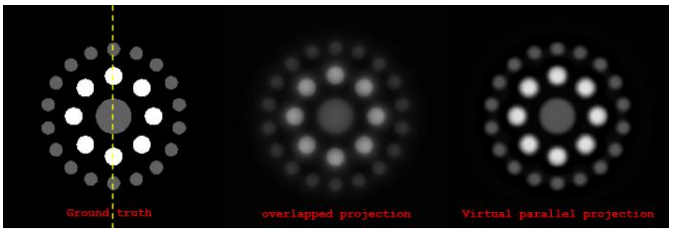


Fig. 5. The result of phantom2. (a) The ground truth image, (b) The overlapped projection (c) The virtual parallel result. The display window set as [1 1.5]. The yellow line denotes the profile used in the quantitive comparison.

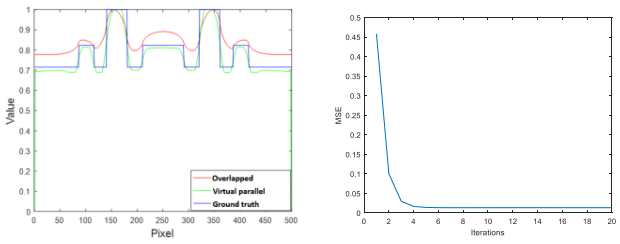


Fig. 6. (a)profile of the yellow line position. (b) the convergence curve

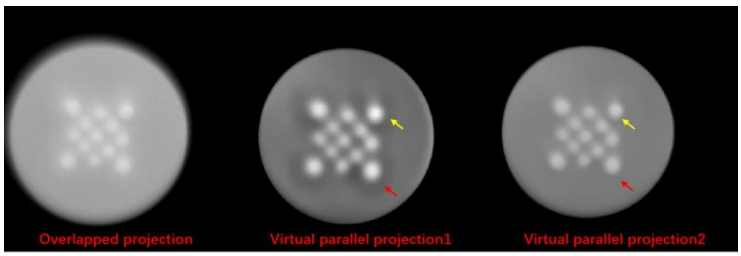


Fig. 7. The result of real data (a) The overlapped projection (b) The recover result with fixed weight. (c) The recovery result with adaptive weight. The display window set as [1 2].

Whereas, in the simplification mentioned above, we assume the object is uniform, which is not suitable in practice. Therefore, we adaptively update the weight w in iterations to add additional structure information about object materials. The update process can be formularized as:

$$w_{pq}^{new} = w_{pq}^{old} - \alpha(w_{pq}^{new} - w_{pq}^{old}) \quad (7)$$

where α is the update interval. In this paper, we all set α=0.3. And Eq.6 can use the gradient descent method to iteratively optimize.

III. Experiments and analysis

1) The imaging system setting

The imaging system based on the flat panel sources is described in Fig.2. The flat panel sources contain 48×48 sources with a 1mm interval between every two sources. Moreover, the collimators set in front of the flat panel sources in order to satisfy the imaging requirements. The emit angle of the sources is 22.6° due to the collimators. The detail is showed in Fig.5. The distance between the flat panel sources and detectors is 250mm. For simplicity, the object is directly placed on the detector. The size of the detector is 1548×1548 with $0.1mm\times0.1mm$ resolution. In practice, the beam of the X-ray source is not even. Thus, before carrying out the experiment based on the flat-panel source array, we first estimated the beam shape of the flat-panel sources. And then, the estimated beam shape was used in the following rebinning algorithm. Here, we measured the air scan under the listed scan protocol: 34Kev, 190 uA, 10s exposure time, as depicted in Fig.2(b).

In the fitting process, we assume the beam shape of the single source can be modeled with a fourth-order polynomial. We used the least-squares method to obtain the coefficients of the polynomial. We designed two kinds of phantoms in the simulation experiments. One size is smaller than the lighting range of the flat panel sources. The other is bigger than the lighting range with different materials. We narrate the details of these phantoms in Table I and Fig.3. In the real data experiment, we used the sketch of phatom1 to create the real phantom, showed in Fig.3 (c).

TABLE I. Phantom information

Phantom	*size*	*materials*
Phantom1	*40*40*50mm*	*2 (PMMA, PE)*
Phantom2	*100mm*100mm*50mm*	*3 (PMMA, PE, PS)*

2) Results and analysis

A series of experiments were carried out to validate the availability of the imaging system and method. In the restoration process, a TV regularization term is added to restrain the solution. The recover result of the phantom1 is described in Fig.4. Because the size of the phantom1 is smaller than the lighting range of the flat panel sources, we first estimated the height of the phantoms. Under this circumstance, the oscillation of the edge can be suppressed, compared with no height information, as shown in Fig.4(b) and (c). Fig.5 showed the recovery results of phantom2. Fig.6 is the quantitive comparison result and convergence illustration. From the comparison, we found that after the recovery process, the value in the yellow line position is closer than the initial projection. Moreover, it has sharper edges than the initial projection with contrast remained. These indicate the efficiency and robustness of the proposed method. In real data experiment, we scan the phantom under the listed scan protocol: 37Kev, 375 uA, 5s exposure time. Moreover, we also used fixed weight to recover the initial projection as the comparison with adaptive weight. We can see that the adaptive weight can avoid the artifacts pointed by red and yellow arrows. It is easy to comprehend this phenomenon. The fixed weight ignores the interior of the phantom is not uniform. Hence, the artifact may appear when using residual propagation. In simulation experiments, the center of the object is aligned with the center of the ray source. Thus, the drawbacks of the fixed weight are not prominent. In the real system setting, these two centers are no longer aligned which means the tails around the object are not uniform. It explains the artifacts showed in Fig.7(b).

IV. Discussion and conclusion

From the quantitative and qualitative results, our proposed method can recover the virtual parallel projection from the overlapped projection. And the profile and convergence results indicate the efficiency and robustness of the proposed method. The proposed method can maintain good contrast compared with initially mixed projection. While in the real data experiment, the result may not be very satisfactory. We are going on improving the algorithm for better recovery effects from the mixed projections, and investigating correction methods for non-uniform among sources. Meanwhile, with the improvement of the current device, the resolution can be elevated when addressable lighting mode can be generated. In the future, we will also promote the image quality based on the addressable lighting mode and try other imaging modalities such as tomosynthesis.

Acknowledgment

This work was supported in part by the National Program on Key Research Project (No.2016YFA0202003).

References

[1] G. Travish, F. Rangel, M. Evans *et al.*, *Addressable flat-panel x-ray sources for medical, security, and industrial applications*, p.^pp. OPO: SPIE, 2012.

[2] D. Chen, X. Song, Z. Zhang *et al.*, "Transmission type flat-panel X-ray source using ZnO nanowire field emitters," *Applied Physics Letters,* vol. 107, no. 24, pp. 243105.

[3] R. J. Parmee, C. M. Collins, W. I. Milne *et al.*, "X-ray generation using carbon nanotubes," *Nano Convergence,* vol. 2, no. 1, pp. 1.

[4] M. T. Cole, R. J. Parmee, and W. I. Milne, "Nanomaterial-based x-ray sources," *Nanotechnology,* vol. 27, no. 8, pp. 082501.

[5] X. Qian, A. Tucker, E. Gidcumb *et al.*, "High resolution stationary digital breast tomosynthesis using distributed carbon nanotube x-ray source array," *Medical Physics,* vol. 39, no. 4, pp. 2090.

[6] G. Yang, R. Rajaram, G. Cao *et al.*, "Stationary digital breast tomosynthesis system with a multi-beam field emission x-ray source array."

[7] G. Cao, Y. Z. Lee, R. Peng *et al.*, "A dynamic micro-CT scanner based on a carbon nanotube field emission x-ray source," *Physics in Medicine & Biology,* vol. 54, no. 8, pp. 2323-2340.

[8] E. J. Grant, C. M. Posada, C. H. Casta?o *et al.*, "A Monte Carlo simulation study of a flat-panel X-ray source," *Applied Radiation & Isotopes,* vol. 70, no. 8.

The Effect of CT Dose Reduction on Proton Therapy Dose Calculation

Masoud Elhamiasl, Sara Teruel Rivas, Koen Salvo, Edmond Sterpin, and Johan Nuyts

***Abstract*—In adaptive proton therapy, a CT scan is acquired in each treatment session to enable detection of and correction for changes in the anatomy which would otherwise cause significant changes to the dose distribution. This CT is not used to delineate lesions or tissues at risk, it is only used to compute proton stopping powers for dose calculations. Because a treatment typically involves 30 to 40 treatment sessions, the total dose associated with these CT-scans is significant, in particular, if 4D CT is used to account for breathing effects. We hypothesized that the signal-to-noise ratio provided by conventional CT protocols is higher than needed for this application. To verify this hypothesis and enable a patient dependent reduction of the CT dose, a CT dose reduction simulation tool is developed to simulate lower-dose CT scans from an existing standard-dose scan. The simulated lower-dose scans are then used for treatment planning and the results are compared with that of the standard dose scan. The preliminary results show that the dose reduction by a factor up to 10 does not have a significant effect on proton dose distribution.**

***Index Terms*—CT, dose reduction, low-dose simulation, proton therapy, treatment planing.**

I. Introduction

PROTON therapy is an advanced form of radiotherapy which utilizes proton beams to destroy cancer cells. In contrast to photons, protons can deliver their maximum dose to the tumor cells which results in a significant dose reduction to healthy tissue [1]. However, systematic and random errors on the stopping powers can have a significant effect on the estimated dose distribution [2]. For example, a small motion of the tumor inside the patient's body can have a significant effect on the dose distribution, because the stopping power of the tissues traversed by the protons can vary substantially, and even relatively small changes of the trajectory of the protons through the body can have an important effect on the position of the Bragg peak. Consequently, it is important to do adaptive planning to make sure that the treatment is based on an accurate estimate of the tissue compositions and densities.

Proton therapy treatment planning is typically based on stopping power estimation from CT images. To minimize range uncertainties due to anatomical variations, daily CT acquisition would be paramount. Unfortunately, the series of CT scans result in a substantial accumulated patient dose, especially for lung tumors where a 4D CT-scan is required. Low-dose CT is desirable, however, lowering the dose results in a lower signal-to-noise ratio (SNR) and therefore in a reduced image quality. Because these CT scans are not used for tumors delineation, we believe that their required signal-to-noise ratio is lower than that needed for typical radiological images, and therefore there is room for dose reduction. In this research, we aim to determine the lowest exposure scan which still produces sufficient information for proton therapy dose calculations. As argued in [2], the estimated proton stopping powers are a non-linear function of the reconstructed Hounsfield units. Therefore, it is not only important the reconstructed images have a low bias, but also their noise magnitude should be sufficiently small to avoid noise-induced bias in the derived stopping powers. Accordingly, a lower-dose CT simulation tool has been developed and validated to simulate lower-dose scans from an existing standard dose scan. The interested reader is referred to [3] for further details on the proposed model. The simulated lower-dose scan will then be used for treatment planning and the results will be compared with that of the standard-dose scan.

M. Elhamiasl and J. Nuyts are with the Department of Imaging and Pathology, Nuclear Medicine & Molecular imaging, Medical Imaging Research Center (MIRC), KU Leuven, Leuven 3000, Belgium.

S. Teruel Rivas is with the Institut de Recherche Experimentale et Clinique (IREC), Center of Molecular Imaging, Radiotherapy and Oncology (MIRO), Universite catholique de Louvain, Brussels 1200, Belgium.

K. Salvo is with the Department of Radiotherapy, UZ Leuven, Leuven 3000, Belgium.

E. Sterpin is with the Department of Oncology, Laboratory of Experimental Radiotherapy, KU Leuven, Leuven 3000, Belgium.

II. Low-dose CT Simulation

A. Low-dose Simulation Algorithm

Reducing the X-ray tube load is the most effective way of radiation reduction dose to the patients. Tube load reduction can be modeled by estimating the (noise equivalent) number of photons in the high exposure scan and then applying a thinning technique to it (i.e. to randomly eliminate photons using a well-chosen survival probability). Therefore, the log-converted CT raw data should first be converted into the form of transmission data. Assuming a monochromatic beam, the mean number of detected photons in the high dose scan can be written as

$$I^{(\beta)} = I_0^{(\beta)} \exp(-\rho) \tag{1}$$

where the superscript (β) identifies the tube load level in mAs. I_0 and ρ represent incident X-ray intensity and log-converted raw data, respectively. We assume that I_0 depends linearly on the tube load, i.e. $I_0^{(\beta)} = \beta I_0^{(1)}$. Accordingly, estimation of the number of photons from the log-converted data requires a noise-free estimate of $I_0^{(1)}$, the incident X-ray intensity for a unit tube load. The variance of an air scan without any attenuating object can be used to estimate $I_0^{(1)}$. Assuming the CT transmission data as a combination of a Poisson and a

normal distribution, the variance of the transmission data in an air scan is given by

$$\mathrm{var}[\exp(-\rho_{air}^{(\alpha)})] = \mathrm{var}[\frac{I_{air}^{(\alpha)}}{I_0^{(\alpha)}}] = \frac{\alpha I_0^{(1)} + \sigma_e^2}{(\alpha I_0^{(1)})^2} \qquad (2)$$

where α and σ_e^2 represent the tube load and the variance of electronic noise, respectively. $I_0^{(1)}$ and σ_e^2 can be estimated by performing some air scans at different levels of tube load. Therefore, the number of photons in the transmission scan with tube load β mAs can be estimated using (1). Thinning with survival probability α/β can then be applied to produce a (more) noisy low-dose scan of α mAs. Thinning means the random elimination of some of the detected photons where the chance of not being eliminated is the same for each photon. A binomial random generator was used to compute the effect of the thinning. This approach accounts for the fact that the higher exposure scan is already noisy. Electronic noise is modeled as a shifted Poisson distribution and the thinning is then applied to all photons. The thinning also reduces the contribution of the electronic noise, which can be restored by adding a noisy number of additional photons. After that, the final Poisson distribution is shifted back. This yields the following equation:

$$I_{sim}^{(\alpha)} = \mathcal{B}\Big(\beta I_0^{(1)} \exp(-\rho) + \sigma_e^2, \frac{\alpha}{\beta}\Big) + \mathcal{P}\{(1-\frac{\alpha}{\beta})\sigma_e^2\} - \sigma_e^2 \quad (3)$$

The first term reduces the number of photons observed in the converted raw transmission data. The second term adds a noisy number of photons to compensate for the thinning of the electronic noise and the last term subtracts the mean number of electronic noise photons, to ensure that the electronic noise has zero mean.

The relation between beam intensity and the effective number of photons has been derived using blank scans. However, in patient scans, the photon beams reaching the detector are harder than in blank scans, because of the beam hardening effect. A harder beam contains fewer photons for the same intensity and is therefore subject to a higher amount of quantum noise [4]. Knowing the attenuation value and also the bowtie thickness in each beam path and assuming that all patient attenuation is due to water [5], the (noise equivalent) number of photons can be adjusted to account for the beam hardening effect.

In the above, the noise contribution of any detector element was assumed to be independent of that of the other detectors. However, the noise of neighboring CT detector pixels is correlated due to crosstalk between detector pixels [4]. Since the noise added by thinning is uncorrelated, it reduces the noise correlations. It is important to restore this noise correlation in the lower-dose simulated scan because the noise propagation through the image reconstruction is not the same for correlated and uncorrelated noise. To obtain the appropriate correlation in the simulated scan, we convolve the thinning noise with a smoothing mask w, which is designed to impose the desired covariance without changing the variance. It can be proved that w should satisfy $[w \otimes w]_{i-j} = r_{ij}$, where r_{ij} represents the correlation matrix between neighboring pixels.

The CT scanners are usually equipped with a preprocessing software to avoid negative values to be passed to the log when the detected signal is too small [6]. In agreement with this, we observed that the correlation between neighboring pixels increases when the number of photons is below a particular threshold. The number of photons for which the filter must be activated can be determined by analyzing the correlation between neighboring pixels. To obtain a similar result as that produced by the unknown vendor software, we introduced a simple intensity-dependent filter: the detected signal is smoothed by a simple 3×3 top-hat filter and a weighted sum of the original and smoothed views is computed. The intensity-dependent weight is designed to prevent the variance to increase beyond a predefined threshold. In [3] the method is described in more detail.

B. Validation of the Low-dose Simulation

The lower-dose CT simulation model was validated by performing phantom studies. In each experiment, a standard-dose scan along with corresponding lower-dose scans were performed. The standard-dose scan was used as an input to simulate the lower-dose scans. The simulated scans were then reconstructed and compared with corresponding real lower-dose acquired scans.

Two phantom studies were conducted to evaluate the accuracy of the lower-dose simulation tool. The tube voltage was set at 120 kVp in all experiments. In the first experiment, a realistic head phantom (Proton Therapy Dosimetry Head, Model 731-HN) was scanned on Siemens SOMATOM Force scanner with a routine clinical protocol [helical, tube current modulation, Pitch: 0.55]. The high-dose scan of 192 effective mAs was used as the input standard-dose scan to simulate the lower-dose scans. The simulated scans were then reconstructed using Siemens reconstruction software and compared with corresponding real lower-dose images. In the second experiment, the low-dose CT simulation tool was validated on Siemens SOMATOM DRIVE scanner. A water-filled PMMA cylinder with a diameter of 21 cm was scanned using a sequential CT head-neck protocol. The lower-dose scans were simulated from a high-dose scan of 300 mAs and reconstructed using in-house FDK reconstruction.

The standard deviations of the reconstructed image in uniform ROIs and were used for the comparison. Table I compares the noise level for the acquired and the simulated images. The standard deviation of the reconstructed images in uniform ROIs revealed a close agreement of the noise level in the observed and simulated lower-dose scans where the relative error was on the order of $\sim$1% for sequential and $\sim$3% for helical scans. Fig. 1 shows the high-dose scan of 192 mAs and its corresponding measured and simulated lower-dose scans. It can be seen that the noise level in the reconstructed images has been increased by reducing the tube load. In [3], [7], [8] additional validation experiments are presented.

III. Effect of CT Dose Reduction on Dose Calculations

The high-dose and simulated lower-dose scans of the realistic head phantom were used to investigate the effect of CT dose

TABLE I
THE DETAILED COMPARISON OF NOISE LEVELS OF THE ACQUIRED AND THE SIMULATED LOW-DOSE IMAGES AT DIFFERENT LEVELS OF TUBE LOAD FOR HEAD AND 21 CM WATER PHANTOMS. THE STANDARD DEVIATION OF THE NOISE IN UNIFORM ROIS WERE USED FOR COMPARISON.

head phantom				21 cm water phantom						
	region 1				center			periphery		
Tube load	real	sim	error	Tube load	real	sim	error	real	sim	error
192	3.38	-	-	300	25.46	-	-	22.19	-	-
117	4.25	4.23	0.2%	250	28.17	27.82	0.8%	25.02	24.91	0.4%
93	4.80	4.68	2.5%	200	31.03	30.96	0.2%	27.88	27.79	0.3%
70	5.48	5.41	1.2%	150	36.07	35.79	0.7%	32.24	32.25	0.0%
47	6.76	6.56	2.9%	100	44.12	44.12	0.0%	39.55	39.70	0.4%
24	9.48	9.18	3.1%	50	62.83	63.66	1.3%	55.94	57.25	2.3%
20	10.44	9.95	4.6%	20	85.16	83.88	1.5%	81.78	80.84	1.1%

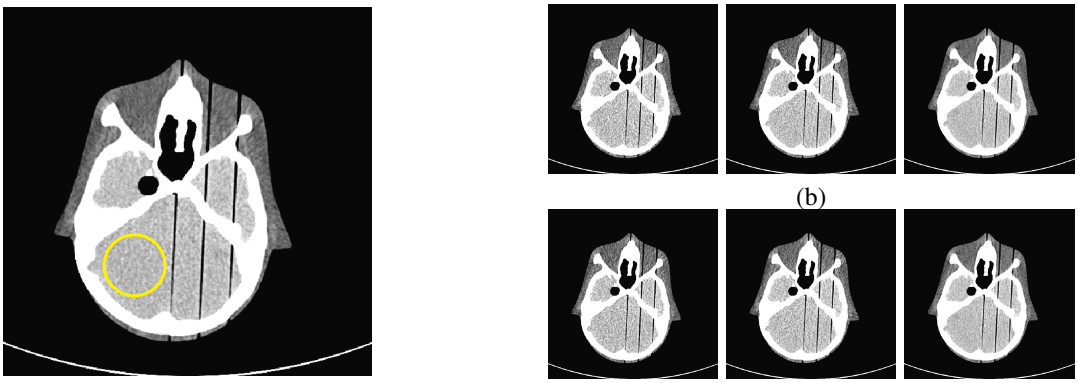


Fig. 1. Presentation of (a) the measured high-dose image of 192 effective mAs and (b) corresponding acquired and (c) simulated lower-dose images of 20, 24, and 47, effective mAs (from left to right) using the proposed framework (window level: 30 HU, window width: 100 HU). The yellow circle represents the uniform ROI.

reduction on proton therapy dose calculations. The standard-dose scan of 192 mAs was used for dose calculation and considered as the ground truth. A small region was defined as the tumor and a treatment plan was generated. The prescription dose was 70 Gy to Planning Target Volume (PTV) and reoptimized on every CT with Monte Carlo simulation, using a 3 mm dose grid resolution, 2 beams 90° (left) and 270° (right) beam. The simulations were performed on RayStation 8B(R) SP1, using the beam model of the clinical beam line of UMCG (University Medical Center Groningen). Subsequently, dose distributions were recalculated for the simulated lower-dose images and compared with the ground truth.

The Dose Volume Histogram (DVH) metrics for the standard-dose scan and the simulated scans are reported in Table II. Fig. 2(a) represents the DVH curve for the high-dose scan of 192 mAs and the real and the simulated scans of 20 mAs. It can be seen that the dose reduction by a factor up to 10 has a negligible effect on the dose calculation where the DVH curve of the standard-dose and the lower-dose scans are on top of each other. The difference between the DVH curves of the high-dose scan and the simulated lower-dose scans, for the same volume%, have been presented in Fig. 2(b). Comparing the high-dose scan of 192 mAs with the lower-dose scan of 20 mAs shows that the maximum difference between the DVH curves is about 0.836 Gy which only happens in 2% of CTV volume. Fig. 3 represents the dose distribution maps for the high-dose scan of 192 mAs and the real and the simulated scans of 20 mAs.

IV. CONCLUSION

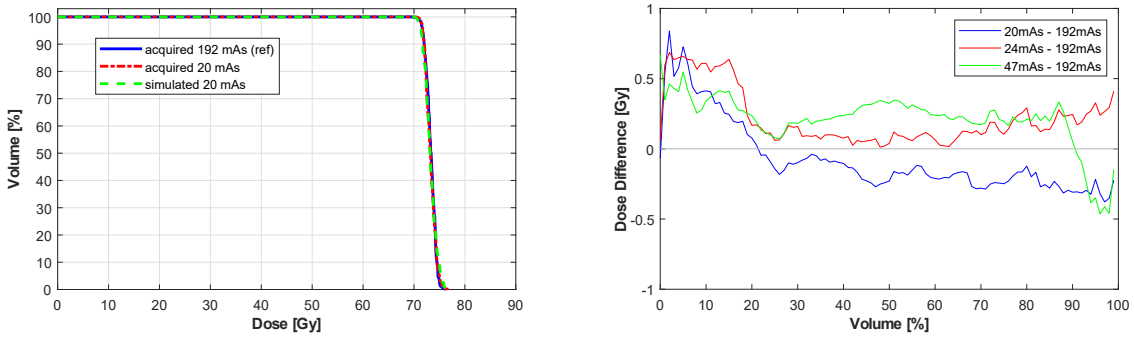
In this paper, we investigated the effect of CT dose reduction on proton therapy treatment planning. A lower-dose CT simulation tool was used to simulate low-dose CT scans from a standard-dose scan. The bowtie filter, the electrical noise, the correlated noise, the beam hardening effect, and the signal-dependent filter were included in the proposed model. The simulated lower-dose scans were used for proton therapy treatment planning. Our preliminary results show that the CT dose reduction does not have a significant effect on the estimated dose distributions. A more systematic and quantitative analysis of the effect of dose reduction on proton therapy treatment planning for more complex inhomogeneous anatomical structures is ongoing.

ACKNOWLEDGMENT

This project is supported by Fonds Baillet-Latour. The authors would like to warmly thank Xavier Geets for the initial design of the research, Walter Coudyzer for his contribution in data collection and Karl Stierstorfer and Frederic Noo for their valuable comments and suggestions.

TABLE II
THE DETAILED COMPARISON OF DVH METRICS FOR THE ACQUIRED STANDARD-DOSE AND THE SIMULATED LOWER DOSE SCANS.

Tube load	PTV D99	PTV D98	PTV D95	Average	PTV D50	PTV D2	PTV D1
192 (REF)	70.95	71.41	71.62	73.31	73.37	75.04	75.37
140 (sim)	70.91	71.22	71.44	73.24	73.24	75.61	76.36
117 (sim)	71.05	71.41	71.68	73.53	73.54	75.72	76.05
93 (sim)	71.10	71.64	71.77	73.37	73.32	75.40	75.72
70 (sim)	70.99	71.26	71.67	73.30	73.38	74.85	75.21
47 (sim)	70.80	70.95	71.27	73.51	73.69	75.51	75.72
24 (sim)	71.36	71.71	71.95	73.53	73.41	75.73	75.97
20 (sim)	70.72	71.06	71.41	73.24	73.14	75.88	75.90
20 (real)	71.19	71.33	71.70	73.26	73.25	75.49	75.67



(a) (b)

Fig. 2. An illustration of the effect of CT dose reduction on the DVH curves. (a) Presentation of the DVH curves for the high-dose scan of 192 mAs, the acquired low-dose scan of 20 mAs, and the simulated scan of 20 mAs. (b) Demonstration of the difference between the DVH curve of the high dose scan and the simulated lower-dose scans.

(a) (b) (c)

Fig. 3. Dose distribution maps corresponding to (a) acquired low-dose scan of 20 mAs, (b) simulated low-dose scan of 20 mAs, and (c) acquired standard-dose scan of 192 mAs (ground truth).

REFERENCES

[1] H. Paganetti, *Proton Beam Therapy*, ser. 2399-2891. IOP Publishing, 2017. [Online]. Available: http://dx.doi.org/10.1088/978-0-7503-1370-4

[2] S. Brousmiche, K. Souris, J. O. de Xivry, J. A. Lee, B. Macq, and J. Seco, "Combined influence of ct random noise and hu-rsp calibration curve nonlinearities on proton range systematic errors," *Physics in Medicine & Biology*, vol. 62, no. 21, p. 8226, 2017.

[3] M. Elhamiasl and J. Nuyts, "Low-dose x-ray CT simulation from an available higher-dose scan," *Physics in Medicine & Biology*, 2020. [Online]. Available: http://iopscience.iop.org/10.1088/1361-6560/ab8953

[4] B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson, "Properties of preprocessed sinogram data in x-ray computed tomography," *Medical physics*, vol. 33, no. 9, pp. 3290–3303, 2006.

[5] T. M. Benson and B. K. De Man, "Synthetic ct noise emulation in the raw data domain," in *IEEE Nuclear Science Symposuim & Medical Imaging Conference*. IEEE, 2010, pp. 3169–3171.

[6] L. Yu, M. Shiung, D. Jondal, and C. H. McCollough, "Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols," *Journal of computer assisted tomography*, vol. 36, no. 4, pp. 477–487, 2012.

[7] M. Elhamiasl, K. Salvo, W. Coudyzer, and J. Nuyts, "Low-dose CT simulation from an available higher dose CT scan," in *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2019, pp. 1–3, ISSN: 2577-0829.

[8] M. Elhamiasl and J. Nuyts, "Simulating lower-dose scans from an available CT scan," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 11072. International Society for Optics and Photonics, 2019, p. 110720X.

Deep Learning-Based Real-Time Estimation of Patient Dose Distributions for Various Medical CT Scan Protocols

Sarah Muller, Joscha Maier, Elias Eulig, Michael Knaup,
Stefan Sawall, and Marc Kachelrieß

***Abstract*—With the increased number of CT examinations in clinical practice on the one hand and the potential risk of ionizing radiation on the other, reporting patient-specific CT dose estimates becomes more and more important. However, conventional metrics such as the volume CT dose index (CTDIvol) or the dose-length product (DLP) are rather based on phantom measurements and do not appropriately account for the patient's anatomy. To incorporate the latter, several approaches have been proposed. Here, the gold standard is to perform a Monte Carlo (MC) simulation modeling the physics of the CT dose deposition. Being computationally expensive, MC simulations cannot be used routinely while faster alternatives have the downside of a reduced accuracy. To combine both, accuracy and computational performance, the so-called deep dose estimation (DDE) was proposed recently by Maier et al. [1]. It uses a deep convolutional neural network that is trained to reproduce MC simulations given a first-order dose estimate and a CT image as input. While the first experiments were restricted to 120 kV circle scans, this study provides a more extensive evaluation of DDE including its application to different tube voltages, circular and spiral trajectories as well as tube current modulation and shaped filters. For any case, DDE yields dose estimates that differ by less than 3 % from the ground truth MC reference while being orders of magnitudes faster.**

***Index Terms*—CT dose estimation, convolutional neural network, Monte Carlo simulation, deep learning**

I. Introduction

CT examinations are widely used in medical practice. Their gain of importance in the clinical environment came along with the upgrade in resolution and speed due to the development of multi-row detectors and fast spiral CT scans [2]. Due to the increased number of CT examinations, the potential risks of ionizing radiation came into focus and the research in patient dose estimation was fueled. Some of the then developed and most commonly used metrics are the weighted CT dose index (CTDIw), the CT dose index averaged over the volume (CTDIvol) and the dose-length product (DLP). Referring to the dose absorbed in cylindrical phantoms, these metrics have the drawback of not being able to account for the patient anatomy, and therefore, provide only a rough estimate of the actual patient dose. Today, the gold standard for an accurate patient-specific dose estimation is Monte Carlo (MC) simulation which has, however, the substantial disadvantage of being computationally expensive. Alternatives which are faster, but pay with accuracy, are based on a database of MC simulated organ doses of different phantoms [3], or reconstruct the absorbed dose directly from CT images using analytical approximations [4]. Recently, we had demonstrated that a fast and accurate solution can be achieved by using a deep convolutional neural network (DCNN) that is trained to reproduce MC dose estimates given a CT image and a first-order dose estimate (FOD) as input [1]. In the present work, we extend our former study [1] of the so-called deep dose estimation (DDE) and analyze its potential to generalize to different anatomies, different tube voltages as well as different acquisition protocols.

Sarah Muller, Dr. Joscha Maier, Elias Eulig, Dr. Stefan Sawall, and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Ruprecht-Karls University, Heidelberg, Germany.

Dr. Michael Knaup is with the German Cancer Research Center (DKFZ), Heidelberg, Germany.

Corresponding author: sarah.muller@dkfz.de

II. Methods

A. First-order dose deposition

We consider the first-order dose deposition as the energy per mass that is transferred to electrons by the photoelectric effect and single Compton scattering. Here, it is given as input to the DDE network to provide additional information that is not directly encoded in the CT image such as the bow-tie filter or the start angle of the x-ray tube. Given the differential photon fluence $d^2N/(d\Omega dE)$, the first-order dose D_1 deposited in a volume element dV at position $\boldsymbol{r}$ can be calculated as:

$$D_1(\boldsymbol{r}) = \frac{dV}{\rho(\boldsymbol{r})dV} \int \frac{d^2N}{d\Omega dE} \sum_{i\in\{\text{PE , CS}\}} P_i(\boldsymbol{r}, E) \cdot E_i(E)\, dE, \quad (1)$$

with ρ denoting the mass density, P_{PE} denoting the probability density for a photoelectric effect (PE) and P_{CS} describing the probability density for a Compton scattering (CS) interaction. With the line integral $\int_0^r \mu(\boldsymbol{s}+r'\cdot\boldsymbol{t}, E)dr'$ over the attenuation coefficient μ from the x-ray source position $\boldsymbol{s}$ to the position of the volume element $\boldsymbol{r}$, the latter is given as:

$$P_i(\boldsymbol{r}, E) = \mu_i(\boldsymbol{r}, E) \cdot e^{-\int_0^r \mu(\boldsymbol{s}+r'\cdot\boldsymbol{t}, E)dr'}. \quad (2)$$

Here, μ_{PE} and μ_{CS} describe the attenuation caused by the photoelectric effect (PE) and by Compton scattering (CS), respectively. In case of a photoelectric interaction, the whole energy E is assumed to be deposed locally. Thus,

$$E_{\text{PE}}(E) = E. \quad (3)$$

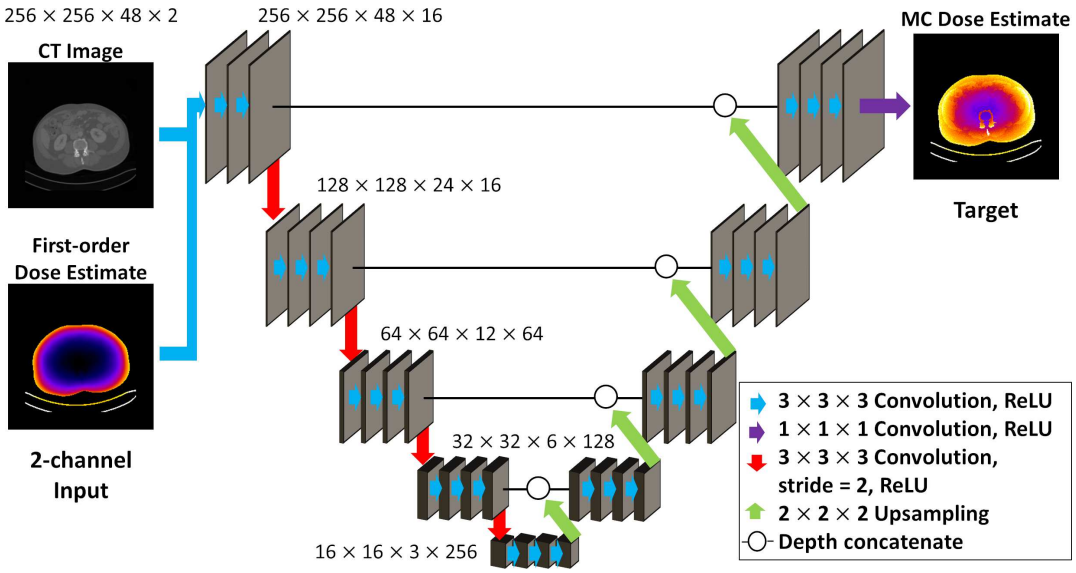


Fig. 1: Architecture of the DDE network adapted from reference [1]. A CT reconstruction and a first-order dose estimate are given as a 2-channel input to the network. The network is trained to predict patient-specific dose estimates using MC simulations as labels.

Considering Compton scattering, the energy transfer for a certain scatter angle ϑ is given by the Klein-Nishina formula while the angular distribution is given by the differential cross section $d\sigma/d\Omega$. Integration over the solid angle yields the mean energy transfer:

$$E_{\mathrm{CS}}(E) = \int \frac{d\sigma}{d\Omega}(E) \left(E - \frac{E}{1 + \frac{E}{m_e c^2}(1 - \cos(\vartheta))} \right) d\Omega. \tag{4}$$

For the sake of simplicity and computational performance, the patient is assumed to have a water-equivalent x-ray attenuation characteristic. Thus, the calculation of the first-order dose deposition can be performed using a precalculated look-up table that evaluates equation (1) for a given number of water-equivalent intersection lengths.

B. Deep Dose Estimation

1) Data generation: For our analysis, we use clinical CT scans (Siemens Somatom Definition Flash) of 35 patients as prior for the generation of simulated data. Circular and spiral CT acquisitions were simulated for different z-positions and tube voltages (80 kV, 100 kV and 120 kV) assuming the geometry of the Siemens Somatom Definition Flash CT scanner. All patients were simulated with and without bow-tie filtration with the bow-tie chosen according to reference [5]. Additionally, scans with bow-tie and angular tube current modulation were simulated. The tube current modulation was chosen according to reference [6], [7]. For each sample, a first-order dose estimate and a MC dose estimate were computed on a 256 × 256 × 128 voxel grid with an isotropic voxel size of 2 mm. We separated the 35 patients into 28 patients for training and 7 patients for validation. In total, we have 4006 training samples and 1185 samples for testing.

2) Model and training: DDE is implemented using a 3D U-net architecture as depicted in figure 1 [8]. Unlike in the original U-net [8], here, the downsampling is achieved by a 3 × 3 × 3 convolution with a stride of 2. This results in a reduction of the dimensions by a factor of 2 after every downward path. As input to the network, we use a water-precorrected CT reconstruction and the first-order dose estimate (see section II-A). The ground truth and target for the network is the corresponding MC dose distribution that was generated with our in-house MC software [9].

DDE was implemented using the Keras framework and the TensorFlow backended. The training was performed on an Nvidia Geforce RTX 2080TI GPU using the Adam optimizer and a batch size of one. As loss function, a weighted mean absolute percentage error (WMAPE) was used. Thereby, a twenty-fold weight was assigned to all bone voxels, to compensate for bone being underrepresented in all CT scans.

III. Results

The results are analyzed by comparing the DDE predictions of the validation data to the ground truth MC simulation in terms of the mean absolute percentage error (MAPE). Exemplary results for 120 kV abdominal scans acquired with a circular scan protocol with and without bow-tie filtration are shown in figure 2 (middle row). Here, DDE is able to provide dose distributions that are almost equal to the MC simulation. This also holds true outside of the z-collimation where the first-order dose is zero. Slightly higher deviations arise at the edges of high contrast structures, probably due to the blurring induced by the use of a pixel-wise loss function.

The ability of DDE reproducing MC dose estimates for 120 kV circle scans has already been shown by reference [1]. This work extends the study by simulating, additionally to the circle scan, a spiral scan protocol with a pitch of 1.0 (see

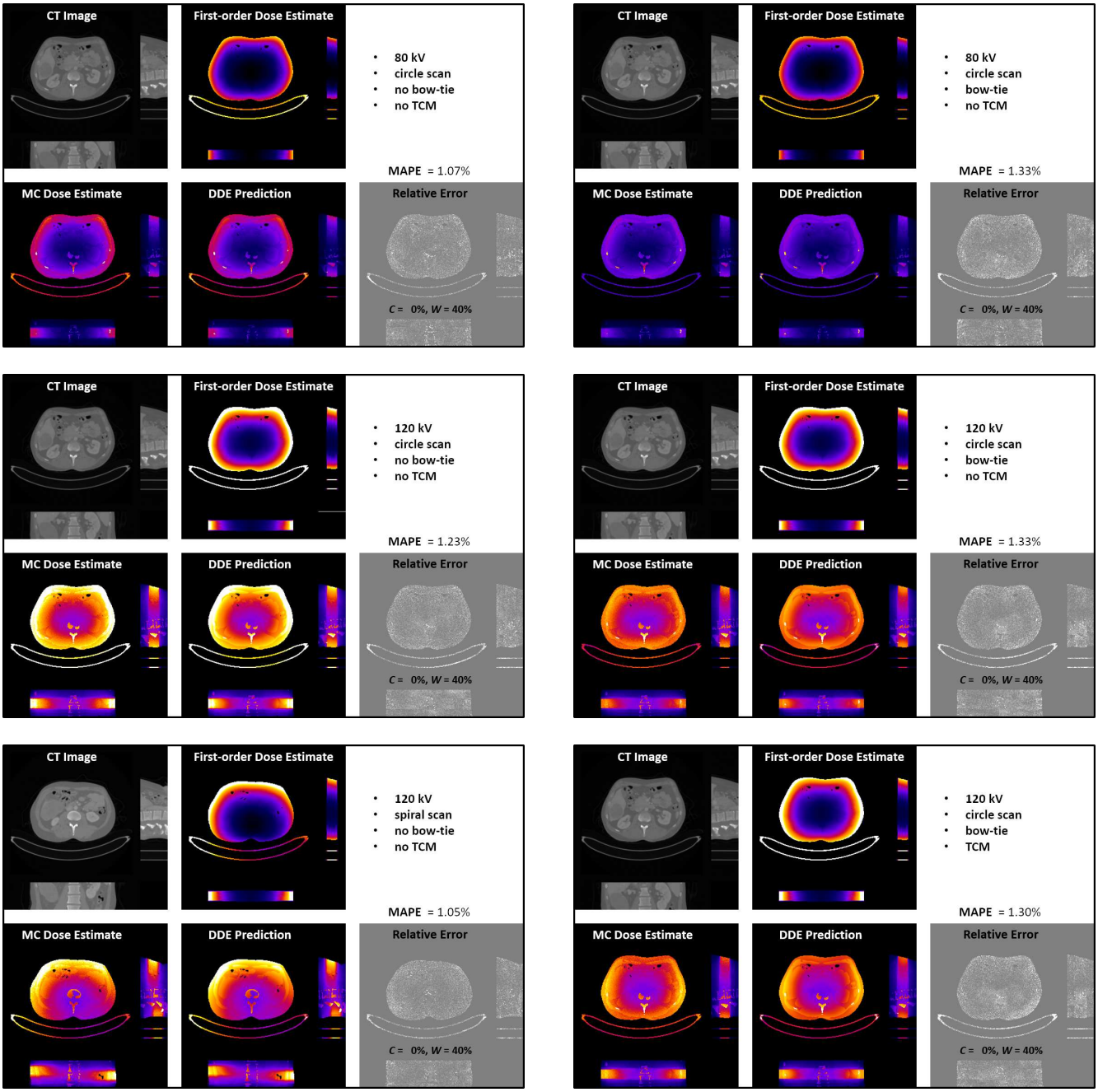


Fig. 2: Input (CT image and first-order dose estimate), prediction, ground truth (MC dose estimate) and the relative error computed between the prediction and the MC ground truth in axial, sagittal and coronal view for different tube voltages, scan modes, with and without shaped filter and tube current modulation (TCM).

figure 2, lower left). The comparison shows that for the spiral CT acquisitions we achieve equally good results as for the circle scans. Unlike in the former study, the network is trained to handle different tube voltages (80 kV, 100 kV and 120 kV). Exemplary samples of 80 kV circle scans are presented in figure 2 (top row). A result with tube current modulation, bow-tie filtration and a tube voltage of 80 kV can be found in the lower right box of figure 2. We observe that the network is able to predict the MC dose estimates with high accuracy. This holds for all tested combinations of parameters. The evaluation of all 1185 validation samples results in a MAPE of 3.33%.

In order to find out if the network is able to reproduce the MC dose estimates independent of the parameter selection, we analyze the error in partitions of the validation data grouped by the parameters tube voltage, anatomical region, scan mode, prefiltration (on / off) and tube current modulation (on / off). The errors are reported in table I. Note, the error is computed in the whole volume where air voxels are set to zero. Grouping the 1185 validation samples into the different voltages (80 kV, 100 kV and 120 kV), e.g. the group “80 kV” contains all

scans simulated with a tube voltage of 80 kV regardless of the configuration of the other parameters. The other groups are created accordingly. An exception is the error calculation for the tube current modulation (on / off), in table I, which only considers samples simulated with bow-tie filter, inasmuch as tube current modulation has only been simulated together with prefiltration.

No noticeable difference in terms of the error could be observed between the tube voltages in table I. This implies that the network is able to predict different ranges of values with equal accuracy. The network is also able to handle different anatomies (thorax, abdomen and pelvis) with a slightly higher error for pelvis. This can be explained with the aforementioned deviations at high contrast structures (bone). Hence, the proportion and location of bones and organs does not significantly affect the accuracy.

As it becomes apparent in the comparison of the circle scans with and without bow-tie filtration (see figure 2, upper and middle row), the network is able to reproduce the effects of dose reduction. This can also be observed in the row "Bow-tie Filtration" in table I, where we get an error of 3.47%. The error for the group of samples with bow-tie and tube current modulation is 4.13%.

Overall, variations in the dose distribution introduced by different tube voltages, scan modes, tube current modulation and prefiltration do not affect the correctness of the prediction, as the ranges of accuracy in table I remain near the overall error of 3.33% on the validation set. This underlines the fact that for all tested combinations of parameters we obtain high accuracy. Furthermore, we do not observe major differences for the different parameter groups.

IV. Conclusion

This study extends the former analysis [1] by the usage of different scan modes and tube voltages. It shows that we are able to derive accurate dose estimates for the circle and the spiral trajectory of a third generation CT system. Furthermore, the network is able to reproduce the MC dose estimates for different tube voltages (80 kV, 100 kV and 120 kV), different anatomies (thorax, abdomen and pelvis), with and without shaped filters and tube current modulation. In the different experiments, we obtain similar dose estimates as the MC simulations and after the network is trained (for $\approx$ 3 d for 65 epochs and 4006 training samples), we receive the dose distribution of a $256 \times 256 \times 48$ voxel volume within 250 ms.

TABLE I: Error reported for different groups of simulation parameters.

Group	Configuration	MAPE
Tube Voltage	80 kV	3.24%
	100 kV	3.13%
	120 kV	3.53%
Anatomical Region	Abdomen	3.30%
	Pelvis	4.48%
	Thorax	3.26%
Scan Mode	Circle	3.03%
	Spiral	3.66%
Tube Current Modulation	Off	3.14%
	On	4.13%
Bow-tie Filtration	Off	3.11%
	On	3.47%

Acknowledgment

The Monte Carlo software RayConStruct-MC® and parts of the reconstruction software were provided by RayConStruct GmbH, Nürnberg, Germany.

References

[1] J. Maier, E. Eulig, S. Dorn, S. Sawall, and M. Kachelrieß, "Real–time patient–specific CT dose estimation using a deep convolutional neural network," *IEEE Medical Imaging Conference Record*, pp. M–03–178, Nov. 2018.

[2] W. A. Kalender, "Dose in x-ray computed tomography." *Physics in Medicine and Biology*, vol. 59, no. 3, pp. R129–50, 2014.

[3] A. Ding, Y. Gao, H. Liu, P. F. Caracappa, D. J. Long, W. E. Bolch, B. Liu, and X. G. Xu, "VirtualDose: a software for reporting organ doses from CT for adult and pediatric patients," *Physics in Medicine and Biology*, vol. 60, no. 14, pp. 5601–5625, 2015.

[4] B. De Man, M. Wu, P. FitzGerald, M. Kalra, and Z. Yin, "Dose reconstruction for real-time patient-specific dose estimation in CT." *Medical Physics*, vol. 42, no. 5, pp. 2740–2751, 2015.

[5] M. D. Harpen, "A simple theorem relating noise and patient dose in computed tomography," *Med. Phys.*, vol. 26, no. 11, pp. 2231–2234, Nov. 1999.

[6] M. Gies, W. A. Kalender, H. Wolf, C. Suess, and M. T. Madsen, "Dose reduction in CT by anatomically adapted tube current modulation. I. Simulation studies," *Med. Phys.*, vol. 26, no. 11, pp. 2235–2247, Nov. 1999.

[7] W. A. Kalender, H. Wolf, and C. Suess, "Dose reduction in CT by anatomically adapted tube current modulation. II. Phantom measurements," *Medical physics*, vol. 26, no. 11, pp. 2248–2253, 1999.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, may 2015.

[9] M. Baer and M. Kachelrieß, "Hybrid scatter correction for CT imaging," *Physics in Medicine & Biology*, vol. 57, no. 21, p. 6849, 2012.

Joint Spectrum Estimation and Image Reconstrution in Spectral CT

Shaojie Chang, Qiong Xu, Xi Chen, and Xuanqin Mou

***Abstract*—With the development of photon counting detector (PCD), spectral computed tomography (CT) has been a promising technique due to its ability of material decomposition with narrow energy windows. In spectral CT reconstruction, X-ray spectrum plays a very important role and is difficult to be measured directly in practice. Furthermore, due to the nonideal effects of PCD response, x-ray spectrum of different energy windows are caused to overlap, which will lead error in material decomposition. In order to reduce error in material decomposition, we propose a joint spectrum estimation and image reconstruction algorithm for spectral CT. The proposed method combines spectrum estimation based on model spectra response functions, basis material fraction image reconstruction and regularization for noise suppression into an optimization framework. The widely-used block coordinate descent (BCD) method is used to resolve the multi-variable optimization problem of simultaneously obtaining the spectra and images in iterations. Simulations with a PCD are performed to evaluate the proposed method. Experimental results show the advantages of joint spectrum estimation and image reconstruction method in improving the accuracy of the estimated spectrum for each energy window and the decomposed basis material images including K-edge material.**

***Index Terms*—Spectral computed tomography, spectrum estimation, photon counting detector, iterative reconstruction.**

I. Introduction

BENEFIT from the development of photon counting detectors (PCDs), spectral computed tomography (SpCT) can simultaneously obtain the number of incident photons at multiple energy windows. By taking use of spectral data, energy-independent basis material fraction images can be reconstructed to realize the quantitative material decomposition of the scanned object. Spectral CT has been considered a promising technique in many clinical and industrial applications such as iodine quantification, characterize kidney stone and security inspection.

Methods for SpCT basis material reconstruction can be classified into three categories: projection domain methods, image domain methods and projection-image domain methods [1, 2]. In this work, we focus on projection-image domain methods, which directly reconstruct basis material fraction images from energy window measurements. Considering the energy-independent information and modeling data acquisition geometry, projection-image domain methods greatly improve decomposition accuracy, and several iterative algorithms have been proposed. However, all the above projection-image domain methods need the x-ray spectrum in material decomposition, which is always obtained by certain spectrum estimation method in advance.

SpCT reconstruction is highly sensitive to the estimated spectrum mismatch, which significantly affects the basis material decomposition accuracy. For example, Wu *et al.* reported that the accuracy of the estimated spectrum can compromise the accuracy of the final material reconstructions in SpCT, and the energy spectrum mismatch is the largest limitation of the state-of-the-art material reconstruction algorithms in practical applications [3]. Hence, in our previous work [1], we developed a spectrum estimation-guided (SEG) approach to eliminate spectrum mismatch in iterations, which could simultaneously estimate spectrum and reconstruct the basis material fraction images directly from dual SpCT projection data. The proposed SEG method has shown the advantages of jointly estimating the x-ray spectra and reconstructing the image in improving the accuracy of spectrum estimation and material decomposition. And the detector response in the SEG method is assumed to be ideal.

However, in practice, the energy response of PCDs is not ideal due to numerous factors, such as pulse pileup, charge sharing, K-escape x-rays, Compton scattering, and others. These nonideal effects will lead to error in recorded counts and cause the energy windows spectra to overlap. Hence, several approaches are proposed to model and compensate for these effects of the PCD based spCT. For instance, Schlomaka *et al.* measured the detector spectral response functions using monochromatic synchrotron radiation at several different primary energies and derived a phenomenological model of the complete spectral response [4]. Such monoenergetic calibration measurements are time consuming. And Schmidt *et al.* proposed a constrained one-step spectral CT reconstruction (cOSSCIR) method to reconstruct basis material maps while modeling the nonlinear x-ray detection process and enforcing convex constraints on the basis map images [2]. However, it needs a extra calibration procedure to estimate effective energy-window spectra in implementation.

In order to avoid extra spectra measurement and spectra mismatch, we propose a joint spectrum estimation and image reconstruction method for SpCT. This method is an extension of our previous work SEG method [1] from DECT to PCD based SpCT, which takes multi-energy windows and nonideal effects of PCD detector into consideration, as well including K-edge material. In this work, we first model the effective energy-window x-ray spectrum by introducing an empirical spectral response function modeling the nonideal effects of the PCD response into x-ray spectrum representation. As our previous work [1], the effective energy-window x-ray spectrum can also be expressed by a linear combination of 6 spectrum coefficients with known model spectra. Hence, our previously proposed SEG method [1] is adopted to simultaneously resolve the spectra parameters of each energy window and

This work was supported in part by National Key Research and Development Program of China (No. 2016YFA0202003), National Natural Science Foundation of China (NSFC) (No. 11975250).

Shaojie Chang, Xi Chen and Xuanqin Mou are with the Institute of Image processing and Pattern recognition, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China, and Qiong Xu is with Beijing Engineering Research Center of Radiographic Techniques and Equipment, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China.

decomposed basis material images including K-edge material. Experimental results in this study show that the proposed SpCT reconstruction approach can accurately estimate the effective energy-window spectrum while realizing high quality multi-material decomposition without knowing the spectrum information in advance.

The remainder of this paper is organized as follows. Section II presents our proposed method in detail. In Section III, we scheme the evaluation arrangement of numerical simulations and results. We will discuss some related issues of this study and concludes this article in Section IV.

II. Methods and Materials

A. Forward projection model of SpCT

Considering the K-edge material imaging in SpCT, a linear attenuation coefficient (LAC) can be represented by M types of basis material, e.g. water, bone, Iodine and so on. The LAC function $\mu_j(\varepsilon)$ at the j^{th} pixel of the image is decomposed into two parts:

$$\mu_j(\varepsilon) = \sum_{m=1}^{M} \mu_m(\varepsilon) f_{mj}, \tag{1}$$

where $\mu_m(\varepsilon)$ is the LAC of material m at energy ε, f_{mj} denotes the fraction value at j^{th} pixel of basis material m. M is the total number of material types. In this work, $M = 3$

Hence, considering the basis material decomposition, the detected photons of SpCT at PCD can be formulated as follows:

$$\hat{N}_{wi} = \sum_{\varepsilon} N_{wi}(\varepsilon) \exp\left(- \sum_{mj} r_{ij} \mu_m(\varepsilon) f_{mj} \right), \tag{2}$$

where $\hat{N}_{wi}$ is the detected photon number for x-ray length i in energy window w. $N_{wi}(\varepsilon)$ denotes the photons at energy ε in energy window w. $\mathbf{R} = \{r_{i,j}\} \in \mathbb{R}^{I \times J}$ is the system matrix, $i = 1, ..., I$, $j = 1, ..., J$, where I is the total number of measurements in the scan and J is the total number of pixels. And the total number of photons without an object is expressed as:

$$N_{wi}^0 = \sum_{\varepsilon} N_{wi}(\varepsilon). \tag{3}$$

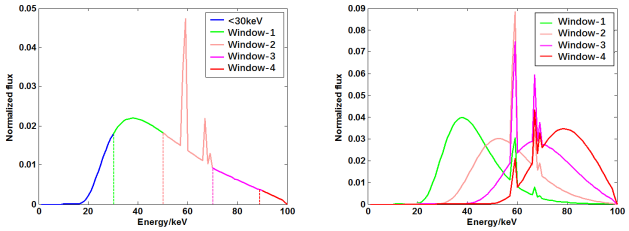
Substituting Eq.(3) into Eq.(2), the normalized photon flux is given as follows:

$$\hat{Y}_{wi} = \sum_{\varepsilon} S_w(\varepsilon) \exp\left(- \sum_{mj} r_{ij} \mu_m(\varepsilon) f_{mj} \right), \tag{4}$$

where $\hat{Y}_{wi} = \frac{\hat{N}_{wi}}{N_{wi}^0}$ $S_w(\varepsilon) = \frac{N_{wi}(\varepsilon)}{N_{wi}^0}$ denotes the normalized photo flux and x-ray spectrum, respectively.

B. Spectrum response function model for PCD

In an energy window, the photon counts can be calculated by subtracting the number of photons between two set energy thresholds. In a PCD with ideal spectral response, the energy-window spectra are separated and non-onverlapping, as shown in Fig.1(a). In practice, due to the effect of pulse pileup, charge sharing and other nonideal effects, the realistic energy-window spectra are overlapped as shown in Fig.1(b). As stated in [4],



(a) (b)

Fig. 1. Ideal and realistic energy-window spectra with thresholds at 30keV, 50keV, 70keV and 90keV. (a) Ideal. (b) Realistic.

Schlomka *et al.* established an empirical model of PCD detector response through experiments and pointed that the detector response presents a Gaussian distribution near the real photon energy. Hence, considering PCD detector response, spectrum response function $\phi_w(\varepsilon)$ in different energy windows can be expressed as Eq.(5), which integrates the x-ray spectrum and detector response in the corresponding energy window.

$$\phi_w(\varepsilon) = \sum_{\varepsilon' = \varepsilon_l^w}^{\varepsilon_h^w} S_0(\varepsilon) R\left(\varepsilon', \varepsilon\right), \tag{5}$$

where ε_l^w and ε_h^w denote the low- and high- threshold for the current energy window. $S_0(\varepsilon)$ represents the emitted spectrum of x-ray source. $R\left(\varepsilon', \varepsilon\right)$ is the detector response of PCD defined as follows:

$$R\left(\varepsilon', \varepsilon\right) = N(\varepsilon) \left(\frac{1}{\sqrt{2\pi}\sigma_1(\varepsilon)} \exp\left(-\frac{1}{2} \frac{\left(\varepsilon' - \varepsilon\right)^2}{\sigma_1(\varepsilon)^2} \right) \right), \tag{6}$$

where $N(\varepsilon)$ is a constant for normalization, which is decided by the incident photon energy. ε and ε' denote the incident photon energy and the photon energy recorded by PCD, respectively. $\sigma_1(\varepsilon)$ describes photon peak, which is validated by the experiment in [4] as a linear relationship with ε as follows,

$$\sigma_1(\varepsilon)^2 = k_e \varepsilon. \tag{7}$$

Where k_e is set to 0.211 and determined by the fitting data of experiment.

C. Joint Spectrum estimation and image reconstruction framework

Based on our previous work [1], the spectrum of an x-ray tube can be represented as a weighted summation of a set of known spectra with different thicknesses of the tube filter. Thus, $S_0(\varepsilon)$ is expressed as $S_0(\varepsilon) = \sum_{k=1}^{K} c_k \hat{S}_k(\varepsilon)$, where c_k denotes the spectrum weight on the k^{th} model spectrum $\hat{S}_k(\varepsilon)$ and K is the number of known model spectra. Hence, the spectrum response function of PCD for energy window w is given by the following:

$$\begin{aligned} \phi_w(\varepsilon) &= \sum_{\varepsilon' = \varepsilon_l^w}^{\varepsilon_h^w} \sum_{k=1}^{K} c_k \hat{S}_k(\varepsilon) R\left(\varepsilon', \varepsilon\right) \\ &= \sum_{k=1}^{K} c_k \sum_{\varepsilon' = \varepsilon_l^w}^{\varepsilon_h^w} \hat{S}_k(\varepsilon) R\left(\varepsilon', \varepsilon\right) \\ &= \sum_{k=1}^{K} c_k \phi_w^k(\varepsilon). \end{aligned} \tag{8}$$

Where $\phi_w^k(\varepsilon) = \sum_{\varepsilon'=\varepsilon_l^w}^{\varepsilon_h^w} \hat{S}_k(\varepsilon) R\left(\varepsilon', \varepsilon\right)$ represents the spectrum response function with the k^{th} model spectrum.

Substituting Eq.8 into 4 by $S_w(\varepsilon) = \phi_w(\varepsilon)$, the current projection measurement $\hat{Y}_{wi}$ can be expressed by the spectrum weights vector $\boldsymbol{C}$ and basis material fraction images vector $\boldsymbol{f}$ as follows:

$$\begin{aligned}\hat{Y}_{wi}(\boldsymbol{C}, \boldsymbol{f}) &= \sum_{\varepsilon} \phi_w(\varepsilon) \exp\left(-\sum_{mj} r_{ij}\mu_m(\varepsilon) f_{mj}\right) \\ &= \sum_{\varepsilon}\sum_{k=1}^{K} c_k \phi_w^k(\varepsilon) \exp\left(-\sum_{mj} r_{ij}\mu_m(\varepsilon) f_{mj}\right) \\ &= \sum_{k=1}^{K} c_k \sum_{\varepsilon} \phi_w^k(\varepsilon) \exp\left(-\sum_{mj} r_{ij}\mu_m(\varepsilon) f_{mj}\right) \\ &= \boldsymbol{C}^w \mathcal{P}_{wi}(\boldsymbol{f}),\end{aligned} \tag{9}$$

where

$$\mathcal{P}_{wi}(\boldsymbol{f}) = \begin{bmatrix} \sum_{\varepsilon} \phi_w^1(\varepsilon) t(\boldsymbol{f}, \varepsilon) \\ \vdots \\ \sum_{E} \phi_w^K(\varepsilon) t(\boldsymbol{f}, \varepsilon) \end{bmatrix}, \tag{10}$$

$$t(\boldsymbol{f}, \varepsilon) = \exp(-\sum_{mj} r_{ij}\mu_m(\varepsilon)\boldsymbol{f}_{mj}). \tag{11}$$

The spectrum weights vector $\boldsymbol{C}$ is given by $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}^1 \\ \vdots \\ \boldsymbol{C}^W \end{bmatrix}$, and for energy window w, $\boldsymbol{C}^w = [c_1, ..., c_k] \in \mathbb{R}^{1\times K}$ a column vector, consisting of the corresponding spectrum weight c_k. $\boldsymbol{f}$ consists of selected basis material fraction images $\boldsymbol{f}_m, (m = 1, ..., M)$.

According to Eq.(9), the projections with an unknown x-ray spectrum can be expressed as a linear combination of several projections with known model effective spectra. With the less spectra parameters involvement, we argue that the spectra parameters can be estimated together with the SpCT reconstruction due to information redundancy of projection data, which could guarantee simultaneously resolving the above two objects. In consistent with our published SEG method [1], considering the measurements follow the Poisson statistical model, the objective function of joint spectrum estimation and image reconstruction for SpCT with the noise removal regularization can be expressed as Eq.(12). It is obtained by performing a second-order Taylor expansion of the Poisson log-likelihood function $g_{w,i}(\hat{Y}_{wi}) = \hat{Y}_{wi} - Y_{wi} ln(\hat{Y}_{wi})$ of the SpCT at the point Y_{wi} .

$$\begin{aligned}&\underset{\boldsymbol{C},\boldsymbol{f}}{\operatorname{argmin}} \sum_{w}\sum_{i} \frac{1}{2Y_{wi}} \left\|\boldsymbol{C}^w \mathcal{P}_{wi}(\boldsymbol{f}) - Y_{wi}\right\|_2^2 + \beta_{TV}\nabla\boldsymbol{f} \\ &+ \lambda \sum_{w} (\left\|\boldsymbol{C}^w\right\|_1 - 1)^2, s.t. \boldsymbol{C}^w \geq 0, \forall w,\end{aligned} \tag{12}$$

where Y_{wi} is the real projection measurement for x-ray length i in energy window w. Here we use total variation (TV) as a regularization term. β_{TV} represents a parameter to balance the strength between fidelity term and the regularization term. To make the solution for Eq.(12) physically meaningful, additional constraints are enforced. That is, the weights in $\boldsymbol{C}^w$ are nonnegative, and their sum is equal to 1. Hence, λ is set to ∞ and $\|.\|_1$ is the L_1-norm of a vector.

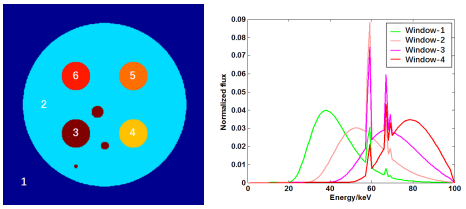
With the block coordinate descent (BCD) method, Eq.(12) can be alternatively optimized if either $\boldsymbol{C}$ or $\boldsymbol{f}$ is fixed. Specifically, Eq.(12) is separated into two subproblems:

$$\min_{\boldsymbol{f}} \sum_{w}\sum_{i} \frac{1}{2Y_{wi}} \left\|\boldsymbol{C}^w \mathcal{P}_{wi}(\boldsymbol{f}) - Y_{wi}\right\|_2^2 + \beta_{TV}\nabla\boldsymbol{f}, \tag{13}$$

$$\begin{aligned}&\min_{\boldsymbol{C}} \sum_{w}\sum_{i} \frac{1}{2Y_{wi}} \left\|\boldsymbol{C}^w \mathcal{P}_{w,i}(\boldsymbol{f}) - Y_{wi}\right\|_2^2 \\ &+ \lambda \sum_{w} (\left\|\boldsymbol{C}^w\right\|_1 - 1)^2, s.t. \boldsymbol{C}^w \geq 0, \forall w.\end{aligned} \tag{14}$$

The optimization details and equations derivation are not presented here due to the limited space, which can be found in [1].

III. Experiments and Results



(a) (b)

Fig. 2. Simulated circle phantom (a) and Effective spectra for each window at 100kVp (b). The material composition of the corresponding numbers are shown in Table I.

TABLE I
Basis material fraction values in Fig.2(a)

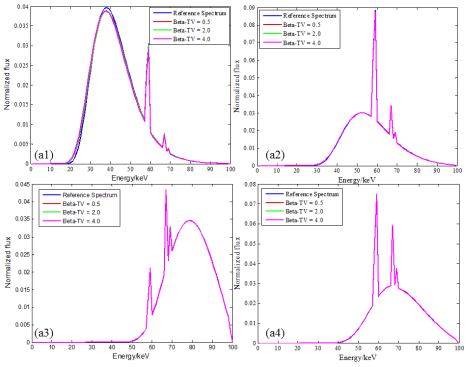
Number	Material	Density(g/cm^3)	f_1	f_2	f_3
1	Air	0.0	0.0	0.0	0.0
2	Water	1.0	1.0	0.0	0.0
3	1.5% (by weight) iodine contrast	0.074	0.0	0.0	0.015
4	3.0% (by weight) iodine contrast	0.148	0.0	0.0	0.03
5	30% (by weight) calcium	0.462	0.0	0.3	0.0
6	50% (by weight) calcium	0.77	0.0	0.5	0.0

TABLE II
The means and STDs of decomposed fraction images with each ROI of different materials in simple phantom

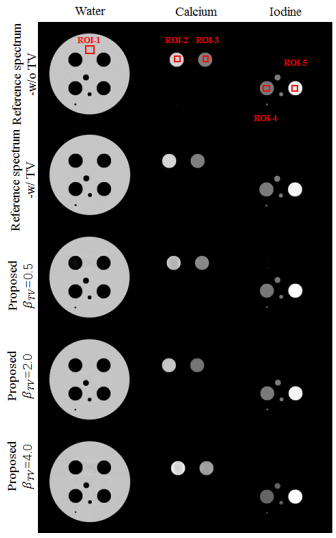
Method	ROI-1	ROI-2	ROI-3	ROI-4	ROI-5
True fractions	1.0000 ± 0.0000	0.5000 ± 0.0000	0.3000 ± 0.0000	0.0150 ± 0.0000	0.0300 ± 0.0000
Reference spectrum w/o TV	0.9978 ± 0.0708	0.5015 ± 0.0220	0.3004 ± 0.0208	$0.0149 \pm 4.9e^{-04}$	$0.0298 \pm 7.1e^{-04}$
Reference spectrum w/ TV	1.005 ± 0.0084	0.4992 ± 0.0012	0.2997 ± 0.0011	$0.0150 \pm 4.9e^{-05}$	$0.0300 \pm 4.9e^{-05}$
The proposed method-$\beta_{TV} = 0.5$	1.012 ± 0.0595	0.4967 ± 0.0038	0.2963 ± 0.0024	$0.0150 \pm 2.1e^{-04}$	$0.0300 \pm 3.9e^{-04}$
The proposed method-$\beta_{TV} = 2.0$	1.005 ± 0.0085	0.5025 ± 0.0016	0.3013 ± 0.0017	$0.0150 \pm 4.2e^{-05}$	$0.0299 \pm 3.9e^{-05}$
The proposed method-$\beta_{TV} = 4.0$	1.004 ± 0.0089	0.4958 ± 0.0041	0.2978 ± 0.0020	$0.0148 \pm 5.7e^{-05}$	$0.0298 \pm 5.4e^{-05}$

In this section, as shown in Fig.2(a), a circle phantom consists of water background with a diameter of 200mm, four circular inserts with a diameter of 35mm and three different size circular inserts is used to evaluate the performance of the proposed algorithm. The phantom consist of different materials and the corresponding fraction values are defined in Table I. The basis materials for circle phantom are selected as water/calcium/iodine. To avoid the contribution of the scatter signal, an equidistant fan-beam geometry is assumed for a SpCT scanner. The distance from x-ray source to the system origin is 132mm and the distance from the source to PCD is 180mm. 640 projections are collected over a full scan range, the detector is composed of 512 detector elements, and each element size is 0.1mm. The reference original x-ray spectrum is generated at 100kVp by SpekCalc software with 2-mm Al filter. And the energy window thresholds are set to 30keV, 50keV, 70keV and 90keV, respectively. The weights for the

initial guess spectrum of each window w are set to the same with $\boldsymbol{C}^w = [0, 0, 0, 0, 0, 1]$. By taking nonideal effects for PCD into consideration, the effective spectra for each window are shown in Fig.2(b). Poisson noise is superimposed onto the measurement by assuming that there are 2×10^5 photons emitted from the x-ray source. All the reconstructed basis material fraction images are 512×512 with each pixel covered an area of $0.075\text{mm} \times 0.075\text{mm}$.



(a)



(b)

Fig. 3. Simulated phantom results. (a) Estimated effective spectra for each window. (b) Reconstructed basis material fraction images. Display windows for water, calcium and iodine are [0.3,1.2], [0,0.6] and [0,0.025], respectively.

In the optimization, we use TV regularization to remove noise. Because the choice of regularization parameter significantly affects the results, we empirically optimize the combination of these parameters to obtain the best reconstructed images and the most accurate spectra. In general, if the selected regularization parameter is too low, image quality suffers from the noise and if the parameter is too high, the reconstructed image is over-smoothed. With a proper selection of the regularization parameter, good reconstructed image quality can be obtained. In our experiment, we choose $\beta_{TV} = 2$ as the optimal regularization parameter by comparing the results of $\beta_{TV} = 0.5, 2, 4$. Fig. 3(a) shows the estimated effective spectra for each window with different regularization parameters at 100kVp. The normalized root mean error (NRMSE) between an estimated spectrum and the reference spectrum for each window with the optimal parameter ($\beta_{TV} = 2$) are 3.70%, 1.11%, 0.48% and 0.32%, respectively. The experimental results indicate that the proposed method can accurately estimate the effective spectrum of each window for PCD from visual and quantitative evaluation.

The reconstructed basis material fraction images are shown in Fig. 3(b). We compare the reconstructed decomposition results of reference spectrum with/without TV regularization and the proposed method with different regularization parameters $\beta_{TV} = 0.5, 2, 4$. In comparison with the first row and the second row of Fig. 3(b), it is indicated that the TV regularization can effectively remove the noise. As shown from the second row and the bottom row of 3(b), with proper selection of regularization parameter ($\beta_{TV} = 2$), the proposed method can provide comparable image quality in comparison with the decomposed basis material fraction images with the reference spectrum. If parameters ($\beta_{TV} = 0.5$) is too low, noisy images are reconstructed and if parameters ($\beta_{TV} = 4.0$) is too high, over-smooth images are obtained. To further quantitatively evaluate the accuracy of decomposed material images, we calculate the mean and standard deviation (STD) of pixels within a uniform region of interest (ROI) in material fraction images as shown in Table II. And the selected ROIs are shown in the red rectangles of Fig. 3. By jointly looking at Fig. 3 and Table II, it is indicated that jointly estimating spectra and basis material fraction images is convenient with no extra process and can provide accurate estimated spectra and decomposed material fraction images. And the spectrum estimation performance of the proposed method has little relation with the selected regularization parameters, which only have influence on the decomposed material images quality.

IV. Discussion and Conclusion

In this work, we proposed a joint spectrum estimation and image reconstruction method for spectral CT. The proposed method can accurately estimate the effective spectrum for each energy window and reconstruct the decomposed basis material images including K-edge material. Due to the strong correlation between each window, in the future, we will introduce the correlation as a prior information into our objective function to suppress the noise. In addition, more practical data based on PCD are needed to verify the proposed method's performance.

References

[1] S. Chang, M. Li, H. Yu, X. Chen, S. Deng, P. Zhang, and X. Mou, "Spectrum estimation-guided iterative reconstruction algorithm for dual energy ct," *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 246–58, 2020.

[2] T. Schmidt, R. Barber, and E. Sidky, "A spectral ct method to directly estimate basis material maps from experimental photon-counting data," *IEEE Transactions on Medical Imaging*, vol. PP, no. 99, pp. 1–1, 2017.

[3] W. Wu, Q. Wang, F. Liu, Y. Zhu, and H. Yu, "Block matching frame based material reconstruction for spectral ct," *CoRR*, vol. abs/1810.10346, 2018.

[4] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. B?umer, C. Herrmann, R. Steadman, and G. Zeitler, "Experimental feasibility of multi-energy photon-counting k-edge imaging in pre-clinical computed tomography," *Physics in Medicine Biology*, vol. 53, no. 15, pp. 4031–4047.

Ringing Artefact Removal From Sparse View Tomosynthesis using Deep Neural Networks

Shabab Bazrafkan, Vincent Van Nieuwenhove, Joris Soons, Jan De Beenhouwer, and Jan Sijbers,

***Abstract*—Tomosynthesis imaging is a special type of X-ray-based image acquisition wherein several projections are acquired from a limited angular range around the object or in a line above it. Subsequently, from these projection images, cross-section images parallel to the detector are reconstructed. However, if only a small number of projection images is acquired on a limited angular range, the reconstruction images suffer from ringing artefacts. This issue has not been addressed in the literature due to the lack of any model-based approach explaining this phenomenon. In the clinical setting, these artefacts can hinder a correct diagnosis. In this work, a deep learning-based ringing artifact reduction algorithm is proposed. The deep learning network was trained on 45786 medical images, resulting in a substantial reduction of ringing artefacts in the tomosynthesis reconstructions. Based on the numerical and visual evaluations, a conclusion is made on the positive effect of a deblurring Deep Neural Network in getting higher quality outputs.**

***Index Terms*—Deep Neural Networks, Deep Learning, Tomosynthesis, Ring Artefact.**

I. Introduction

Tomosynthesis involves the reconstruction of a 3D object imaged with a relatively small number of views acquired over a limited angular range, resulting in typical image artifacts. Indeed, reconstructions with (filtered) backprojection ((F)BP) algorithms generate out-of-plane artifacts that appear as blurred versions of the in-plane objects. Under certain conditions, this blurring changes into ringing, in particular when the perpendicular distance from a ringing source to the imaging plane increases above a certain threshold. Indeed, ringing occurs by a similar mechanism as blurring but is caused by the small number of projections in a sweep. It is caused by a high-contrast structure (e.g. rib structure) that is outside the plane in focus and whose contribution to that imaging plane is not sufficiently blurred. As a result, each interval between two successive projections in the imaging plane is wider than the tomographic blurring, so that separate instances of the ringing source appear in the imaging plane.
In order to avoid the FPB reconstruction disadvantages, other numerical methods known as iterative reconstruction algorithms such as SIRT (Simultaneous Iterative Reconstruction Technique) and CGLS with total variation minimization are used to generate the tomosynthesis image. However the main problem of ringing artefacts persists even with these reconstruction methods. Fig 1 illustrates this issue wherein using a low number of projections induces a disturbing ringing artefact in the final reconstruction which is not describable by analytical models. Hence, there are no successful model-based algorithms that are able to suppress these type of artefacts. In the current study, a data-driven machine learning method known as Deep Learning is utilized to restore the tomosynthesis image. To the best of our knowledge, there are no other data-driven or model-based methods addressing this issue.
Deep learning approaches have been successfully applied to the medical image processing. One of the widely investigated applications of the DNNs in medical image processing is known as image restoration. In image restoration, the input and the output of the network are from the same class of data. They include the same content and represent the same object. Noise removal is an example of image restoration. Other work focussed on the removal of reconstruction artefacts in limited angle or few-view CT with Deep learning techniques as in [6], [7]. In the current study, a similar approach is applied to the tomosynthesis reconstructions in order to remove the ringing artefact caused by a limited number of projections.

In the next section, the Neural Network model used in the current study is explained followed by a discussion on the database and training procedure in section III. In section IV, the numerical and visual results are presented and conclusions and future works is discussed in section V.

II. Deep Neural Networks

A. Network Design

Convolutional Deep Neural Networks consist of consecutive processing units also known as layers which can be fully connected, convolution, deconvolution, pooling, or un-pooling operations [1]. Based on the application, the network designs could take advantage of different regularization units such as batch normalization [2], weight normalization and/or drop-out [3] techniques. One of the widely used network designs in image processing is known as Fully Convolutional Deep Neural Networks (FCDNN). These are DNNs wherein no Fully Connected operation is applied throughout the network design. All the layers are convolution, deconvolution, pooling, and un-pooling units. FCDNNs are used in applications in which the input and the output of the network are images. Restoration problems are good examples of the uses cases of FCDNNs. In our work, two FCDNNs known as Mixed Scale Dense (MSD) Deep Neural Network [4] and a deblurring DNN [5] are utilized to remove ringing artefacts in tomosynthesis reconstruction. These network architectures are explained in the following sections.

S. Bazrafkan, J. De Beenhouwer and J. Sijbers are with imec-Visionlab, Department of Physics, University of Antwerp, Antwerp, Belgium e-mail: {shabab.bazrafkan},{jan.debeenhouwer},{jan.sijbers}@uantwerpen.be.

V. Van Nieuwenhove and J. Soons are with Agfa NV, Mortsel, Belgium email: {vincent.vannieuwenhove},{joris.soons}@agfa.com.

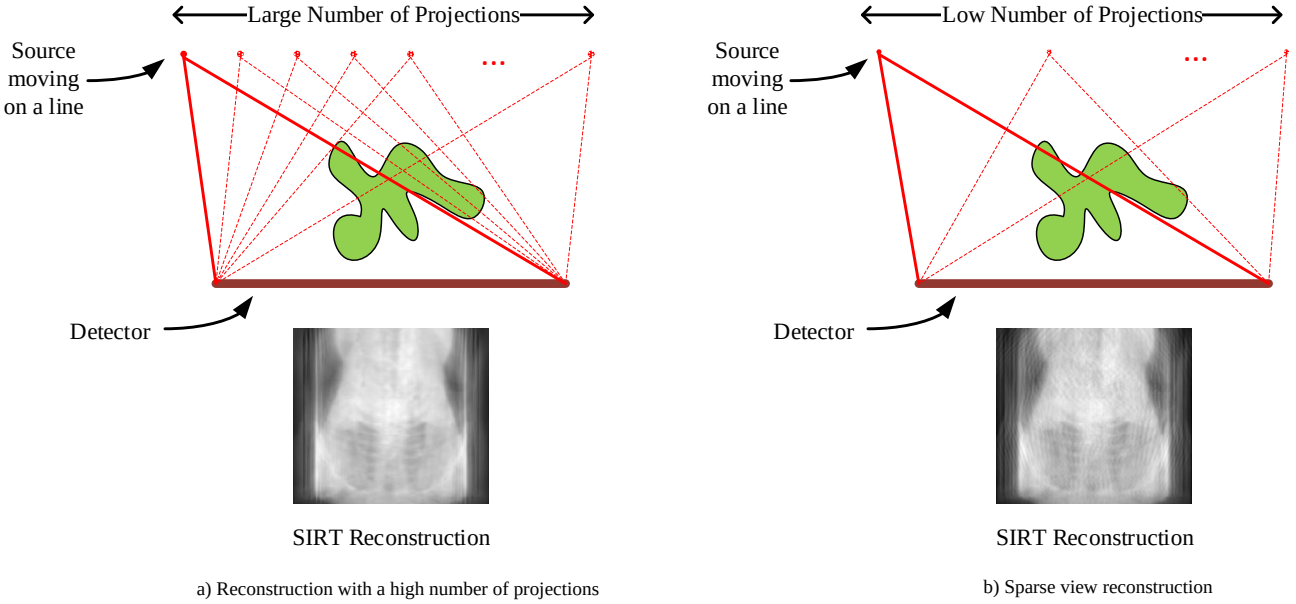


Fig. 1. a) A high number of projections returns a high quality reconstruction. b) Sparse view acquisition induces ring artefact in the final reconstruction.

1) Mixed Scale Dense (MSD) DNN: In order to explain the MSD network, one needs to understand the dilation property of the convolution operation. Dilation allows the convolution kernel to cover a larger area without using a higher number of parameters in larger kernels and/or applying a pooling operation. This is computationally more efficient because using bigger non-dilated kernels leads to a higher number of parameters which is susceptible to over-fitting, and using the pooling operation causes blurring in the final outcome. The dilation idea was first introduced in [4] and is illustrated in Fig. 4a. The operation is simply implemented by dilating the convolution kernel and filling the gaps with zero value.

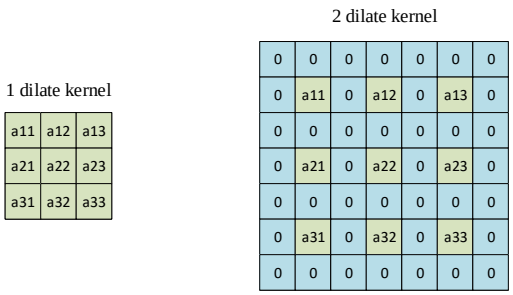


Fig. 2. 3×3 kernels. Left: 1 dilate. Right: 2 dilate.

The Mixed Scale Dense DNN as introduced in [4], is taking advantage of the dilation operation followed by concatenating the output of each layer with its input in the channel dimension. The network architecture is shown in Fig. 4a. The kernel dilation value is chosen by the layer number. This network has been already used to remove low dose CT artefacts from reconstructed images in [6], [7]. In the current study, this architecture is utilized to remove the ringing artefacts from low-dose tomosynthesis reconstructions.

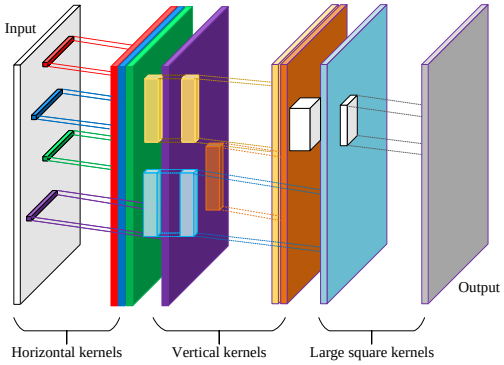
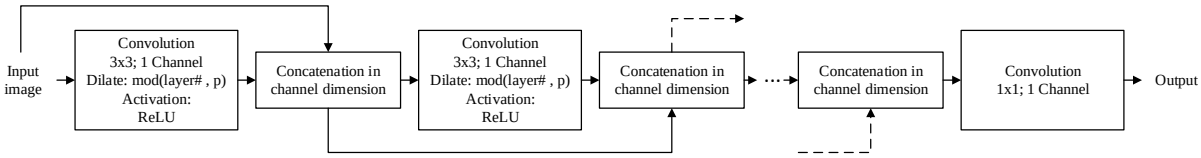
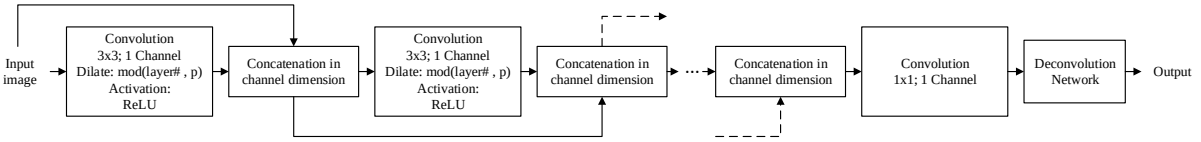


Fig. 3. Deconvolution network presented in [5].

2) Deblurring Convolutional Neural Network: Convolutional Deep Neural Networks tend to introduce a certain amount of blurring into the output especially when used with the Mean Squared Error as the loss function. However, Deblurring or Deconvolutional Convolutional Neural Networks have been proposed that are designed to reduce the blurring artefact in their input image. These networks are widely used to increase the quality of the image in super-resolution applications. In this article, a Deblurring network is used after the MSD network to improve the sharpness of the output. In [5], the authors proposed a network architecture designed for image deblurring, which is shown in Fig. 3. The first two layers consist of horizontal and vertical kernels and the last layer is a large square kernel. This design resembles the Singular Value Decomposition technique used in conventional deblurring methods, with the difference that here these filters are learned during training.



(a) MSD



(b) MSD+Deconv

Fig. 4. a) Mixed Scale Dense network. b) Mixed Scale Dense network followed by deblurring network.

B. Network Architectures

In the current study, two main architectures have been evaluated, and compared for the ring removal application.

1) **MSD**: The network is shown in Fig 4a. This is an MSD network with 25 layers and dilate rage p=5.
2) **MSD+Deconv**: The network is shown in Fig. 4b. This is an MSD network with 25 layers and dilate rage p=5 followed by a deblurring network shown in Fig. 3. The deblurring network uses the horizontal kernel size (1,129), vertical kernel size (21,1), and square kernel size (17,17). The number of channels in the first and second layers is 128 and 32, respectively.

III. Database and Training

In order to simulate the ringing artefacts and the corresponding ground truth, two different scanning geometries were defined, similar to Fig. 1. The first geometry takes 10 projections which eventually leads to reconstructions with the ringing artefacts and the second geometry uses 100 projections and the reconstruction is used as the ground truth. All simulations are accomplished using the ASTRA toolbox [8]. The reconstruction volumes were calculated with 25 iterations of SIRT. Two separate databases are used for training and testing purposes. Using different datasets is essential in evaluating the generality of the solution [9]. The training set is divided into Training and Evaluation subsets with the ratio 4 to 1, and the network is blind to the test set during the training process. Note that the networks are trained on the 2D samples in the transverse plane and results are shown in coronal plane where the ring artefact is visible.

A. Training Database

CPTAC-PDA: National Cancer Institutes Clinical Proteomic Tumor Analysis Consortium Pancreatic Ductal Adenocarcinoma (CPTAC-PDA)[1] is a publicly available database containing 45786 Pancreas images from CPTAC phase 3 patients. It consists of 45 radiology and 77 pathology subjects. This database contains several modalities including CT, Computed Radiography (CR) and MRI samples. The images are from different sizes but in the current work, they were resized to 30×512 (transverse plane).

B. Test Database

Visible Human Project CT Datasets: Visible Human Project CT Datasets[2] contains 2989 images from 10 CT imaging cases. This dataset is publicly available. The images are 512×512 while in the current study they were all resized to 30×512 (transverse plane).

C. Training

The Mean Squared Error between the output of the network and the ground truth has been used as the loss function at the training stage. An ADAM optimizer has been utilized to update the parameters with learning rate, β_1, β_2 and ϵ equal to 0.0001, 0.9, 0.999, and 10^{-8} respectively. The MXNET 1.3.0[3] [10] framework has been used to train the network on top of python 2.7. The training and validation losses are shown in Fig 5. As it is shown in this figure, the deblurring network introduces a large gap between the loss functions which means that the MSD+Deconv network converges to a better solution at the training stage. In the next section, both networks are evaluated on the test database.

IV. Results

Table I shows the Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity index (SSIM) for both networks tested on 10 subsets of the test database. The column labeled "Input" corresponds to the comparisons between input and the ground truth. The MSD network returns outputs with up to 5dB improvement compared to the input signal and, the numerical results show 2 to 3 dB improvement for the MSD+Deconv network compared to MSD. This amount of improvement in the ranges higher than 40 dB implies that the deblurring network has a significant effect on the output quality.
Fig. 6 illustrates visual results from the Shoulder M subset and clearly shows that the output of the MSD+Deconv network is sharper and more consistent with the ground truth image.

[1] https://wiki.cancerimagingarchive.net/display/Public/CPTAC-PD

[2] https://mri.radiology.uiowa.edu/visible_human_datasets.html

[3] https://mxnet.apache.org/

	Input				MSD				MSD+Deconv			
	PSNR		SSIM		PSNR		SSIM		PSNR		SSIM	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Hip M	38.88	0.88	0.944	7.1e-5	44.05	0.97	0.980	2.5e-3	**47.08**	**0.95**	**0.989**	**1.5e-3**
Pelvis M	37.73	0.97	0.930	9.5e-3	41.71	1.13	0.969	6.2e-3	**44.25**	**1.0**	**0.982**	**2.5e-3**
Shoulder M	36.60	1.30	0.921	1.0e-2	40.25	**1.43**	0.960	7.1e-3	**43.65**	1.88	**0.980**	**7.1e-3**
Head M	43.98	1.60	0.969	7.0e-3	48.18	1.36	0.980	3.2e-3	**50.78**	**1.17**	**0.991**	**8.8e-4**
Hip F	40.97	2.57	0.957	1.4e-2	45.51	2.16	0.981	5.7e-3	**48.42**	**1.82**	**0.989**	**2.9e-3**
Pelvis F	38.31	0.20	0.935	4.4e-3	42.28	0.59	0.969	2.8e-3	**44.99**	**0.51**	**0.983**	**1.3e-3**
Shoulder F	36.98	1.42	0.929	9.9e-3	40.89	**1.67**	0.964	7.3e-3	**43.70**	2.16	**0.979**	**6.5e-3**
Head F	41.80	1.90	0.960	9.4e-3	46.44	1.65	0.979	3.9e-3	**49.41**	**1.64**	**0.991**	**1.4e-3**
Ankle F	42.60	2.30	0.964	8.1e-3	46.63	1.70	0.980	4.6e-3	**49.11**	**1.66**	**0.989**	**1.7e-3**
Knee F	42.38	0.74	0.967	4.3e-3	46.94	0.50	0.986	1.5e-3	**49.42**	**0.42**	**0.990**	**1.1e-3**

TABLE I
TEST RESULTS FOR BOTH MSD AND MSD+DECONV MODELS ON SUBSETS OF VISIBLE HUMAN PROJECT CT DATASETS

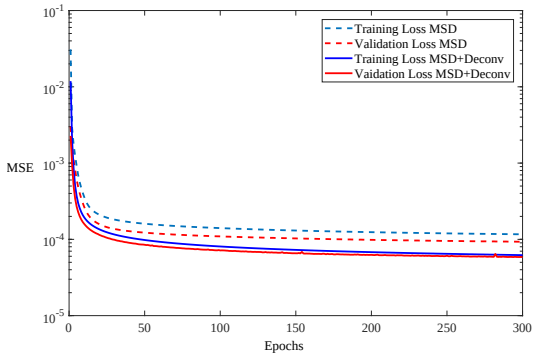


Fig. 5. Training and validation loss for both MSD and MSD+Deconv models.

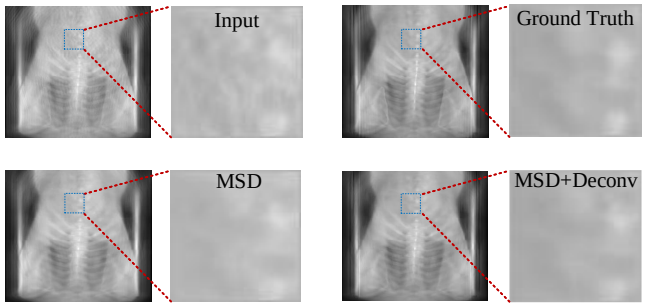


Fig. 6. Top left: The input to each network. Top right: Ground truth. Bottom left : The output of MSD network. Bottom right: the output of MSD+Deconv network. The visual results shows sharper and higher quality results for MSD+Deconv network.

V. CONCLUSION AND FUTURE WORK

In this article, a deep learning approach has been introduced to remove the ring artefacts from sparse view tomosynthesis reconstruction images. To the best of our knowledge, no other method presented in the literature addresses this problem and the current study shows the effectiveness of Deblurring networks in obtaining sharper and higher quality results using both numerical and visual assessments. Fully convolutional DNNs applied to restoration problems do not induce severe localization errors due to the convolution nature. In our initial visual evaluations (an example is depicted in Fig. 6), the network does not induce localization inconsistencies but our future works include educated evaluations by radiologists to ensure that the network does not alter the diagnosis procedure.

ACKNOWLEDGMENT

This work is financially supported by VLAIO (Flemish Agency for Innovation and Entrepreneurship), through the ANNTOM project HBC.2017.0595 and the Research Foundation Flanders (FWO) through the MetroFlex project S004217N.

REFERENCES

[1] J. Lemley, S. Bazrafkan, and P. Corcoran, "Deep learning for consumer devices and services: Pushing the limits for machine learning, artificial intelligence, and computer vision." *IEEE Consumer Electronics Magazine*, vol. 6, no. 2, pp. 48–56, 2017.

[2] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[3] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[4] D. M. Pelt and J. A. Sethian, "A mixed-scale dense convolutional neural network for image analysis," *Proceedings of the National Academy of Sciences*, vol. 115, no. 2, pp. 254–259, 2018.

[5] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Advances in neural information processing systems*, 2014, pp. 1790–1798.

[6] S. Bazrafkan, V. Van Nieuwenhove, J. Soons, J. De Beenhouwer, and J. Sijbers, "Deep neural network assisted iterative reconstruction method for low dose ct," *arXiv preprint arXiv:1906.00650*, 2019.

[7] D. Pelt, K. Batenburg, and J. Sethian, "Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks," *Journal of Imaging*, vol. 4, no. 11, p. 128, 2018.

[8] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the astra toolbox," *Optics express*, vol. 24, no. 22, pp. 25 129–25 147, 2016.

[9] S. Bazrafkan, T. Nedelcu, P. Filipczuk, and P. Corcoran, "Deep learning for facial expression recognition: A step closer to a smartphone that knows your moods," in *Consumer Electronics (ICCE), 2017 IEEE International Conference on*. IEEE, 2017, pp. 217–220.

[10] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang, "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems," *arXiv preprint arXiv:1512.01274*, 2015.

A CT image feature space (CTIS) loss for reconstruction with deep learning methods

Ao Zheng, Zhiqiang Chen, Kaichao Liang, Li Zhang and Yuxiang Xing*

***Abstract*—Deep learning-based methods have attracted a lot of attention in medical imaging field because of its great power for image processing. Many researches focus on using deep learning methods to solve ill-conditioned CT reconstruction problems, such as low-dose and sparse view. For supervised learning methods in CT reconstruction, different loss functions will lead to different image quality which may affect diagnosis. In this paper, we proposed a loss in CT image feature space (CTIS) that is defined in a feature space learned from a big dataset of high-quality CT images by an autoencoder. The CTIS can be combined with per-pixel loss to form a total loss function in our experiment. In the study, the output of the bottom dense layer in the autoencoder is treated as a high dimensional vector in CTIS. In this way, our loss function is not only global, but also contains prior knowledges from CT image dataset. We demonstrated the special structures that learned from CT images by an autoencoder in an intuitive way. Then, we evaluated its performance in a low-dose CT reconstruction task. We compared the reconstructions using CTIS loss + L2-loss with other conventional losses, our method gets encouraging preliminary results on quantitative comparison and visual examination.**

***Index Terms*—prior, feature space, loss function, ill-conditioned CT reconstruction, autoencoder.**

I. Introduction

Computed tomography (CT) has been widely used in clinical diagnosis. With the concern of radiation dose, people tried to lower dose in various ways, such as reduce the tube current or scanning angles [1]. However, in these ill-conditioned situations, the image quality of directly analytical reconstruction cannot meet the requirement of clinical diagnosis. Conventional ways to deal with this situation are composed of sinogram filtration, iterative reconstruction and image post-processing [2].

Recently, many deep learning methods have been proposed to solve ill-conditioned reconstruction problems [3], [4], [5]. For supervised learning, we can minimize the loss function defined by the distance between the result and ground truth. In general, loss functions can be divided into two categories, low-level loss functions and high-level loss functions. The low-level loss functions are simple loss functions which contain per-pixel loss and some regularizations, such as mean absolute error (MAE), mean squared error (MSE) and total variation (TV) regularization. High-level loss functions are based on high-level feature space and it includes perceptual loss and adversarial loss. Compared to low-level loss functions, it focuses more on semantic and structural information. How to choose the metric of distance is an important issue. MSE optimizes per-pixel L2-distance and often leads to over-smoothed images. Generative adversarial network (GAN) [6] can avoid over-smooth and give visually pleasant images. However, it optimizes the distribution of images and often hard to train. Perceptual loss [7] is another way to avoid over-smooth, which measures the distance in the feature space of VGG [8] pretrained on ImageNet. But it does not constraint the output to match each pixel exactly and might be not optimal for CT images.

In this paper, inspired by perceptual loss, we proposed a feature space loss by taking advantage of the information from a large set of high-quality CT images. We use the distance in high-level feature space tailored for CT images to form a CT image feature space (CTIS) loss. For CT reconstruction, we want our result not only have a low perceptual distance, but also match the ground truth for each pixel. Therefore, we combine our CTIS loss with per-pixel loss. The feature space is defined by an autoencoder [9]. In order to show the effectiveness of the autoencoder, we tried to testify that the output of the latent space (bottom dense layer) in autoencoder represent the structures of CT images. This is the key point of our method. In experiment, we compared our CTIS loss with conventional loss functions for fan-beam low-dose reconstruction which used the same framework as previous work [10], [11] of our group. From quantitative comparison and visual examination, we showed our CTIS loss could perform better than most conventional loss functions. At the end of our work, we give our discussion on the drawbacks and future work.

II. Method

A. The feature space specifically tailored for CT images

Our idea is to find a feature space representative to the general knowledge of CT images. For solving this problem, autoencoder is a natural option, which optimize:

Asterisk indicates corresponding author.

This work was supported in part by the National Natural Science Foundation of China (No. 61771279 and No. 61527807) and National Science Fund for Distinguished Young Scholars of China (No. 11525521).

Ao Zheng, Zhiqiang Chen, Kaichao Liang, Li Zhang and Yuxiang Xing are with Department of engineering physics, Tsinghua University, Beijing, 100084, China. (e-mail: xingyx@mail.tsinghua.edu.cn).

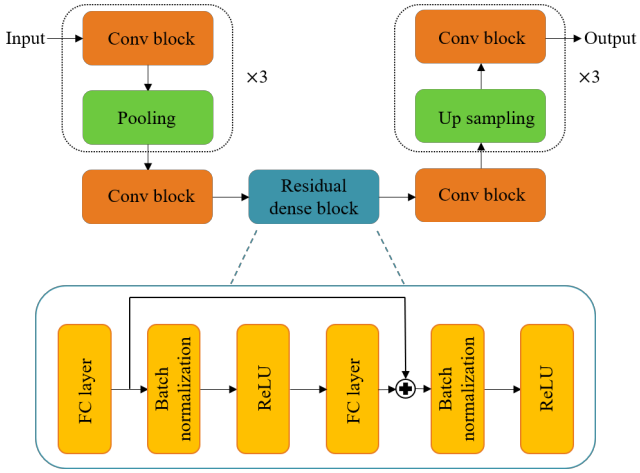


Fig. 1. The network architecture of autoencoder. The Conv block contains two cascaded convolutional layers, batch normalization layers and activation layers. Residual dense block is composed of two cascaded fully-connected layers, batch normalization layers and activation layers with a shortcut.

$$E, D = \arg\min \sum_k \|\boldsymbol{\mu}_k - D(E(\boldsymbol{\mu}_k))\|_2 \tag{1}$$

where E and D represent encoder and decoder respectively, $\boldsymbol{\mu}_k$ is one sample of normal dose CT images and $\|\cdot\|_2$ means the L2-norm. Encoder learns to extract the features of input and map them to a latent space, while decoder use the output of latent space to reconstruct an input. The architecture of autoencoder we used is shown in Fig. 1. For the encoder part, we use alternate convolutional block and pooling layer to expand receptive field. In order to map the features to a latent space, a residual dense layer is introduced in the bottleneck of the autoencoder. In this way, the output of the latent space is globally related to the input, which can also make the autoencoder learn more complicated features. However, considering the huge parameters of dense layers, we use small patches for input, for example, 128×128 patches. The decoder part is symmetrical to the encoder. Once the output of autoencoder can perfectly match the input, the network has learned the inherent features of CT images, then we can use it to compute our CTIS loss.

B. *Visualization of latent space*

For a given CT image $\boldsymbol{\mu}_k$, we denote the output of the encoder to be an n-dimensional vector:

$$[f_{k,1}, f_{k,2}, \ldots, f_{k,n}] = E(\boldsymbol{\mu}_k) \tag{2}$$

Trying to explain what $f_{k,j}, j \in \{1, 2, \ldots, n\}$ represent in an intuitive way, we add $\boldsymbol{\delta}_j$ to $E(\boldsymbol{\mu}_k)$ and then compare the output of decoder. It can be denoted as (3). By changing j from 1 to n and then compute the result, we can know what the latent space represents for a given CT image.

$$D(E(\boldsymbol{\mu}_k)+\boldsymbol{\delta}_j) - D(E(\boldsymbol{\mu}_k)), \boldsymbol{\delta}_j = \begin{cases} 1 & \text{the } j\text{-th dimension} \\ 0 & \text{others} \end{cases} \tag{3}$$

C. *Problem modeling and CTIS loss evaluation*

For supervised learning in CT reconstruction, different loss functions will lead to different image quality. Suppose the result is $\hat{\boldsymbol{\mu}}_k$ and ground truth is $\boldsymbol{\mu}_k^*$, where k indexes images in training dataset.

$$\text{MSE}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}^*) = \frac{1}{K}\sum_k \|\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k^*\|_2^2 \tag{4}$$

As we described in the introduction, MSE focuses on per-pixel distance, and will give an over-smoothed result. To solve this problem, some people use adversarial loss in GAN framework. While, conventional GAN suffers from several problems such as convergence problems, mode collapse and poor stability. Another popular way to avoid over-smooth is to use perceptual loss which is more stable than GAN. It compares results and ground truth in high level feature space of VGG pretrained on ImageNet. But there is a gap between RGB images and medical images. For our CTIS loss, we compare results and ground truth in the latent space of autoencoder pretrained on high-quality CT images. And it can be written as:

$$\text{CTIS Loss}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}^*) = \frac{1}{K}\sum_k \|E(\hat{\boldsymbol{\mu}}_k) - E(\boldsymbol{\mu}_k^*)\|_2^2 \tag{5}$$

To ensure the accuracy in pixel level, we combine CTIS loss and MSE with equal weight as our final loss function. In this way, we can constrain the result not only in pixel level, but also in latent space of autoencoder which was pretrained on CT images rather than ImageNet.

For a CT reconstruction problem, we have:

$$\mathbf{p} = \mathbf{H}\boldsymbol{\mu} + \mathbf{N} \tag{6}$$

where $\mathbf{p}$ represents measured projection data, $\mathbf{H}$ is system projection operator, $\boldsymbol{\mu}$ is a linear attenuation map and $\mathbf{N}$ is noise term. Using the dual-domain reconstruction network proposed by our group (please refer to [10] and [11] for more details), the reconstruction problem is modeled as follow:

$$\begin{aligned} \boldsymbol{\theta}_{\text{prj}}, \boldsymbol{\theta}_{\text{img}} = \arg\min \{&\text{CTIS Loss}(f(\mathbf{p}, \boldsymbol{\theta}_{\text{prj}}, \boldsymbol{\theta}_{\text{img}}), \boldsymbol{\mu}^*) \\ &+\text{MSE}(f(\mathbf{p}, \boldsymbol{\theta}_{\text{prj}}, \boldsymbol{\theta}_{\text{img}}), \boldsymbol{\mu}^*)\} \end{aligned} \tag{7}$$

In (7), f denotes the reconstruction network operator, $\boldsymbol{\theta}_{\text{prj}}$ and $\boldsymbol{\theta}_{\text{img}}$ are the network parameters of projection domain CNN and image domain CNN respectively. We conduct a fan-beam low-dose reconstruction experiment. The dual-domain reconstruction network is composed of projection domain CNN, analytical reconstruction layer and image domain CNN. Projection domain CNN and image domain CNN extract feature maps of their inputs and denoise in their own domains respectively. Analytical reconstruction layer is to convert data from projection domain to image domain. It contains three sub-steps, weighting, filtering and weighted backprojection. We can use FBP operator to initialize it and fix them. In this way, the loss can back-propagate from image domain to projection domain and we can get a better result than image post-processing methods with same number of parameters.

III. Experimental Study

The training and test datasets of autoencoder was downloaded from the website "The Cancer Imaging Archive" (TCIA), which contains head, chest and abdomen. We use 10K slices in total and 90% as training dataset. We use ADAM [12] as optimizer with learning rate setting to 1e-4. The autoencoder is trained in a self-supervised way by optimizing the MSE between input and output. Training converges after 50 epochs.

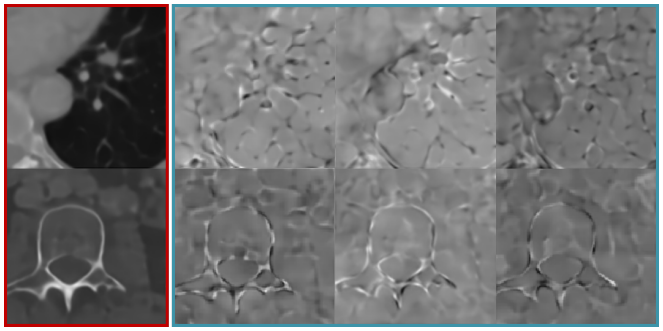


Fig. 2. Visualization of latent space (second fully-connected layer in the residual dense block). Each row represents one sample. The first column is the input of autoencoder. And for the right three columns, each column represents the difference of output after adding different $\boldsymbol{\delta}_j$ to $E(\boldsymbol{\mu}_k)$. Images in same rectangle are displayed in the same self-adaptive display windows.

A. Visualization of latent space

We use the method described in Section II to visualize the latent space of autoencoder. Some samples which contain complex structures are shown in Fig. 2. From the results we can see that each parameter in latent space represent one kind of structure of inputs. It is a little similar to principal component analysis (PCA), but in a non-linear way. Since the autoencoder has learned to extract high-level feature of CT images as prior knowledge, in other supervised tasks, we can compare results and ground truth in the latent space. In this way, we constrain our results more similar to ground truth from structural level rather than pixel level, which will effectively avoid over-smooth.

B. Evaluation in reconstruction

We evaluate our CTIS loss in a fan-beam low-dose reconstruction experiment. Lung CT dataset from Give A Scan® are used for generating training and test datasets. The size of image and projection are 512×512 and 512×720 respectively. 5000 incident photons per ray are simulated for low dose situation. We use the dual-domain network architecture described in Section II. To examine the effectiveness of our CTIS loss, we compare our results with those of MSE and perceptual loss. For perceptual loss, we choose the output of "block3_conv3" layer in VGG-16 to compute loss function and also add MSE with equal weight of perceptual loss. Two cases are shown in Fig. 3.

From Fig. 3, we can see that the reconstructed image with MSE loss is over-smooth as expected. The result of perceptual loss solved this problem, but some details are not completely satisfying. We can check local structures in the zoom-in results. Seeing the results in blue rectangle, our method can get a more accurate result and recover more texture on tissues than perceptual loss. Compared in the red rectangle, we can find the result of perceptual loss is also slightly over smoothed, but our CTIS loss can recover more detailed structures. We think this is caused by the difference between RGB images and CT images. CT images contain many complex and tiny structures. In addition, it also has a lot of textures on tissues. But RGB images do not have these characters. We also use peak signal to noise ratio (PSNR) and structural similarity index (SSIM) to compare

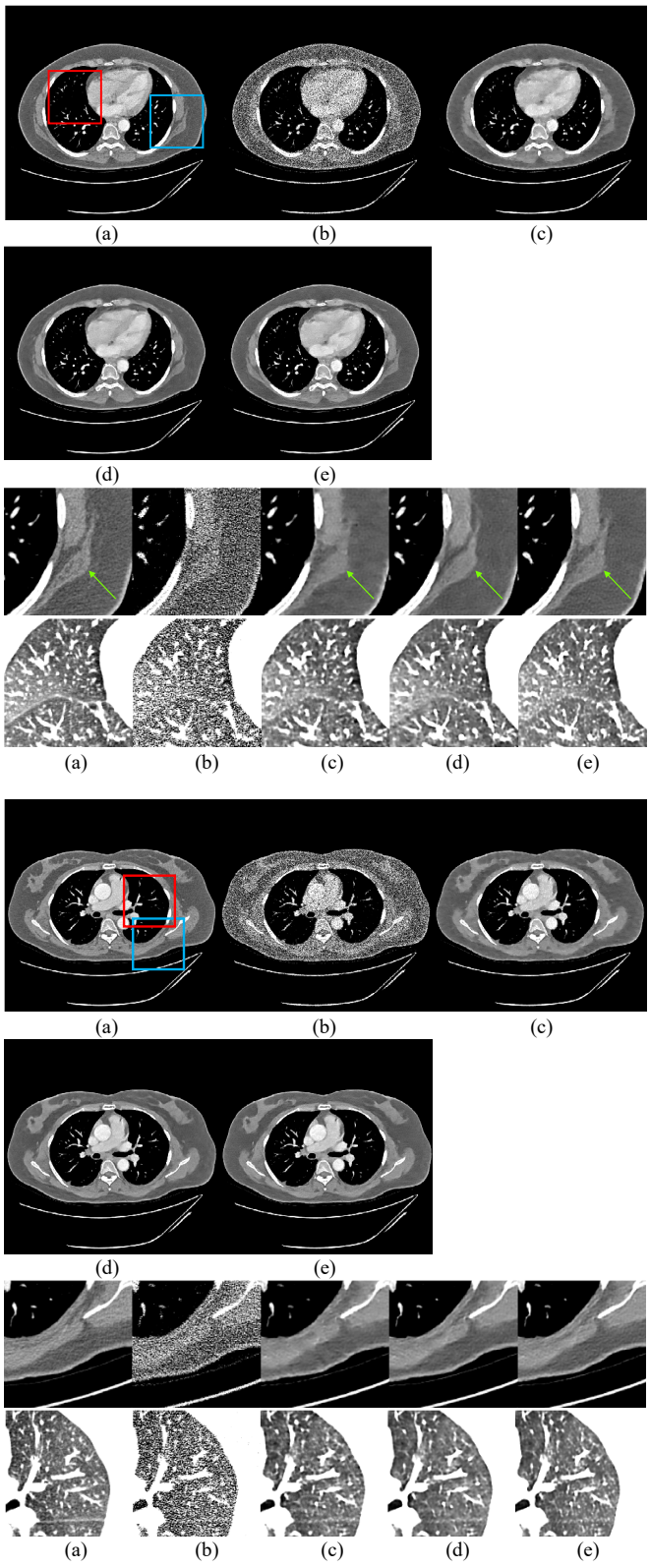


Fig. 3. Two cases in test dataset for evaluation. The first two rows of each case are reconstructed images. The below shows the zoom-in structures marked in the rectangles of the first image. Different letters represent: (a) Normal dose FBP; (b) Low dose FBP; (c) MSE; (d) Perceptual loss + MSE; (e) CTIS loss + MSE. The display window of first three rows for each case is [-300, 300] HU and the last row is [-1000, -500] HU.

TABLE I
QUANTITATIVE COMPARISON

Method	PSNR(dB)	SSIM
Low dose FBP	24.32	71.49%
MSE	34.84	95.90%
Perceptual loss + MSE	35.22	96.34%
CTIS loss + MSE	**35.32**	**96.45%**

test results quantitatively, which can be shown in Table 1.

$$\mathrm{PSNR}=10\times\log_{10}\left(\frac{\max(\boldsymbol{\mu}^*)^2}{\mathrm{MSE}}\right) \tag{8}$$

$$\mathrm{SSIM}=\frac{(2\bar{\hat{\boldsymbol{\mu}}}\bar{\boldsymbol{\mu}}^*+C_1)(2\sigma_{\hat{\boldsymbol{\mu}}\boldsymbol{\mu}^*}+C_2)}{(\bar{\hat{\boldsymbol{\mu}}}^2+\bar{\boldsymbol{\mu}}^{*2}+C_1)(\sigma_{\hat{\boldsymbol{\mu}}}^2+\sigma_{\boldsymbol{\mu}^*}^2+C_2)} \tag{9}$$

where $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}^*$ are the output of the network and the ground truth. $\bar{\hat{\boldsymbol{\mu}}}$ and $\bar{\boldsymbol{\mu}}^*$ are means of $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}^*$ respectively, $\sigma_{\hat{\boldsymbol{\mu}}}$ and $\sigma_{\boldsymbol{\mu}^*}$ are standard deviations of $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}^*$, and $\sigma_{\hat{\boldsymbol{\mu}}\boldsymbol{\mu}^*}$ is covariance. $C_1=(k_1L)^2$ and $C_2=(k_2L)^2$ with L denoting the dynamic range of pixel intensity. $k_1=0.01$ and $k_2=0.03$.

The PSNR and SSIM results prove the advantage of our method. This shows incorporating CT images prior knowledge as loss function can make the method perform better and more stable in both detailed structures and visual effect. The MSE loss get a poor result which shows over-smoothed area in reconstructed images. And we should also notice that perceptual loss performs nearly close to our CTIS loss. Therefore, we randomly choose 25 samples from test dataset and plot the SSIM of them to furtherly compare perceptual loss and CTIS loss. From Fig. 4, the results of CTIS loss perform well for most samples, and we admit the difference is not very big. This is due to the huge dataset ImageNet and the optimized pretrain weight of VGG. Considering the size of dataset used for training our autoencoder and the optimization of network parameters, our CTIS loss could have space for further improvement. Because the result of adversarial loss (GAN) depend heavily on training details. Therefore, we didn't compare with it in our experiment.

IV. Discussion and Conclusion

In this paper, we propose a CTIS loss based on autoencoder trained from normal dose CT images. It is easy to use it for supervised learning in CT reconstruction. The autoencoder learns the high-level feature of high-quality CT images as prior knowledge in the latent space. We can combine our CTIS loss with per-pixel loss to achieve good performance not only in pixel level, but also in high-level feature space. Compared to conventional loss functions, it can get a better result than MSE and perceptual loss, but still needs further improvement. However, there are also some drawbacks. The result depends on autoencoder pretrained on CT images, which means it would only work when there is plenty of data in target domain. In future, we will keep optimizing our autoencoder and do a comprehensive comparison for this CTIS loss. In addition, we hope to build a specialized feature extractor trained on "CT ImageNet".

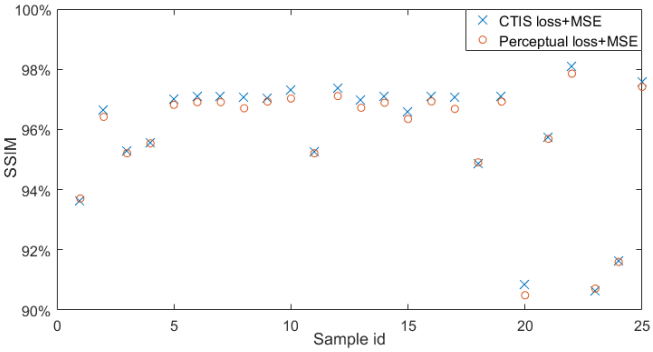


Fig. 4. SSIM between results and ground truth of different loss functions.

References

[1] M. M. Lell and M. Kachelrieß, "Recent and Upcoming Technological Developments in Computed Tomography: High Speed, Low Dose, Deep Learning, Multienergy," *Investigative Radiology*, vol. 55, no. 1, pp. 8-19, 2020.

[2] H. Shan et al., "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Machine Intelligence*, vol. 1, no. 6, p. 269, 2019.

[3] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509-4522, 2017.

[4] H. Shan et al., "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1522-1534, 2018.

[5] D. Wu, K. Kim, G. El Fakhri, and Q. Li, "Iterative low-dose CT reconstruction with priors trained by artificial neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2479-2486, 2017.

[6] I. Goodfellow et al., "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672-2680.

[7] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*, 2016, pp. 694-711: Springer.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[9] A. Ng, "Sparse autoencoder," *CS294A Lecture notes*, vol. 72, no. 2011, pp. 1-19, 2011.

[10] K. Liang, H. Yang, and Y. Xing, "Comparision of projection domain, image domain, and comprehensive deep learning for sparse-view X-ray CT image reconstruction," *arXiv preprint arXiv:1804.04289*, 2018.

[11] H.-K. Yang, K.-C. Liang, K.-J. Kang, and Y.-X. Xing, "Slice-wise reconstruction for low-dose cone-beam CT using a deep residual convolutional neural network," *Nuclear Science and Techniques*, vol. 30, no. 4, p. 59, 2019.

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

3D landmark annotation for automatic 3D cephalometry

Hye Sun Yun, Tae Jun Jang, Sung Min Lee, Sang-Hwy Lee and Jin Keun Seo

***Abstract*—Annotation of three-dimensional (3D) cephalometric landmarks in 3D low dose cone beam computerized tomography (CBCT) has become an essential part of cephalometric analysis, which is used for diagnosis, surgical planning, and treatment evaluation. The automation of 3D landmarking with high-precision remains challenging due to the limited availability of training data and the high computational burden. This paper addresses these challenges by proposing a hierarchical deep-learning method consisting of four stages: 1) a deep-learning-based coarse-to-fine landmark annotator on the 2D cephalogram, 2) a basic landmark annotator for reconstructed 3D skull, 3) a low-dimensional representation of total landmarks using variational autoencoder (VAE), and 4) a local-to-global landmark annotator. The implementation of the VAE allows two-dimensional-image-based 3D morphological feature learning and similarity/dissimilarity representation learning of the concatenated vectors of cephalometric landmarks. The proposed method achieves an average 3D point-to-point error of 3.63 mm for 93 cephalometric landmarks with a small number of training CBCT datasets.**

***Index Terms*—Cone beam computerized tomography, cephalometry, landmark detection, deep learning**

I. Introduction

The cephalometric analysis offers morphometrical guideline for diagnosis, planning, and treatment of craniofacial disease or for evaluations in anthropological research. Recent advances in image processing techniques have allowed the annotation of three-dimensional (3D) cephalometric landmarks in 3D computerized tomography to become an essential clinical task. Since cephalometric landmarking is mostly done by manual identification, it requires a high level of expertise, time, and labor, even for experts. Therefore, there is increasing demand for a computer-aided automatic landmarking system capable of reducing the labor-intensiveness of this task and improving workflow.

Recent developments in deep learning for medical imaging have led to several attempts to apply it to an automatic cephalometric landmarking system [6] [8]. The main advantage of these deep learning approaches over conventional image processing is that the experts' experience can be reflected in the algorithm through learning training datasets. The architecture for learning datasets labeled by experts has allowed deep-learning-based methods to located landmarks in 2D cephalograms with impressive performance [1]. However, difficulties remain in applying these methods to 3D cephalometry because of the required number of datasets. According to Barron's observation [2] about approximations with shallow neural networks, the number of training datasets needed for learning grows tremendously as the input dimensions increase. Due to the high input dimensionality for dental CBCT data (typically $800 \times 800 \times 400$ matrix size), a lot more training datasets are required than are currently available, especially given the legal and ethical restrictions attached to medical data. This represents a significant challenge to the development of high-precision, automatic, 3D landmarking systems.

This paper reports the development and evaluation of an automatic annotation system for 3D cephalometry using limited training datasets. To overcome the dimensionality issue and limited data availability, a hierarchical deep learning method is developed which consists of four stages.The first stage employs deep learning-based detection of 2D landmarks on cephalogram based on coarse-to-fine detection. In the second stage, we fuse truncated 3D data and 2D cephalogram to reconstruct the information of whole skull data. Stage 3 applies the VAE to the landmark vectors to extract a disentangled low-dimensional representation. In stage 4, this low-dimensional representation then is used to map from fractional information of the landmarks (obtained in stages 1 and 2) to total information of the landmarks.

In this paper, the proposed method was evaluated by comparing the positional discrepancy between its results and the experts' results. Using a small number of training CBCT datasets, the proposed method achieved an average 3D point-to-point error of 3.63 mm for 93 cephalometric landmarks. The proposed method is therefore acceptable point-to-point error for assisting medical practice.

II. Method

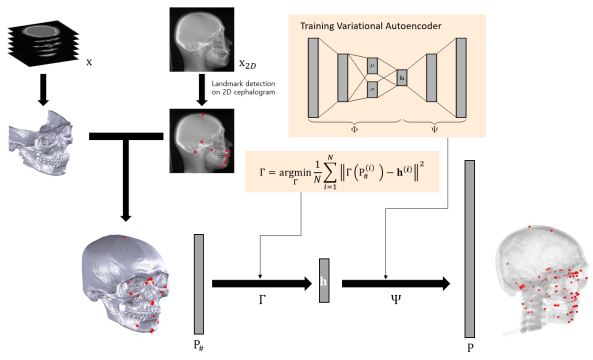


Fig. 1. Schematic diagram of the proposed method for 3D landmark annotation system.

Hye Sun Yun was with the Department of Computational Science and Engineering, Yonsei University, Seoul, South Korea, e-mail: (sally7711548@gmail.com)

Let $\mathbf{x}$ represents a three-dimensional CBCT data with voxel grid $\Omega := \{\mathbf{v} = (v_1, v_2, v_3) : v_j = 1, \cdots, d_j, j = 1, 2, 3\}$ with d_j being the number of voxels in direction v_j. In our case, CBCT data size is $800 \times 800 \times 400$. The value $\mathbf{x}(\mathbf{v})$ at voxel position $\mathbf{v}$ can be viewed as the attenuation coefficient. The goal is to find a map f from 3D CBCT image $\mathbf{x}$ to a group of 93 landmarks $\mathbf{P}$, where $\mathbf{P} = (\boldsymbol{p}_1, \cdots, \boldsymbol{p}_{93})$ and each $\boldsymbol{p}_j = (p_{j,1}, p_{j,2}, p_{j,3})$ denotes the 3D position of j-th landmark. $\mathbf{P}$ can be expressed as follows :

$$\mathbf{P} = ((p_{1,1}, p_{1,2}, p_{1,3}), (p_{2,1}, p_{2,2}, p_{2,3}), \cdots, (p_{93,1}, p_{93,2}, p_{93,3})) . \quad (1)$$

The group of landmarks $\mathbf{P}$ can be viewed as a geometric feature vector describing craniofacial skeletal morphology. Since direct automatic detection of $\mathbf{P}$ seems to be quite challenging, it would be desirable to infer $\mathbf{P}$ from its minimal information which is more convenient to detect automatically. Taking advantage of similarity of facial morphology, it would be reasonable to have a low dimensional latent representation for feature vectors. This could be achievable provided that the geometric feature vector can be expressed as low dimensional representation, which keeps crucial morphological features describing dissimilarities between data. Each step of the proposed method is illustrated in following subsections (also see Fig 1).

A. Stage 1: Detecting landmarks on 2D cephalogram

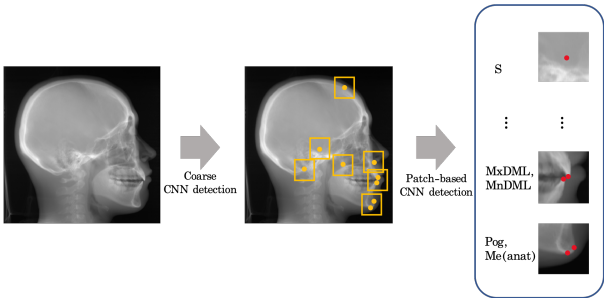


Fig. 2. Coarse-to-fine detection using entire image-based and patch-based CNN. 9 landmarks are rougly detected using the entire image-based CNN (coarse detection). Then, patch-based CNN (fine detection) is used to detect each landmark on the corresponding local patch obtained from output of entire image-based CNN.

The first step is to detect 9 landmarks on 2D cephalogram, (see Fig 3). The nine landmarks are all on or near the midsagittal plane and they are also used to estimate the total landmarks of the skull. Let $\mathbf{x}_{2D} \in \mathbb{R}^{512\times512}$ denote the 2D cephalogram and $\mathrm{P}_{\mathrm{loc}} = (v_{2,1}, v_{3,1}, \cdots, v_{2,9}, v_{3,9})$ denote 9 landmarks on the 2D image. With the dataset $\{(\mathbf{x}_{2D}^{(i)}, \mathrm{P}_{\mathrm{loc}}^{(i)}) \mid i = 1, 2, ..., N\}$, we train a network which detects 9 landmarks on 2D image using convolutional neural network (CNN). However, the number of data is insufficient to exactly detect the landmarks. To address this problem, we use coarse-to-fine detection approach using entire image-based CNN and patch-based CNN. The architectures of CNNs will be described in result section in Fig 7. The entire image-based CNN allows to detect $\mathrm{P}_{\mathrm{loc}}$ roughly by capturing global information. This coarse detection output is used to generate local patches, which are input of the patch-based CNN. The patch-based CNN provides $\mathrm{P}_{\mathrm{loc}}$ with improved accuracy. See Fig 2.

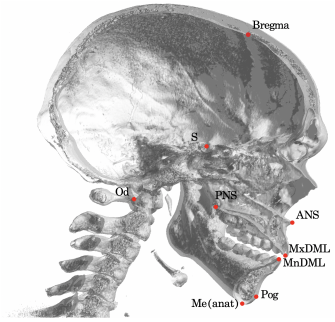
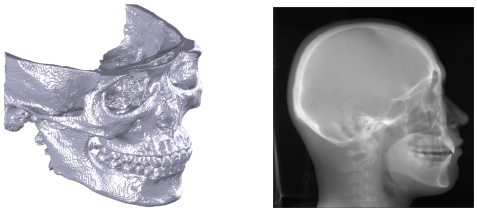


Fig. 3. Landmarks near the midsagittal plane. These landmarks are to be identified by using coarse-to-fine CNN.

B. Stage 2 : Data fusion



(a) (b)

Fig. 4. (a)Truncated 3D skull data obtained from dental CBCT and (b)2D cephalogram. These two data are fused to reconstruct the whole skull data.

The proposed method will be performed under dental CBCT hardware limitations such as the field of view (FOV) truncation and offset detector. As shown in Fig 4, cranial bone part is missing for the dental CBCT data without stitching. To compensate this missing part in CBCT, we use the 2D cephalogram that will be combined with CBCT to get the complete skull image. We developed a novel method integrating 2D cephalogram with truncated skull image, in order to fill in the missing cranial part. Then the reference landmarks on the complete skull image can be obtained by an existing approach of multiple shadowed 2D-image-based landmarking [6].

With the reconstructed skull, we normalized the data for effective feature learning. The normalization is done through suitable anisotropic scaling based on facial width (the distance between the x-coordinate of L Po and R Po), facial depth (the distance between y-coordinate of L Po and nasion), and facial height (the distance between z-coordinate of CFM and bregma). This normalization reveals facial deformities, and enables efficient feature learning of similarity/dissimilarity in the third stage when applying VAE.

C. Stage 3: Learning a low-dimensional latent representation of landmarks

In this stage, a low-dimensional latent representation of the total landmark vector $\mathbf{P}$ is obtained by applying the VAE [3]

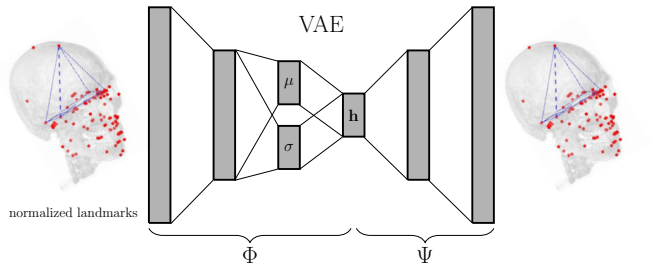


Fig. 5. Architecture of VAE-based low dimensional representation. Normalized landmarks are used to generate semantically useful latent space.

to the landmark feature vector. This facilitates the extraction of exploitable morphological factors.

In mathematical terms, the VAE is a deep learning technique that finds a nonlinear expression of the concatenated landmark vector $\mathbf{P} \in \mathbb{R}^k$ by variables $\mathbf{h} \in \mathbb{R}^d$ $(d \ll k)$ in the low-dimensional latent space. In our experiments, we use $k = 279$ and $d = 25$. The VAE uses the traning datasets $\{\mathbf{P}^{(i)} : i = 1, \cdots, N\}$ to learn two functions, the encoder $\Phi : \mathbf{P} \to \mathbf{h}$ and the decoder $\Psi : \mathbf{h} \to \mathbf{P}$ by the following loss minimization over the training data:

$$(\Psi, \Phi) = \underset{(\Psi,\Phi)\in \mathbb{VAE}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} [\|\Psi \circ \Phi(\mathbf{P}^{(i)}) - \mathbf{P}^{(i)}\|^2 + D_{KL}(\mathcal{N}(\mu^{(i)}, \Sigma^{(i)}) \parallel \mathcal{N}(0, I))], \quad (2)$$

where $\mathbb{VAE}$ describes a class of functions in the form of the deep learning network described in Fig 5. Note that the covariance Σ and $D_{KL}(\mathcal{N}(\mu, \Sigma) \parallel \mathcal{N}(0, I))$ are used for smooth interpolation and compactly encoding.

The decoder $\Psi : \mathbf{h} \to \mathbf{P}$ in (2) provides a low-dimensional disentanglement representation, so that each latent variable is sensitive to changes in individual morphological factors, while being relatively insensitive to other changes.

D. Stage 4: Local-to-global landmark annotation for automatic 3D cephalometry

In this final step, we detect the total landmark vector ($\mathbf{P}$) from the fractional information ($\mathrm{P}_\sharp$), where $\mathrm{P}_\sharp$ is the vector composed of $\mathrm{P}_{\mathrm{loc}}$ obtained in stage 2 and the reference landmarks in stage 1. In stage 3, the VAE can find a low-dimensional latent representation of the total landmarks, i.e., $\Psi(\mathbf{h}) = \mathbf{P}$. Stage 2 detected a part of the landmarks $\mathrm{P}_{\mathrm{loc}}$ near the midsagittal plane. Using the encoder map $\mathbf{h}^{(i)} = \Phi(\mathbf{P}^{(i)})$ in the result of stage 3, the training data $\{(\mathbf{h}^{(i)}, \mathrm{P}_\sharp^{(i)}) : i = 1, 2, ..., N\}$ can be generated. Then, the training data can be used to learn a nonlinear map $\Gamma : \mathrm{P}_\sharp \to \mathbf{h}$, which connects the latent variables $\mathbf{h}$ and the fractional data $\mathrm{P}_\sharp$. The nonlinear regression map $\Gamma : \mathrm{P}_\sharp \to \mathbf{h}$ is obtained by minimizing the loss $\frac{1}{N}\sum_{i=1}^{N} ||\Gamma(\mathrm{P}_\sharp^{(i)}) - \mathbf{h}^{(i)}||^2$. Then, the local-to-global landmark annotation is obtained from

$$(\Psi \circ \Gamma)(\mathrm{P}_\sharp) = \mathbf{P}. \quad (3)$$

See Fig 6.

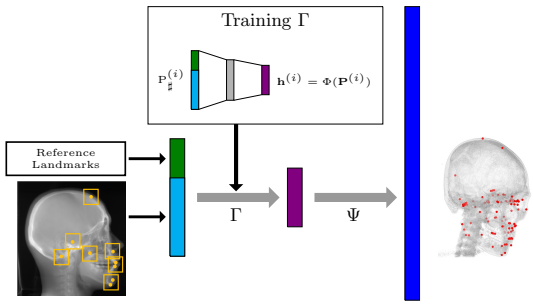


Fig. 6. Local-to-global landmark annotation $\Psi \circ \Gamma$. The map $\Gamma : \mathrm{P}_\sharp \to \mathbf{h}$ is trained using the dataset $\{(\mathbf{h}^{(i)} = \Phi(\mathbf{P}^{(i)}), \mathrm{P}_\sharp^{(i)}) : i = 1, \cdots, N\}$ generated from the encoder Φ in stage 3.

III. Results

A. Dataset and Experimental setting

We used two datasets, provided by Dentri, HDXWILL, Korea; the first dataset contains 26 anonymized CBCT data and 2D cephalogram corresponding to the same subjects with cephalometric landmarks for each data. The size of the original CBCT image was $800 \times 800 \times 400$. The labeling for the landmarks on each data was done through manual marking. The second dataset with 3D positions of landmarks from anonymized 229 subjects with dentofacial deformity and malocclusion was also used, and they were acquired in excel formats with 3D coordinates of landmarks from the original data source. The labeling of the landmarks for both data sets were all done by one author (LSH) with more than 20 years of experience in 3D cephalometry. When training CNN, we used the first dataset of 22 subjects as training data and four as test data. Translation was applied for data augmentation. For VAE and nonlinear regression, we used both the first and the second dataset to set 230 subjects as training data and 25 for test data. For each experimental setting, we set learning rate as 0.0001 and batch size 8, and went through 3000 iterations when training CNN. For VAE and nonlinear regression, we set learning rate as 0.0001 and batch size 50, and went through 30,000 iterations.

B. Experimental results

In our proposed method, we aimed to locate 3D coordinates of 93 landmarks from fractional information composed of $\mathrm{P}_{\mathrm{loc}}$ obtained in stage 1 and the reference landmarks in stage 2.

For the detection of 9 landmarks on 2D cephalogram, the architecture for image-based CNN are shown in Fig 7(a). With the input data of size 512×512, the architecture consists of 3×3 convolution with stride 1, 2×2 convolution with stride 2 for spatial downsampling, and fully connected layers. Rectified Linear Unit (ReLU) activation was performed after each pooling layer to solve vanishing gradient problem, dropout rate of 0.75 was chosen to alleviate overfitting problem [7], and Adam [4], which is an adaptive gradient method, was chosen as the optimization algorithm. Additionally, we extracted local patch using the output obtained from image-based CNN. For fine detection of each landmark, CNN architecture was again designed, similar to that of image-based CNN (see Fig 7(b)).

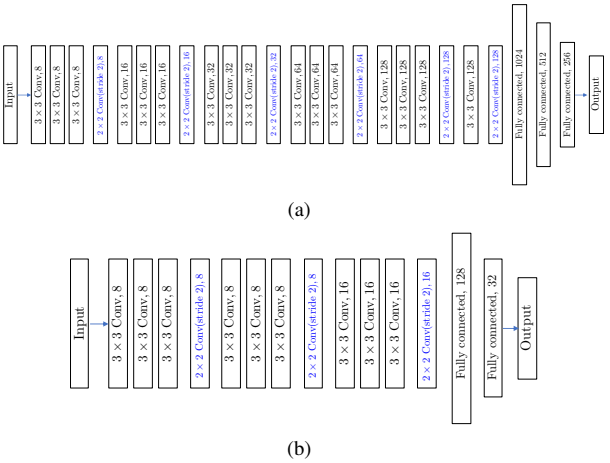


Fig. 7. Architectures of (a) entire image-based CNN and (b) patch-based CNN.

The size of patch (as input for patch-based CNN) was chosen by the characteristics of morphological structures around the landmarks. With small amount of data at our hands, it was inevitable to apply landmark detection additionally on the small patches for better feature learning for each landmark. Table I shows the prediction error value of each landmark.

Landmark	2D distance error (mm)
ANS	0.88
Bregma	0.98
Me(anat)	0.62
MnDML	1.45
MxDML	1.39
Od	2.73
PNS	2.63
Pog	1.51
S	3.75

TABLE I
MEAN OF TEST ERROR FOR 9 LANDMARK DETECTION USING COARSE-TO-FINE ANNOTATOR.

As a core work of our study, we used VAE to find the latent representation of landmark feature vector. We aim to find the low dimensional representation of high dimensional landmark feature vector. The latent dimension of latent space is chosen to be 25 empirically, which tells that the landmark feature vectors ($\in \mathbb{R}^{279}$) can be nonlinearly expressed into 25 variables. Next, we connected the landmarks detected in previous steps to the trained representation by nonlinear regression with multilayer perceptron having 45-30-25 neurons in each layer. After finishing training VAE and nonlinear regression, the reconstructed landmark vectors are given by $(\Psi \circ \Gamma)(\mathrm{P}_\sharp)$. As a result, we achieved an average point-to-point error of 3.63 mm for 93 cephalometric landmarks with standard deviation 1.41 mm. Fig 8 shows test error of each landmark. Total points showed the errors ranging within 8 mm and 60 landmarks were within 4 mm.

IV. CONCLUSION

This paper proposes a multi-stage deep learning framework for automatic 3D cephalometric landmark annotation. Its novel contribution is the use of a VAE for 2D image-based 3D

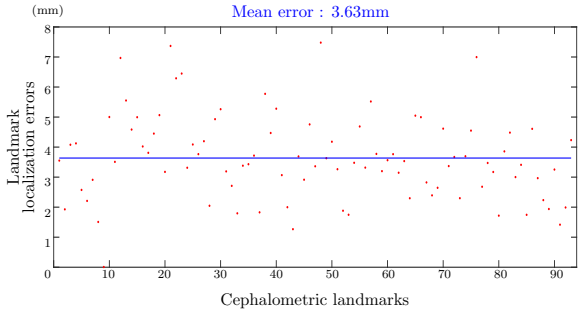


Fig. 8. The landmark localization errors (mm) for 93 cephalometric landmarks for the test dataset. Red dots denote the 3D distance error for each landmark, and the blue line represents the average point-to-point error for 93 landmarks (3.63 mm).

feature learning by representing the high-dimensional 3D landmark feature vectors with much-lower-dimensional latent variables. It allows all 3D landmarks to be annotated from partial information obtained through 2D landmarking on 2D cephalogram and the detection of 3D landmarks using the existing method. The experiments confirmed the capability of the proposed method, even when given a limited number of training datasets. The proposed method has the potential to alleviate experts' time-consuming workflow by dramatically reducing the time taken for landmarking while preserving high accuracy.

This automatic 3D cephalometric annotation system is in an early stage of development, and there is potential for further improvement. It would be desirable to integrate the multi-stage hierarchical learning framework of this study into a unified learning framework, because errors in each step can affect the successive steps in the hierarchical structure.

ACKNOWLEDGEMENTS

This work was supported by the Samsung Science & Technology Foundation.

REFERENCES

[1] S.Ö. Arik, B. Ibragimov and L. Xing 2017 Fully automated quantitative cephalometry using convolutional neural networks *J Med Imaging (Bellingham)* **4(1)** 014501
[2] A.R. Barron 1994 Approximation and estimation bounds for artificial neural networks *Machine learning* **14(1)** 115-133
[3] D.P. Kingma, M. Welling, 2013 Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*
[4] D. P. Kingma, J. Ba, 2014 Adam: A method for stochastic optimization *arXiv preprint arXiv:1412.6980*
[5] S.-H. Lee, T.-J. Kil, K.-R. Park, B.C. Kim, Z. Piao, and P. Corre 2014 Three-dimensional architectural and structural analysis-a transition in concept and design from Delaire's cephalometric analysis *Int J Oral Maxillofac Surg* **43** 1154-1160
[6] S.M. Lee, H.P. Kim, K. Jeon, S.H. Lee and J.K. Seo 2019 Automatic 3D cephalometric annotation system using shadowed 2D image-based machine learning *Physics in medicine and biology*
[7] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and , R. Salakhutdinov 2014. Dropout: a simple way to prevent neural networks from overfitting *The Journal of Machine Learning Research*, **15(1)** 1929-1958
[8] H.S. Yun, T.J. Jang, S.M. Lee, S.H. Lee and J.K. Seo 2020 Learning-based local-to-global landmark annotation for automatic 3D cephalometry *submitted to Physics in medicine and biology*

Disassemblable Fieldwork CT Scanner Using a 3D-printed Calibration Phantom

Florian Schiffers, Thomas Bochynek, André Aichert, Tobias Würfl, Michael Rubenstein, Oliver Cossairt

***Abstract*—The use of computed tomography (CT) imaging has become of increasing interest to academic areas outside of the field of medical imaging and industrial inspection, e.g. to biology and cultural heritage research. The pecularities of these fields, however, sometimes require that objects need to be imaged on-site, e.g., in field-work conditions or in museum collections. Under these circumstances, it is often not possible to use a commercial device and a custom solution is the only viable option. In order to achieve high image quality under adverse conditions, reliable calibration and trajectory reproduction are usually key requirements for any custom CT scanning system. Here, we introduce the construction of a low-cost disassemblable CT scanner that allows calibration even when trajectory reproduction is not possible due to the limitations imposed by the project conditions. Using 3D-printed in-image calibration phantoms, we compute a projection matrix directly from each captured X-ray projection. We describe our method in detail and show successful tomographic reconstructions of several specimen as proof of concept.**

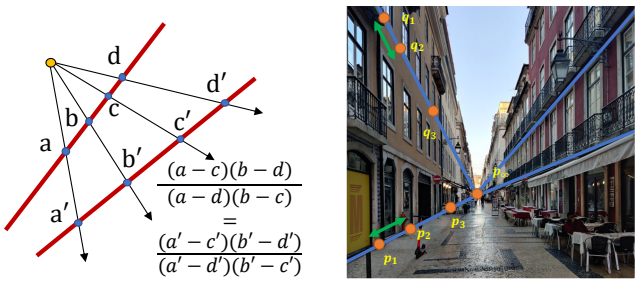
***Index Terms*—Industrial CT, low-cost calibration, disassemblable CT scanner.**

I. Introduction

WHEN computed tomography (CT) was invented in 1971 it quickly became an essential tool for non-destructive testing and medical diagnosis. The proliferation of CT hardware and software in recent decades has promoted widespread interest from the greater scientific community, finding new applications in a broad range of fields such as cultural heritage [1], [2] and biology [3].

While industrial CT applications often feature optimized and standardized imaging setups that allow high-throughput imaging, e.g. in a hospital setting or a production line, the scientific community operates under a fundamentally different set of constraints than medical and industrial applications. For instance, biological and cultural heritage research often prohibits samples from being transported to the nearest CT facilities, requiring scanning on-site.

Under these circumstances existing commercial products are unsuitable, and a custom-built, portable system is necessary. Of particular interest here is the case where the imaging system must be shipped between sessions, and repeated calibration is necessary. As a potential solution to this problem, Aichert *et al.* [4] proposed an online geometry calibration method using in-image fiducial markers for flat-panel computed tomography based on projective invariants that can be easily realized with a 3D-printed metal bead phantom. They demonstrate a simulation-based proof-of-concept for a calibration phantom designed for computing both extrinisic and intrinsic camera parameters from a single projection image.



(a) Cross-Ratio (b) Cross-section to calculate relative distances

Fig. 1: Schematic principle of the projective invariant.

This paper describes the construction of a novel and disassemblable cone-beam computed tomography (CBCT) scanner, which for the first time, applies the calibration method proposed in [4]. We show reconstructions of biological specimen imaged on Barro Colorado Island, a remote, tropical island in Panama. This work demonstrates that customized imaging and calibration setups allow for the acquisition of industry-standard quality images under demanding imaging environments.

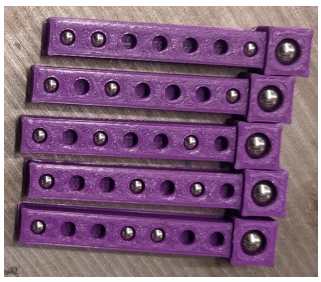
II. Methods

A. Cone Beam Projection Model

The most popular trajectory takes the shape of an arc of a circle. Such trajectories depend on few parameters compared to the calibration of the geometry of each projection image [5], [6]. Therefore it is appealing to assume an ideal smooth trajectory and calibrate only those parameters as opposed to estimating a geometry for each projection image independently. This simplification, however, comes at the cost of no longer being able to compensate for deviations from this ideal smooth trajectory. An alternative is to describe each projection matrix independently e.g. [5]. For those methods, the image formation of a conventional X-Ray source with a cone beam and flat-panel detector is typically modeled as a projection matrix.

In an industrial CT setup, the focal length of the system and the principal point, which is the intersection of the orthogonal projection of the source to the detector, are stable because

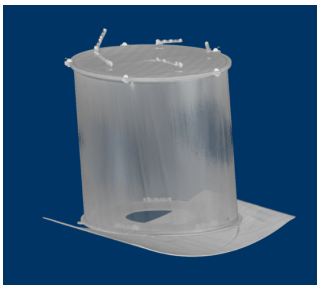
F. Schiffers and T. Bochynek are with the Department of Computer Science, Northwestern University, IL, Evanston, 60201, USA
M. Rubenstein and O. Cossairt are with the Department of Computer Science and Electrical and Computer Engineering, Northwestern University, IL, Evanston, 60201, USA
A. Aichert is with Digital Technology Innovation, Siemens Healthineers, Erlangen, Germany
T. Wuerfl is with the Chair of Pattern Recognition, FAU Erlangen, Germany.
Corresponding author: florian.schiffers@northwestern.edu



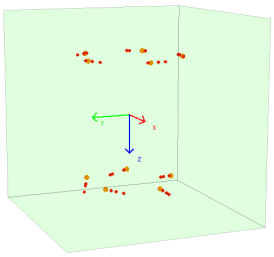
(a) 3D-printed markers with unique metal bead placement to achieve different cross-ratios.

(b) Sample container with calibration phantom consisting of ten calibration markers

Fig. 2: 3D-printed calibration markers, in detail and assembly.



(a) CT of 3D-printed calibration phantom with metal beads.



(b) Extracted 3D positions of metal beads.

Fig. 3: The 3D-printed calibration is imaged with a commercial CT scanner at a hospital. The 3D locations of the metal beads are extracted from DICOM-series.

the source and detector do not move during the acquisition. Therefore the intrinsic camera parameters can be measured as fixed in the whole system, and only the extrinsic camera parameters need to be determined for each projection.

B. Estimation of Geometric Parameters

A significant challenge for tomographic calibration is to build a 3D-phantom with interest points that can be uniquely identified under a projective transformation in different projections. A pair of coordinates of the same 3D interest point in the projections are called corresponding points and the problem of computing them is commonly referred to as the correspondence problem. Strobel *et al.* [7] proposed the *PDS-2* phantom where binary codes of small metal beads are encoded to solve the correspondence problem. Choi *et al.* [8] solve this problem using the assumption that markers only move a small distance in successive projections.

These and similar methods usually acquire a dedicated calibration scan, since CT scanners are manufactured to consistently reproduce the same motion.

In contrast, Aichert *et al.* [4], [9] present a method relying on the cross-ratio to solve the correspondence problem. If the phantom is present in the acquisition the method can be used to calculate a projection matrix for each X-Ray projection, which removes the necessity of reproducible trajectories. This approach meets the requirement of robust reconstruction under demanding field conditions. Each phantom consists of multiple, short calibration elements of four uniquely spaced metal beads each.

These elements can be arranged arbitrarily within the projection space but in fixed relation to the sample (e.g. by attaching them to a sample container). The respective cross-ratio of the metal beads, depicted in Fig. 1, is used as a scalar descriptor for each such calibration element. It is defined by the distances $a,b,c,d \in \mathbb{R}$ between the four metal beads in a linear arrangement.

$$\mathrm{cr}(a,b,c,d) := \frac{(a-c)\cdot(b-d)}{(a-d)\cdot(b-c)}. \qquad (1)$$

Because this number is invariant to perspective projection, it can be used to identify corresponding points in different projections. A RANSAC-based algorithm is used to estimate all geometric parameters from a single image. For a detailed discussion of the mathematical methodology and how the cross-ratio is used for calibration, we refer to [4].

C. Phantom Design and Manufacture

We fabricated ten calibration elements of six-centimeter length, each containing a unique arrangement (and thereby cross-section) of four metal beads. To provide directionality, the diameter of the lower-most bead was twice that of the remaining beads (i.e., 3.2 mm and 1.6 mm, respectively).

We 3D-printed the base marker shape from Polylactic acid. Metal ball bearings were press-fitted into predefined recesses. Fig. 2a exemplary shows five of those calibration sticks. The assemblage of the ten sticks in 3D space is such that at least three are fully visible on the detector for any projection image of the trajectory.

We placed five of these calibration elements, each to both the top and bottom of the sample container, see Fig. 2b. Equidistant arrangement of the elements ensured that there would be three markers identifiable for any projection image of the trajectory, removing marker overlap as a possible complicating factor. To avoid overlap of markers and the sample specimen, we reserve the top and bottom of the projection for the calibration phantom. Note that this effectively trades image information for a reliable and accurate online estimation of geometric parameters.

D. Hospital CT-scan of Calibration Phantom

The algorithm proposed in [4] requires knowledge of the exact 3D positions of all beads in the phantom. While it is known from computer vision that these positions could be inferred algorithmically, we instead acquired a reference scan of the whole calibration phantom with a commercial CT scanner in the Northwestern Memorial Hospital (see Fig. 3a).

The spheres are identified from the DICOM series via the application of the connected components method. Subsequently we calculate the center of mass of each sphere as shown in Fig. 3b. Note, that the marker accuracy is given not by the accuracy of the 3D-printing process, but by the identification of the circle centers.

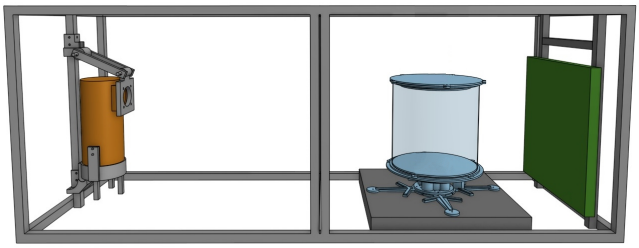


Fig. 4: Computer model of the scanner setup. Mount for the X-Ray source (left) and rotation stage (center right) are fabricated from 3D-printed components. See text for details on the used components.

E. Experimental Setup

The CT scanner setup (see Fig. 4) is composed of four main components: The imaging container with the calibration phantom is located inside an 80/20 T-slot aluminum-frame cabinet, which is shielded by fused lead/steel plates for radiation protection. We used a *Imaging M-113T X-Ray tube* and a *Varex Imaging XRD 4343 Digital X-Ray detector* for radiation generation and detection. A rotation stage, driven by a *1000:1 Micro Metal Gearmotor HP 6V*, rotated the imaging container around its axis. Rotation stage, X-Ray tube, and X-Ray detector were controlled and synchronized via an *Arduino Uno microprocessor*.

A single 3D scan consisted of approximately 450 exposures of the rotating imaging container at a rate of 4.5 exposures per second. A single exposure consisted of a 10.6 ms X-Ray burst at 40 keV energy and 125 mA current supplied by a *Spellman PMX 5kW X-Ray generator*.

III. Reconstruction Pipeline

We implemented the reconstruction pipeline in Python and embedded our pipeline into the open-source framework of pyCONRAD *(https://git5.cs.fau.de/PyConrad/pyCONRAD)* which natively supports arbitrary trajectories via projection matrices [10]. Since pyCONRAD has the full GPU functionality of CONRAD, fast reconstruction with the GPU-accelerated back-projector is possible. For reconstruction, we use the Feldkamp algorithm [11].

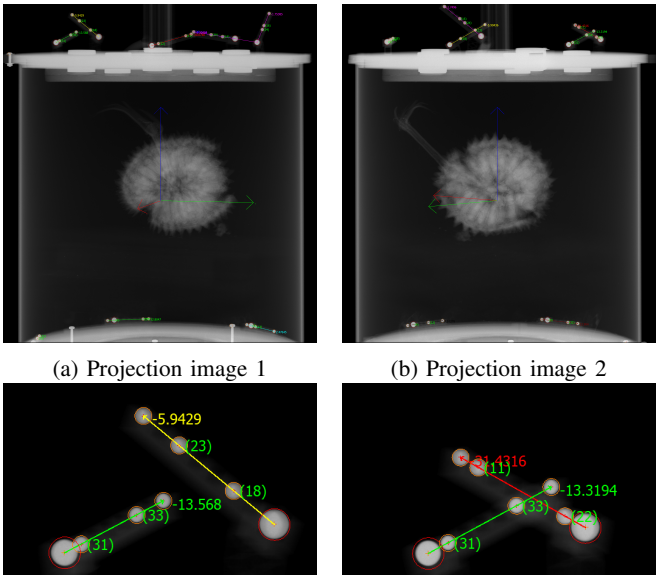
A. Image Pre-Processing

Raw images are normalized using the following normalization equation:

$$f = \frac{I - I_{Dark}}{I_0 - I_{Dark}} \tag{2}$$

where I denotes the measured intensity with the object, I_0 the reference intensity without object, and I_{Dark} the measured dark-image.

Our detector has approximately 100 permanent defect pixels which either report full or zero intensity in every exposure. If not accounted for, these pixels will result in strong ring-artifacts in the reconstructions. We hence detect those pixels with a prior calibration measurement and correct them using OpenCV's inpainting method.



(a) Projection image 1 (b) Projection image 2

(c) Close-up showing reprojection error 1 (d) Close-up showing reprojection error 2

Fig. 5: (a,b) Projection images with biological specimen (*Annona purpurea* fruit) and calibration markers. Detected calibration elements and the derived local coordinate system are shown in the center of each image. (c,d) show a zoomed version for better visualization of the detected beads' reprojection.

B. High resolution volume reconstruction

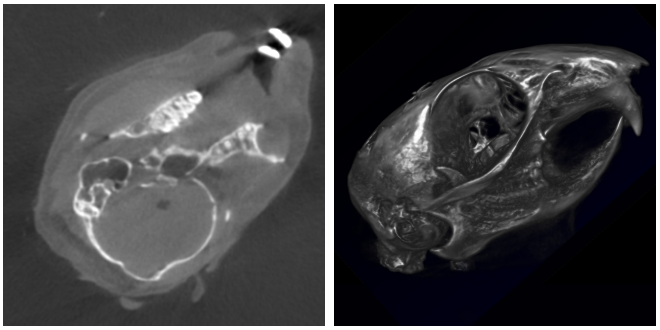
Our detector allows for a resolution of up to 3000×3000 pixels, generating large volumes of data of both raw data and reconstructions. For example, a reconstructed volume of 2000^3 with float precision requires approximately 32 GB of storage space. PyCONRAD does not natively support reconstruction volumes of this size because it is limited by the GPU memory. Reconstruction of large high-resolution volumes was achieved by successively reconstructing and storing smaller volume subsets.

C. Results

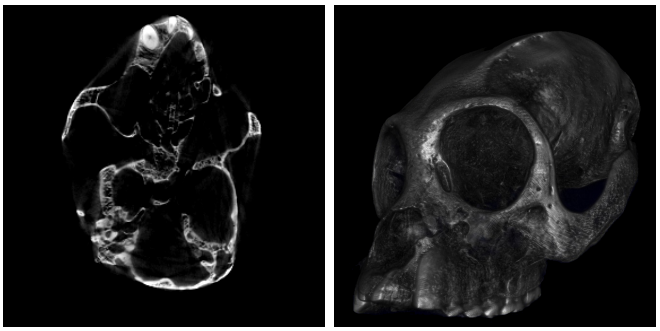
Figure 5 shows an example of an online-calibration with bead detection, feature matching using the cross-ratios, and successful geometric calibration for two different views. The complete trajectory was calibrated and is visualized in Fig. 7. Example tomographic reconstructions and renderings for four specimens found on-site at the Barro Colorado Island are shown in Fig. 6.

IV. Discussion and Conclusion

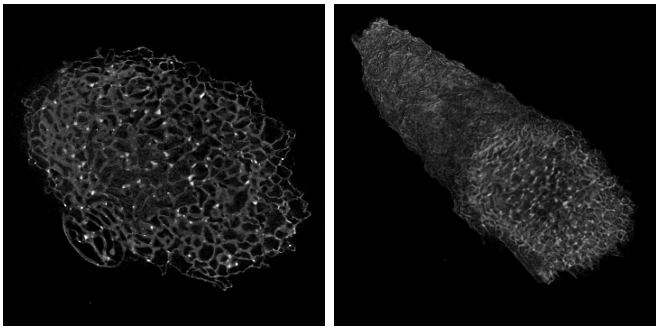
We constructed a 3D CT scanner and online calibration phantom from commonly available materials using 3D printing as a key manufacturing method. While the methods we employ in the construction process might not meet the high-accuracy manufacturing requirements applied to medical devices, we still obtain high-quality 3D reconstructions. In contrast to one-time calibrated and commercial scanners, our setup is



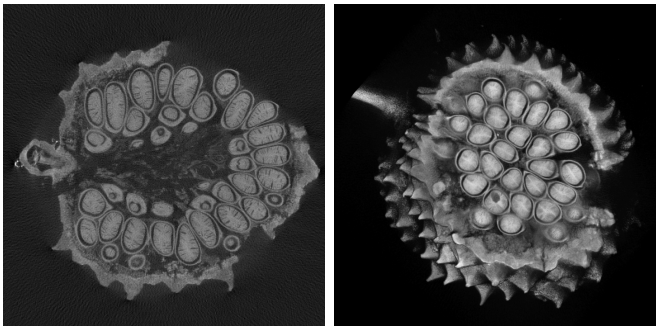
(a) *Tamiasciurus sp.* squirrel skull



(b) *Alouatta palliata:* howler monkey skull



(c) *Nausititermes sp.* termite nest

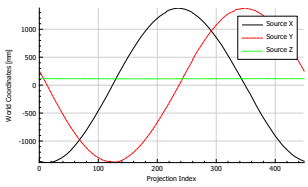


(d) *Alouatta palliata:* tropical fruit

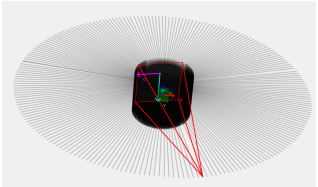
Fig. 6: Example reconstruction from our tomographic scanner. Left: Tomographic slice; right: volume rendering.

resilient to adverse conditions, disassemblable for transport to remote locations, and was successfully operated for field-work in a tropical environment. We intend that this paper can serve as a guideline for interested parties to develop their own applications.

Future research directions include reducing the amount



(a) Calculated source position



(b) Calculated trajectory

Fig. 7: Visualization of the trajectory calculated online during a tomographic scan. The image reveals that the source positions do not form perfect sinusoids, and the angular spacing is not equidistant. While this would be a challenge to a one-time calibrated setup, the online calibration we employ compensates for these inaccuracies.

of detector space required for online calibration, employing bundle-adjustment to lift the requirement of a prior CT-scan of the calibration phantom, using our methodology on a theoretically complete trajectory like a helix, additional artifact reduction techniques, such as beam-hardening correction, as well as the use of image-based calibration correction methods.

References

[1] D. Stromer, V. Christlein, X. Huang, P. Zippert, T. Hausotte, and A. Maier, "Virtual cleaning and unwrapping of non-invasively digitized soiled bamboo scrolls," *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019.

[2] M. P. Morigi, F. Casali, M. Bettuzzi, R. Brancaccio, and V. D'Errico, "Application of X-ray Computed Tomography to Cultural Heritage diagnostics," *Applied Physics A: Materials Science and Processing*, vol. 100, no. 3, pp. 653–661, 2010.

[3] K. Singh, B. P. Muljadi, A. Q. Raeini, C. Jost, V. Vandeginste, M. J. Blunt, G. Theraulaz, and P. Degond, "The architectural design of smart ventilation and drainage systems in termite nests," *Science Advances*, vol. 5, no. 3, pp. 1–12, 2019.

[4] A. Aichert, B. Bier, L. Rist, and A. Maier, "Projective invariants for geometric calibration in flat panel computed tomography," *Proc. 5th Intl. Mtg. on Image Formation in X-ray CT*, pp. 69–72, 2018.

[5] L. Von Smekal, M. Kachelrieß, E. Stepina, and W. A. Kalender, "Geometric misalignment and calibration in cone-beam tomography," *Medical Physics*, vol. 31, no. 12, pp. 3242–3266, 2004.

[6] M. J. Daly, J. H. Siewerdsen, Y. B. Cho, D. A. Jaffray, and J. C. Irish, "Geometric calibration of a mobile C-arm for intraoperative cone-beam CT," *Medical Physics*, vol. 35, no. 5, pp. 2124–2136, 2008.

[7] N. K. Strobel, B. Heigl, T. M. Brunner, O. Schuetz, M. M. Mitschke, K. Wiesent, and T. Mertelmeier, "Improving 3D image quality of x-ray C-arm imaging systems by using properly designed pose determination systems for calibrating the projection geometry," *Medical Imaging 2003: Physics of Medical Imaging*, vol. 5030, no. June 2003, p. 943, 2003.

[8] J. H. Choi, A. Maier, A. Keil, S. Pal, E. J. McWalter, G. S. Beaupré, G. E. Gold, and R. Fahrig, "Fiducial marker-based correction for involuntary motion in weight-bearing C-arm CT scanning of knees. II. Experiment," *Medical Physics*, vol. 41, no. 6, pp. 1–16, 2014.

[9] A. Aichert, T. Würfl, and A. K. Maier, "Method and device for determining a geometry calibration for an imaging device and also method for determining assignment data for the geometry calibration," 2019. US Patent Application US2019/0355156 A1.

[10] A. Maier, H. G. Hofmann, M. Berger, P. Fischer, C. Schwemmer, H. Wu, K. Müller, J. Hornegger, J. H. Choi, C. Riess, A. Keil, and R. Fahrig, "CONRAD - A software framework for cone-beam imaging in radiology," *Medical Physics*, vol. 40, no. 11, 2013.

[11] L. a. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, p. 612, jun 1984.

Metal Artifact Reduction for 3D Cone-Beam CT by Non-Local Prior Image Integration

Nele Blum, Maik Stille, Christian Hofmann and Thorsten M. Buzug

***Abstract*—While the quality of CT images has been continuously improved over the years, the presence of high-density objects, such as metal implants, still remains a major problem. Often standard approaches for metal artifact reduction are not able to correct the artifacts caused sufficiently and bright and dark streaks as well as dark shadows remain as artifacts in the image. Based on a non-local regularization, a new iterative reconstruction algorithm for 3D cone-beam CT data is proposed. In the regularization term, 3D image patches of the image and prior image are compared. For optimization, both the regularization term and the derivative of the regularization term are calculated. To avoid long computation time, parts of the algorithm are implemented to be run parallel. The usage of the proposed non-local regularization term shows an improvement in artifact reduction compared to the linear interpolation approach.**

I. Introduction

COMPUTED tomography provides high-quality images of the interior of the human body using x-ray. However, artifacts caused by different physical phenomena such as scatter, beam hardening or noise, reduce the image quality. A major problem in clinical practice are high-density objects, such as metal implants or surgical instruments in the field of view. Such objects cause artifacts that appear as bright and dark streaks in the image and decrease the diagnostic value of the image.

Standard metal artifact reduction (MAR) approaches often can not reduce this artifacts sufficiently or introduce new artifacts. One method to reduce these artifacts is to use an iterative reconstruction approach [1]. This has the advantage of easy integration of additional information. The integration of prior images in the CT image reconstruction has become popular in the reconstruction field since the PICCS algorithm was published by Chen et. al. [2]. A regularization strategy, which integrates information from a prior image by a non-local regularization term, was already introduced for 2D image reconstruction [3]. Since then the regularization term has been improved and presented here for 3D cone-beam reconstruction. During regularization, differences between grey values in the image of the current reconstruction step and the non-local prior image are penalized. The idea of the regularization term is to penalize variations to the prior image, while keeping information of the anatomical structure of the original image. However, the prior image should be similar to the image to be reconstructed in order to save calculation time and guarantee a sufficient penalization. Parts of the algorithm are implemented to be calculated on a GPU to avoid long computation time when using large 3D images. Thus, the regularization term was implemented in such a way that the calculation is executed in parallel and a fast comparison of the image and prior image patches is possible. Together with a fast calculation of forward and back projection operators for 3D cone-beam CT in each iteration step, the algorithm becomes interesting for clinical routine.

N. Blum, M. Stille and T. M. Buzug are with the Institute of Medical Engineering, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, e-mail: (see {blum, stille, buzug}@imt.uni-luebeck.de).

C. Hofmann is with Siemens Healthineers, Henkestraße 127, 91052 Erlangen, e-mail:(see: christian.hofmann@siemens-healthineers.com)

II. Methods

The log-likelihood function for transmission tomography in statistical image reconstruction is defined as

$$l(f) = \sum_{i=1}^{M}(-n_i \ln(n_0) + n_i \sum_{j=1}^{N} a_{ij} f_j + ln(n_i!) + n_0 \exp(-\sum_{j=1}^{N} a_{ij} f_j) \quad (1)$$

for a given set of intensity measurements n_i and an unknown vector $f \in \mathbb{R}^N$, which consist of the expected attenuation coefficients [4]. The number of measured intensities in the absence of an absorber is denoted by n_0, the total number of measured projections is denoted by M and the number of image voxels is denoted by N. In (1) the forward projection step is described by

$$p_i = \sum_{j=1}^{N} a_{ij} f_j \quad (2)$$

where a_{ij} is the element at position (i, j) of the system matrix $A \in \mathbb{R}^{M \times N}$. The log-likelihood function is part of the objective function of the optimization problem, which needs to be solved for the image reconstruction. The corresponding optimization problem is defined as

$$f^* = \arg\min(l(f) + \delta R(f, \Gamma(h, \gamma))) \quad (3)$$

where $R(f, \Gamma(g, \gamma))$ describes a regularization term. The regularization term is added to the log-likelihood function. To control the influence of the regularization on the reconstruction result, a regularization parameter δ is multiplied to the regularization term.

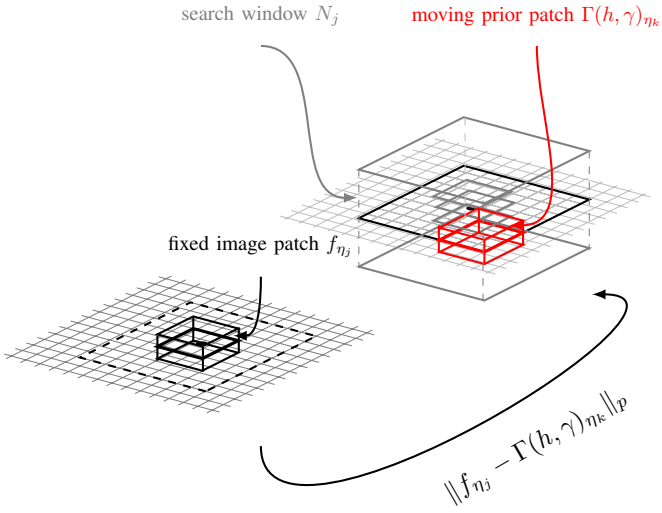


Fig. 1. Comparison of 3D patches between the image to be reconstructed and the prior image. The fixed image patch (black box) is compared to different prior patches (red box) inside a 3D search window of the prior image by calculating the distances between the voxel values inside the patches. The 3D search window in the prior image is visualized by a gray box.

The regularization is here proposed as a non-local regularization

$$R(f,\Gamma(h,\gamma)) = \sqrt{\sum_{j=0}^{N} \iota_j \left(f_j - \frac{1}{\kappa_j} \Psi\left(f,\Gamma(h,\gamma),j\right)\right)^2} \quad (4)$$

with

$$\Psi\left(f,\Gamma(h,\gamma),j\right) = \sum_{k\in N_j} \Gamma(h,\gamma) \exp\left(\frac{-\sigma\|f_{\eta_j} - \Gamma(h,\gamma)_{\eta_k}\|_p}{v^2}\right) \quad (5)$$

where $\Gamma(h,\gamma)$ denotes the transformed prior image h by the transformation parameter γ. Figure 1 shows the principle of the non-local regularization term described in (4). In the illustration, N_j (gray box) denotes a search window in the prior image h around the voxel j. Each patch $\Gamma(h,\gamma)_{\eta_k}, k \in N_j$ (red box) within the search window is compared to a given patch f_{η_j} (black box) in the image that is to be reconstructed. Both the search window and the patches are defined in all three dimensions. The comparison between the image and prior patches is done by calculating the Minkowski distance $\|\cdot\|_p$ of the order p between the intensity values of the voxels of both patches. Additionally, a mask function $\iota \in \{0,1\}^N$ with

$$\iota_x = \begin{cases} 0 & \text{if } \frac{1}{\kappa_j}\Psi\left(f,\Gamma(h,\gamma),j\right) = 0 \\ 1 & \text{else} \end{cases} \quad (6)$$

is applied to force the regularization to ignore voxels from the prior that hold no relevant information.

To solve the optimization problem described in (3) the l-BFGS-b algorithm is used [5]. In order to apply the l-BFGS-b algorithm to the optimization problem (3), the partial derivatives of the likelihood function and the regularization term are needed. The derivative of the likelihood function according to the voxel f_r is given by

$$\frac{\partial l(f)}{\partial f_r} = \sum_{i=1}^{M} (n_i a_{ir} - n_0 a_{ir} \exp(-\sum_{j=1}^{N} a_{ij} f_j)). \quad (7)$$

For the partial derivative of the regularization term

$$\begin{aligned}
\frac{\delta R(f,\Gamma(h,\gamma))}{\delta f_r} = \frac{1}{2}\Bigg(\sum_{j=0}^{N}\iota_j\left(f_j - \frac{1}{\kappa_j}\Psi\left(f,\Gamma(h,\gamma),j\right)\right)^2\Bigg)^{-\frac{1}{2}} 2\Bigg(\iota_r f_r - \Bigg(\iota_r \frac{1}{\kappa_r}\Psi(f,\Gamma(h,\gamma),j) + \\
\sum_{j=0}^{N}\iota_j f_j\Bigg(\frac{1}{\kappa_j}\sum_{k\in\mathcal{N}_j}\Gamma(h,\gamma)\frac{\delta\left(-\sigma\|f_{\eta_j} - \Gamma(h,\gamma)_{\eta_k}\|_p\right)}{\delta f_r v^2} e^{\frac{-\sigma\|f_{\eta_j}-\Gamma(h,\gamma)_{\eta_k}\|_p}{v^2}} - \\
\frac{\left(\sum_{k\in\mathcal{N}_j}\frac{\delta(-\sigma\|f_{\eta_j}-\Gamma(h,\gamma)_{\eta_k}\|_p)}{\delta f_r v^2} e^{\frac{-\sigma\|f_{\eta_j}-\Gamma(h,\gamma)_{\eta_k}\|_p}{v^2}}\right)}{\kappa_j^2}\Psi\left(f,\Gamma\left(h,\gamma\right),j\right)\Bigg)\Bigg) + \\
\sum_{j=0}^{N}\iota_j\Bigg(\frac{1}{\kappa_j^2}\Psi\left(f,\Gamma(h,\gamma),j\right)\sum_{k\in\mathcal{N}_j}\Gamma\left(h,\gamma\right)\frac{\delta\left(-\sigma\|f_{\eta_j} - \Gamma(h,\gamma)_{\eta_k}\|_p\right)}{\delta f_r v^2} e^{\frac{-\sigma\|f_{\eta_j}-\Gamma(h,\gamma)_{\eta_k}\|_p}{v^2}} - \\
\frac{\left(\sum_{k\in\mathcal{N}_j}\frac{\delta\left(-\sigma\|f_{\eta_j}-\Gamma(h,\gamma)_{\eta_k}\|_p\right)}{\delta f_r v^2} e^{\frac{-\sigma\|f_{\eta_j}-\Gamma(h,\gamma)_{\eta_k}\|_p}{v^2}}\right)}{\kappa_j^3}\Psi(f,\Gamma(h,\gamma),j)\Bigg)\Bigg)
\end{aligned} \quad (8)$$

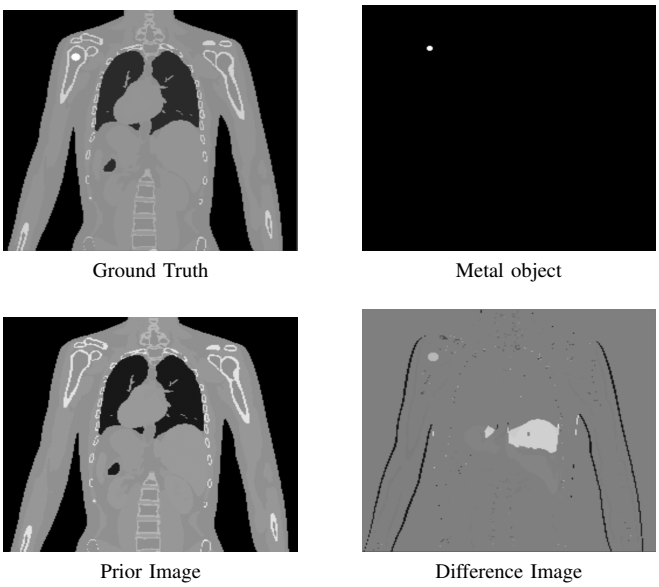


Fig. 2. XCAT Phantom used for image reconstruction in coronal view. Segmentation of the metal object. The used prior image, which was generated by the XCAT software with minor variations of the anatomical parameters to the generated Phantom. The difference image between the ground truth image and the prior image.

one has to consider the occurrence of the voxel f_r in patches $f_{\eta_j}, j \neq r$. The derivative of the p-norm between the patches of the image an the prior can be described by

$$\frac{\delta \| \cdot \|_p}{\delta f_r} = \begin{cases} \frac{1}{p} \left(\sum_{k \in \mathcal{N}_j} (f_{\eta_j} - \Gamma(h,\gamma)_{\eta_k})^p \right)^{\frac{1}{p}-1} \\ \cdot\, p(f_r - \Gamma(h,\gamma)_{k_r})^{p-1} & \text{if } f_r \in f_{\eta_j} \\ 0 & \text{else} \end{cases}$$

the derivative of the Minkowski distance according to the image voxel f_r, if f_r is in the image patch or zero otherwise. To save computation time intermediate results of the distances are temporary saved for all image voxels in the current image patch during the comparison of the patches to avoid unnecessary calculations. As a result, it is possible to calculate all partial derivatives of the regularization with only marginal additional effort.

III. Results

Figure 2 shows a slice in coronal view of the used phantom, which is generated by the XCAT software [6]. To simulate a metal object, an artificial object is added within the right humerus. Additionally, a prior image is generated by modifying anatomical parameters for the bones and the soft tissue. The difference between both images is shown in Figure 2d. In Figure 3 the progress of the reconstruction algorithm is illustrated in different iterations steps. In every fourth iteration step, a new non-local prior and a new derivative of the regularization term for the current image are calculated. The new non-local prior is the result of the comparison between the intermediate result of the reconstruction and the given prior

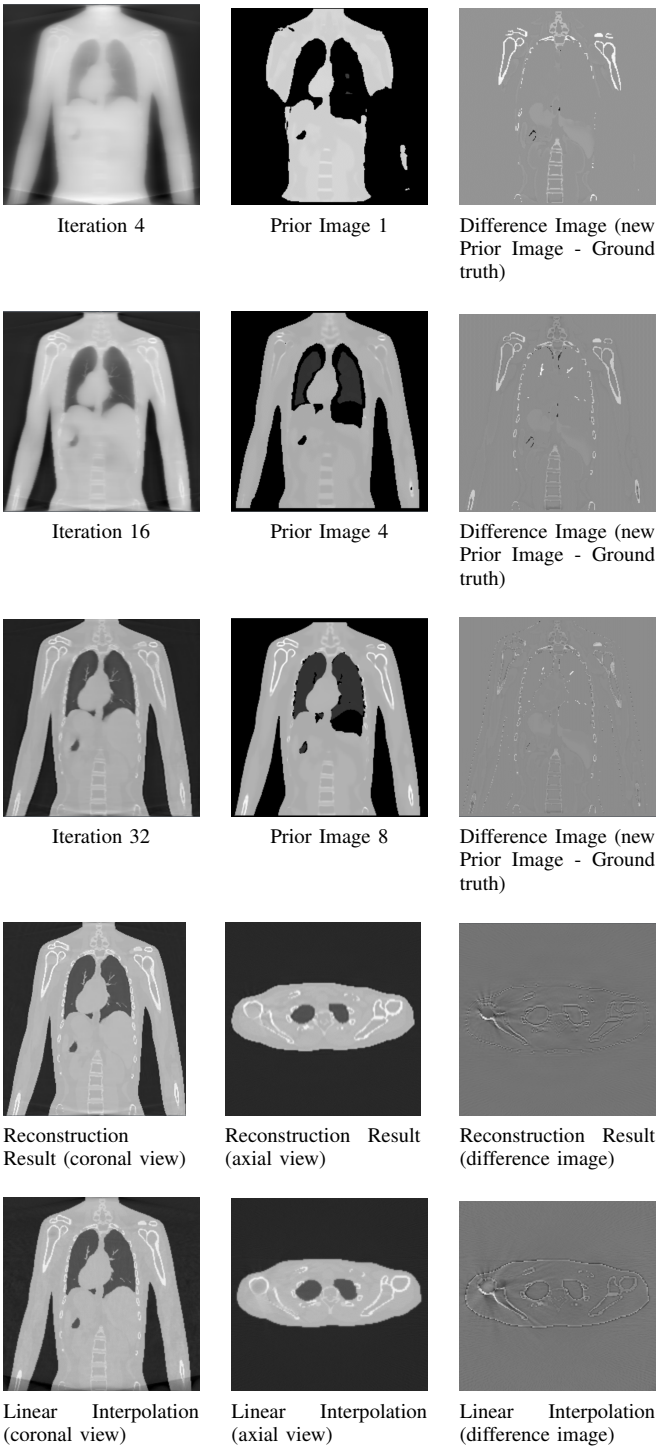


Fig. 3. Reconstruction progress. In every fourth iteration a new prior image is calculated. The deviations from the new calculated prior image to the ground truth is visualized in a difference image. The second last row shows the reconstruction result in coronal and axial view. For a better visualization of the remaining artifacts, in the last image a difference image to the ground truth is calculated in axial view. In the last row the results of the linear interpolation in coronal and axial direction are shown, as well as the difference to the ground truth image in axial view.

image. If a voxel of the new non-local prior image is assigned a zero, the value does not influence the regularization due to the multiplication of the mask function (6) in the regularization term. The same is true for the derivative of the regularization term. The derivative of the regularization term has a direct influence of the new image calculated in the next iteration step. The newly calculated non-local prior images are getting more similar to the intermediate results during the iterations. The last non-local prior shows only small areas of voxels, which are assigned a zero. After approximately 50 iterations the gradient of the objective function is small enough (10^{-9}) to stop the optimization process. In Figure 3 the coronal and axial view of the reconstructed image are shown. The results show that even fine structures are preserved during the reconstruction. In order to better visualize the remaining artifacts, a difference image between the reconstruction result and the ground truth is calculated. In the last row of Figure 3 the proposed algorithm is compared to a linear interpolation approach [7]. Compared to the linear interpolation the final reconstruction result shows significantly less artifacts. This is particularly apparent in the two difference images, which were calculated in relation to the ground truth image. For an objective comparison of the two methods, the sum of squared differences (SSD) to the ground truth image were calculated. With a value of 94 616, the calculated SSD value of the here proposed reconstruction method is significantly lower than the SSD value of the linear interpolation with 207 919. The image values are given in hounsfield units.

IV. Discussion

Many reconstruction approaches for artifact reduction in CT images strongly depend on the similarity of the prior images to the image to be reconstructed. Therefore, the correct choice of a prior image plays an essential role in order to achieve an artifact free reconstruction result. One promising approach is to use clinical progress studies, as these images generally have the same anatomical structures. Nevertheless, due to temporal differences, slight anatomical deviations, caused for instance by organ shifts, can occur. The here proposed non-local regularization takes this issue into account by not only considering direct neighbor pixels. The simulated data show that a reconstruction of the image is possible even with anatomical variations between the ground truth and the prior image. However, more detailed studies are planned to show applicability for clinical data. For this purpose, the procedure presented by Ziemann et. al. can be used as an orientation [8]. Various parameters of the regularization, e.g. the influence of the size of search window and image patches, should be investigated. For example, if the position of various anatomical structures has changed significantly, enlarging the search window can have a positive effect on the reconstruction result. Also, the patch size should not be too large, as this could lead to insufficiently similar structures. Another aspect, which needs to be addressed is that a selection of a larger search window and patch size directly results in a higher computation time. Especially when using large 3D image data, this could be a major challenge. An already parallel implemented calculation of the comparison between the image patches allows a calculation on a GPU, which makes the calculation much faster.
Furthermore, the behavior of noise in clinical image data should be investigated with respect to the calculation of a new non-local prior image. Especially during the comparison of image patches, where a direct weighting between pixel values is performed, noise can strongly affect the calculation of the pixel-specific weighting. In this case, one might want to perform a noise reduction in the image domain before the weighting is calculated.

V. Conclusion

A metal artifact reduction algorithm for 3D cone-beam CT is proposed, which integrates information based on a prior image. A new prior image is calculated using a non-local approach, with a regularization term that penalizes gray-value variations between the non-local prior and the intermediate results of the image reconstruction. A parallel implementation of the calculation of the regularization term and its derivation enables a fast computation on a GPU. For the optimization process, both the regularization term and its derivative are calculated. Compared to the linear interpolation approach the proposed reconstruction shows significantly less artifacts.

References

[1] M. Stille, M. Kleine, J. Haegele, J. Barkhausen, and T. M. Buzug, *Augmented Likelihood Image Reconstruction*, IEEE Transactions on Medical Imaging, vol. 35, no. 1, pp. 150-173, 2015.
[2] G.-H. Chen, J. Tang and S. Leng, *Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets*, 3rd ed. Medical Physics, vol. 35, no. 2, pp. 660–663, 2008.
[3] M. Stille and T. M. Buzug, *Metal artifact reduction by projection replacements and non-local prior image integration*, Current Directions in Biomedical Engineering, vol. 1, pp. 100-103, 2015.
[4] M. Oehler and T. M. Buzug, *Statistical image reconstruction for inconsistent CT projection data*, Methods of Information in Medicine, vol. 46, no. 3, pp. 261-269, 2007.
[5] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, *Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale boundconstrained optimization*, ACM Transactions on Mathematical Software, vol. 23, no. 4, pp. 550-560, 1997.
[6] W. Segars, M. Mahesh, T. Beck, E. Frey, and B. Tsui, *Realistic CT simulation using the 4D XCAT phantom*, Medical Physics, vol. 35, no. 5, pp. 3800–3808, 2008.
[7] W. Kalender and R. Hebel, *Reduction of CT artifacts caused by metallic implants*, Radiology, vol. 164, no. 2, pp. 576-577, 1987.
[8] C. Ziemann, M. Stille, F. Cremers, D. Rades and T.M. Buzug, *The effects of metal artifact reduction on the retrieval of attenuation values*, Journal of Applied Clinical Medical Physics, vol. 18, no. 1, pp. 243-250, 2017.

Metal-Tolerant Noncircular Orbit Design and Implementation on Robotic C-Arm Systems

Grace J. Gang[1], Tom Russ[2], Yiqun Ma[1], Christian Toennes[2], Clifford R. Weiss[1], Jeffrey H. Siewerdsen[1], Lothar R. Schad[2], and J. Webster Stayman[1]

***Abstract*—Metal artifacts are a major confounding factor for image quality in CT, especially in image-guided surgery scenarios where surgical tools and implants frequently occur in the field-of-view. Traditional metal artifact correction methods typically use algorithmic solutions to interpolate over the highly attenuated projection measurements where metal is present but cannot recover the missing information obstructed by the metal. In this work, we treat metal artifacts as a missing data problem and employ noncircular orbits to maximize data completeness in the presence of metal. We first implement a local data completeness metric based on Tuy's condition as the percentage of great circles sampled by a particular orbit and accounted for the presence of metal by discounting any rays that pass through metal. We then compute the metric over many locations and many possible metal locations to reflect data completeness for arbitrary metal placements within a volume of interest. We used this metric to evaluate the effectiveness of sinusoidal orbits of different magnitudes and frequencies in metal artifact reduction. We also evaluated noncircular orbits in two imaging systems for phantoms with different metal objects and metal arrangements. Among a circular, tilted circular, and a sinusoidal orbit of two cycles per rotation, the latter is shown to most effectively remove metal artifacts. The noncircular orbit not only reduce the extent of streaks, but allows better visualization of spatial frequencies that cannot be recovered by metal artifact correction algorithms. These results illustrate the potential of relatively simple noncircular orbits to be robust against metal implants which ordinarily present significant challenges in interventional imaging.**

***Keywords*—Metal artifact, Missing data, Tuy's condition**

I. Introduction

Cone-beam computed tomography (CBCT) acquisitions have traditionally followed a circular source-detector orbit. Despite well-known cone-beam artifacts as a result of insufficient sampling away from the central slice (i.e., the null cone), a circular orbit can be achieved with relatively simple mechanical designs and therefore has remained the dominant scan geometry on CBCT systems. The advent of modern robotic c-arms has enabled greater flexibility in source-detector placements around the patient. This capability has primarily been used for intra-operative fluoroscopy guidance, but it has also provided opportunity to realize a host of noncircular orbits designs for CT acquisition. Many designs have been proposed to combat cone-beam artifacts - e.g., circle-and-line,[1] saddle orbits,[2] circle-plus-arc,[3] etc. These designs often aim to improve or satisfy cone-beam data completeness conditions first proposed by Tuy.[4] Customized noncircular orbits for particular objects or patient anatomy have also been designed for specific image quality objectives, including task-based detectability.[5], [6]

This work is supported by NIH grant R01EB027127.

[1] Johns Hopkins University, Baltimore, MD, USA
[2] Universität Heidelberg, Mannheim, Germany

In this work, we investigate noncircular orbits for metal artifact minimization. Metal artifacts are especially relevant in image-guided surgery scenarios where surgical tools and implants frequently appear in the field-of-view. The traditional solution to metal artifacts is to apply any number of data correction approaches. Projection measurements where x-rays have passed through metal are highly attenuated - often to the point where they can be consider "missing data". Algorithms therefore seek to replace these measurements via interpolation schemes based on neighboring projection data, prior information from previous or other patient scans, or other in-painting approaches.[7], [8] These methods can reduce the apparent severity of the artifacts, but are typically not able to "fill in" the true missing information. That is, these algorithms do not improve the completeness of the data, and anatomical details not sampled in the acquisition will remain missing in the reconstruction. In this work, we aim to address the root cause of metal artifacts by designing orbits that are metal-tolerant by maximizing data completeness in the presence of metal.

II. Theoretical Methods

A. Local Completeness Metric

We adapt Tuy's condition to a local context where data sufficiency is quantified for a single point in the object. A specific "probe" point is completely sampled if all planes passing through it intersects the orbit. We denote points in the object as l with subscript i and discrete source locations on the orbit as S with subscript j. As shown in Fig.1, for each l_i, we define a local spherical coordinate system, $(\theta_{l_i}, \phi_{l_i})$. Each plane passing through l_i intersects a great circles on the unit sphere. We sample an ensemble of great circles (denoted as c with subscript k) at regular intervals of θ_{l_i} and ϕ_{l_i}, which can be equally represented by their unit normals, $\overrightarrow{v_{c_k}}$.

We then calculate whether a particular source location intersects any great circles by finding the angle between $\overrightarrow{v_{c_k}}$ and the vector connecting the source location to the location of interest, $\overrightarrow{v_{l_i}S_j}$. If the angle falls within $90^o \pm \epsilon$, the great circle c_k has been sampled by the orbit. For example, in Fig.1, S_1 intersects c_2 and S_2 intersects c_1 (as well as many others, not shown). We repeat the calculation for all source locations S_j and all great circles c_k. The local completeness metric is defined as the percentage of great circles that have been intersected by the orbit.

In the presence of metal, rays intersecting the metal are highly attenuated - often to the point they provide no useful information. Therefore, we no longer count great circles sampled by such source locations (e.g., S_2 and c_1) towards the completeness metric. Instead, these great circles need to be sampled by an alternative source location (e.g., S_3) that circumvents the metal. In other words, the presence of metal results in a range of $(\theta_{l_i}, \phi_{l_i})$ that represent "missing data" due to high attenuation. When the metal can be approximated as a sphere, this range of $(\theta_{l_i}, \phi_{l_i})$ can be

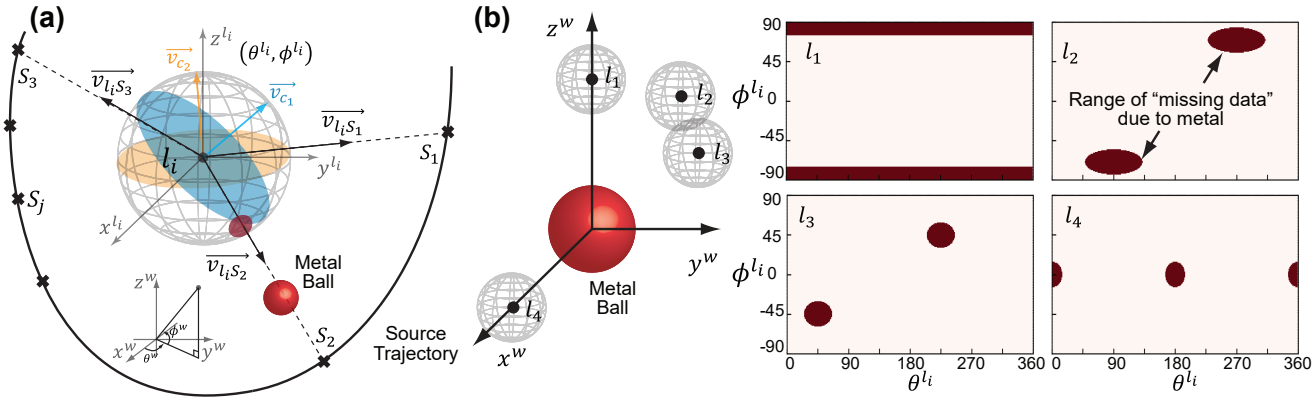


Fig. 1: (a) Local Tuy completeness metric at a specific "probe" point, l_i, as the percentage of great circles intersected by the source trajectory. The presence of metal projects a "missing data" (red) space onto the unit sphere around l_i that no longer contributes to sampling. (b) The "missing data" space in dependent on the size of the metal and the relative location of the metal and the location of interest, illustrated here for four locations around a metal ball.

computed as the equirectangular projection of the metal sphere onto the unit sphere around l_i based on the size and relative location of the metal to l_i. Fig.1(b) illustrates the range of $(\theta_{l_i}, \phi_{l_i})$ to avoid for four example l_i around a metal ball.

B. Noncircular Orbit Design

An orbit that can circumvent metal and achieve a 100% completeness at all probe locations within an object can theoretically eliminate metal (and cone-beam) artifact. We therefore use the completeness metric to identify orbits that are robust against metal artifacts in the presence of metal at arbitrary locations. In this work, we assume the metal object to be a sphere of radius 18 mm, and sampled metal locations within a cylindrical field of view with radius of 16 cm and height of 20 cm. For each metal location, we evaluated the local completeness metric over 50 probe locations distributed over a spherical shell that is 10 mm away from the surface of the metal. We only evaluated locations close to the metal because they suffer from the most severe metal artifacts, i.e., lowest percentage completeness. We can then compute a scalar metric, e.g., the mean, the minimum, etc. to quantify data completeness over the volume at arbitrary metal locations. The metal locations and probe locations around each metal ball used in this work are illustrated in Fig.2.

In previous work,[9] we performed an orbit optimization using the maximum of the minimum local percentages as the objective function. We found that many noncircular orbits can achieve similar levels of completeness. Thus, additional constraints need to be used to select among similar near-complete orbits. In this work, we focus on simple sinusoidal orbits and apply the completeness metric to evaluate the effectiveness of a range of sinusoids of different frequencies, f, and magnitude, ϕ_{max}. In the spherical world coordinates, (θ_w, ϕ_w), the sinusoidal orbits are expressed as:

$$\phi_w = \phi_{max} cos(f\theta_w) \tag{1}$$

where θ_w is the traditional gantry rotation angle, ranging from 0° to 360°, and ϕ_w is the gantry tilt angle [see Fig.1(a) for world coordinate definitions]. The maximum achievable tilt on the imaging system (e.g. due to mechanical constraints, patient positioning, table, etc.) is denoted by ϕ_{max}. Special cases of this definition include $\phi_{max} = 0°$ for a circular orbit and $f = 1$ for a tilted circular orbit.

We computed the completeness metric for sinusoidal orbits of frequencies ranging from 0.25 to 8 at 0.25 increment, and of ϕ_{max} ranging from 0° to 50°.

III. Experimental Methods

A. Imaging phantoms and Imaging Systems

We illustrated the robustness of the noncircular orbits in two phantoms shown in Fig.3(a). Both phantoms are constructed with a cluttered background by mixing plastic balls of different sizes and materials. In this first phantom, on the central axial slice, we inserted three stainless steel balls of 18 mm diameter and surrounded by 3D-printed radial line pairs. These radial line pairs are designed specifically to challenge CT imaging around metal. According to the Fourier slice theorem, the frequencies that will be missed by an in-plane circular orbit are those perpendicular to the metal ball. Thus, line pairs directed to the center of the metal balls would present a significant challenge for interpolation-based correction methods. Two smaller stainless steel balls were also inserted 2.5cm above and below the central axial plane. A second phantom was

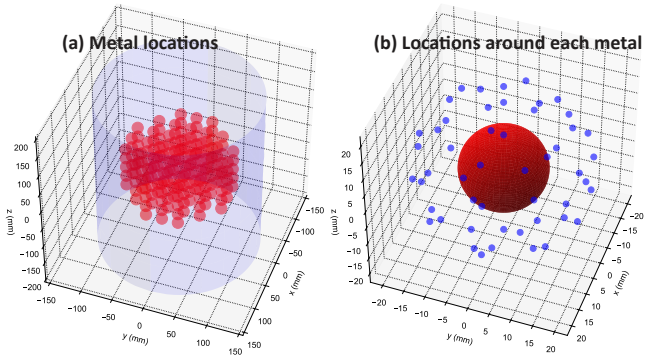


Fig. 2: Sampling locations for (a) metal and (b) probe locations around each metal.

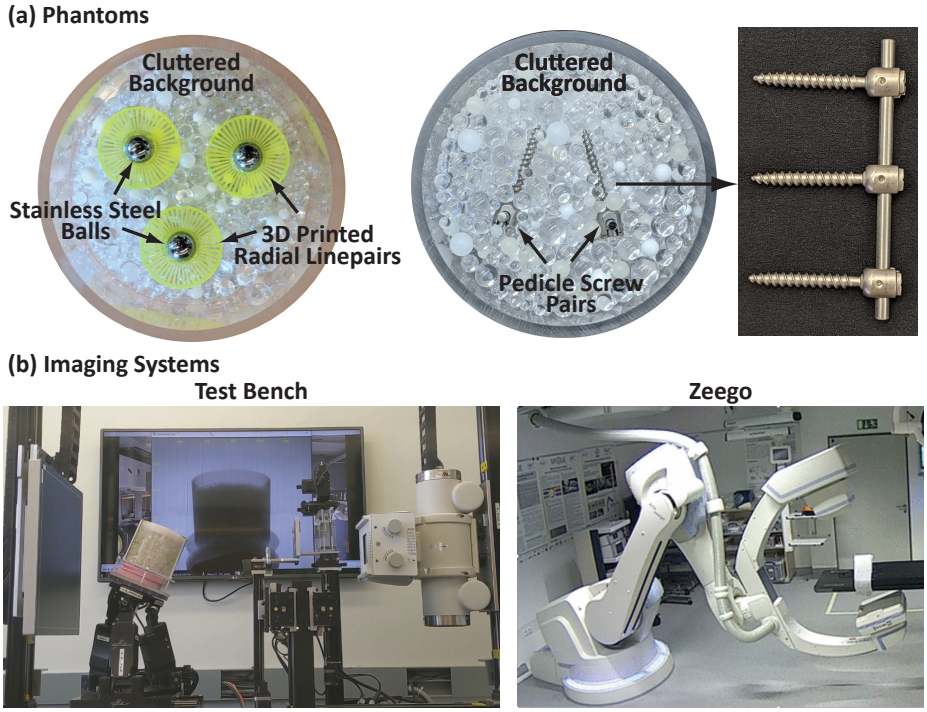


Fig. 3: (a) Imaging phantoms consisting of a cluttered background. We inserted two sets of metal objects: 1) stainless steel balls surrounded by 3D printed radial line pairs, and 2) spine fusion hardware consisting of three pairs of pedicle screws. (b) Noncircular orbits were realized on a CBCT test bench and a Zeego C-arm system.

constructed with the same background clutter but with a spinal fusion hardware complex. Specifically, three pairs of stainless steel pedicle screws were placed in the phantom. Each set of three screws are joined by a stainless steel rod.

Noncircular orbits were realized on two physical systems. First, experiments were conducted on a CBCT test bench comprised of a Varex 4343CB flat-panel detector, Varex Rad-94 x-ray tube and an Alio Industries "hybrid hexapod" system capable of six degree-of-freedom motion, which served as an object stage. This system can emulate a robotic C-arm system with arbitrary orbits. In these studies, source-detector distance was fixed and source, detector, and hexapod motion was used to emulate sinusoidal orbits with tilts up to approximately $24°$. The second system used for investigation was a modified Siemens Zeego robotic C-arm system. This research system was adapted to permit specification of largely arbitrary orbits in a step-and-shoot fashion.[10] This system in shown in Fig.3(b) and also has a maximum tilt limitation of approximately $24°$.

B. Geometry calibration using 2D-3D registration

The noncircular orbit was calibrated using a self-calibration routine described in Ref [11]. Briefly, a reconstruction of the phantom from a circular geometry was used as a known model of the object. Projections of the known model was then registered to each view of the noncircular acquisition to identify the projection matrix. The ensemble of projection matrices were then used to reconstruct data from the noncircular acquisition.

C. Reconstruction

For reconstruction, we chose a model-based algorithm due to its ability to accommodate arbitrary acquisition geometries. We adopted a a penalized weighted least squares objective with a Huber penalty and used a separable footprint projector. No metal artifact correction methods were applied to highlight the extent of metal artifact as a result of orbit alone.

IV. Results

The results of the completeness metric evaluation is summarized in Fig.4 as the minimum over all metal and probe locations. A number of features are important. The percentage completeness is the lowest for circular ($\phi_{max} = 0$) and tilted circular orbit ($f = 1$). In general, the degree of completeness improves with larger ϕ_{max} and increased frequency.

We observe some periodicity where odd frequencies perform worse than even frequencies. This is due to odd frequencies not taking advantage of redundancy in θ to sample different ϕ. That is, the rays (lines of response) that are collected at θ and $\theta + 180^o$ are simlar (for small fan angles) . Even orbits, on the other hand, leverage inherent redundancy of a $360°$ orbit to provide rays sampled at different ϕ and are therefore more robust against metal artifacts. At high frequencies, the space is densely sampled and therefore the periodicity is not as pronounced.

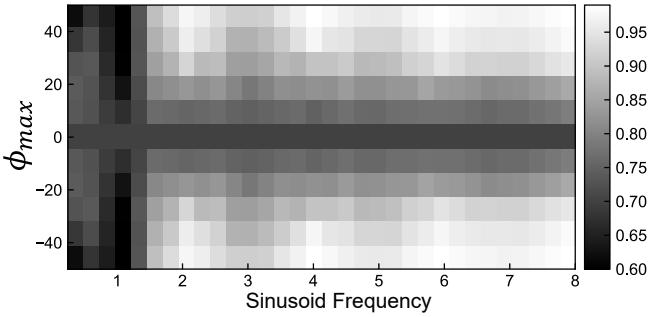


Fig. 4: The minimum percentage complete over all metal and probe locations for sinusoidal orbits of different frequencies and magnitudes.

Reconstructions for the two phantoms in Fig.3 are shown in Fig.5 for a circular, tilted circular, and a sinusoid acquisition at $f = 2$. Reconstructions following circular orbits have severe streaking as a result of multiple in-plane metal objects. Moreover, the radial line pairs cannot be faithfully reproduced in-plane since those frequencies were never sampled. The tilted circular orbit improves in-plane visualization but displaces the metal artifacts to out-of-plane, evident in the saggital slices. The $f = 2$ sinusoid exhibits the least metal artifacts both in-plane, and out-of-plane. The reduction in metal artifact is not only reflected in the reduced streak artifacts, but also in the ability to see sphere details between metals which would not be visible had a metal artifact reduction algorithm been used.

V. Discussions and Conclusions

This works presents noncircular orbit design and implementation for metal artifact minimization. We used a local Tuy completeness metric to quantify the extent of sampling in the presence of metal and investigated the robustness of sinusoidal orbits of different frequencies and magnitudes. The orbits were realized on an experimental CBCT test bench and a Zeego system for two phantoms with metal balls and pedicle screws. We observed that a sinusoidal orbit of frequency equals 2 effectively removes metal artifacts by improving sampling and reducing the missing data effects as a result of metal implants.

Ongoing work is seeking to extend orbital designs for patients with hip implants and to implement such metal-tolerant orbits on other clinical systems including mobile C-arms. Similarly, quantitative comparisons with traditional metal artifact correction approaches are underway. Future work will include new orbital design that seek solutions within the class of complete orbits using additional image quality measures. For example, image quality is also driven by patient anatomy (habitus and bone) that can be used to select from metal-tolerant complete orbits.

This work has illustrated the basic feasibility for using noncircular orbits to significantly reduce metal artifacts by addressing the source of the problem - missing data. Complete orbits as simple as sinusoids can provide sufficiently redundant data to eliminate metal artifacts - providing a potential clinical solution for interventional CT to a longstanding problem for surgical tools and metal implants.

References

[1] G. L. Zeng and G. T. Gullberg, *Physics in Medicine & Biology*, vol. 37, no. 3, p. 563, 1992.

[2] J. P. Shah *et al.*, in *Medical Imaging 2014: Physics of Medical Imaging*, vol. 9033. International Society for Optics and Photonics, 2014, p. 90332L.

[3] X. Wang and R. Ning, *IEEE transactions on medical imaging*, vol. 18, no. 9, pp. 815–824, 1999.

[4] H. K. Tuy, *SIAM Journal on Applied Mathematics*, vol. 43, no. 3, pp. 546–552, 1983.

[5] J. W. Stayman *et al.*, *Journal of Medical Imaging*, vol. 6, no. 2, p. 025002, 2019.

[6] S. Capostagno *et al.*, *Journal of Medical Imaging*, vol. 6, no. 2, p. 025004, 2019.

[7] W. A. Kalender *et al.*, *Radiology*, vol. 164, no. 2, pp. 576–577, 1987.

[8] F. E. Boas and D. Fleischmann, *Imaging in Medicine*, vol. 4, no. 2, pp. 229–240, 2012.

[9] G. J. Gang *et al.*, in *SPIE Medical Imaging*, Houston, TX, 2020.

[10] K. Chung *et al.*, *International journal of computer assisted radiology and surgery*, vol. 13, no. 10, pp. 1481–1495, 2018.

[11] S. Ouadah *et al.*, *Physics in Medicine and Biology*, vol. 61, no. 7, pp. 2613–2632, 2016.

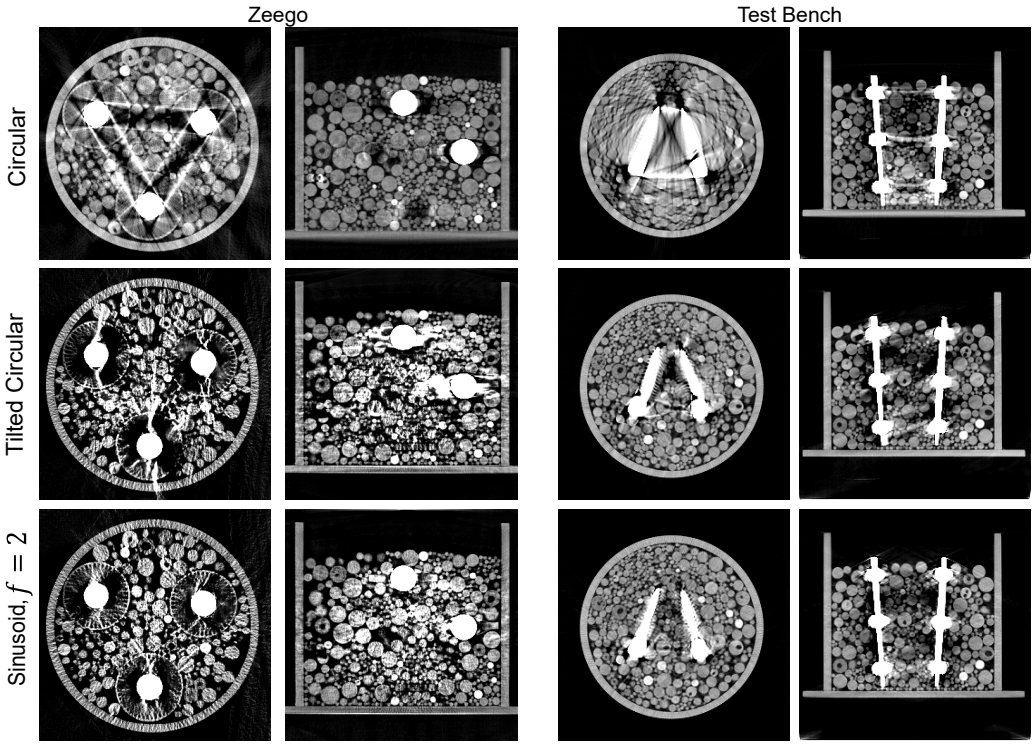


Fig. 5: Reconstructions for the two phantom configurations using a circular orbit, a tilted circular orbit, and a sinusoidal orbit of $f = 2$

Single-pass metal artifact correction using a dual layer flat panel detector

Linxi Shi, N. Robert Bennett, Amy Shiroma, Mingshan Sun, Jin Zhang, Richard E. Colbeth, Josh Star-Lack, Minghui Lu, Adam S. Wang[1]

***Abstract*—Metal artifact remains a challenge in cone-beam CT images. Many image domain-based segmentation methods have been proposed for metal artifact reduction. Most of these methods require a two-pass reconstruction and work fairly well, but are limited when the metal is outside the scan field-of-view (FOV) or when the metal is moving during the scan. In the former, even reconstructing with a larger FOV does not guarantee a good estimate of the metal location in the projections; and in the latter, the metal location in each projection is difficult to identify due to motion. Furthermore, two-pass methods increase the total reconstruction time. A dual layer detector provides dual energy images with perfect temporal and spatial registration, which can aid in the identification of metal in the projection domain. Previously, we used a simple phantom with metal wires (copper) and a needle (steel) to demonstrate the feasibility of this idea. Preliminary results showed an enhanced ability to identify metal regions. In this work, we extend the algorithm to adapt it to more realistic problems. Two objects are evaluated: 1) A liver steering phantom with embedded metal and a biopsy needle crossing the phantom boundary, and 2) a cadaver head with dental fillings and an additional metal tag attached to its surface. Material-specific projections are used to identify metal regions in the projection domain, which are then corrected by interpolation using surrounding pixels. Promising results are seen for both two cases, with substantial metal artifact reduction obtained in reconstructed images.**

***Index Terms*— Cone-beam CT, metal artifact, dual energy, dual layer, material decomposition**

I. Introduction

Cone-beam CT has grown rapidly in clinical use, including for image-guided radiation therapy, image-guided interventions, and for diagnostic applications requiring dedicated head, dental, and extremity scanners. For all these use cases, metal objects, such as interventional devices, implants and dental fillings are often present. The presence of metal severely degrades image quality due to strong beam hardening effects or photon starvation. Metal artifact reduction (MAR) methods have been a popular research topic, many of which apply a two-pass reconstruction method to first segment the metal from a first-pass reconstruction which is then forward-projected, to identify the corresponding region in projections[1]. These methods are effective in most conditions, but are insufficient for some challenging cases, including when the metal is outside the scan FOV, at an edge where the metal (needle) crosses in and out of the FOV, or when the metal is moving[2]. These challenges could be directly solved if metal is detectable in the projection domain, which is difficult due to complicated background signals from tissue and bone, and complications arising from large variations in metal shape and size that are often found.

Dual energy (DE) imaging can generate material-specific images in the projection domain, thus potentially identifying metal. In addition to addressing the aforementioned challenges, it allows for a single-pass reconstruction to be performed thus substantially improving computational efficiency compared to a multi-pass method. There are several methods for obtaining DE data, including fast kV switching that alternates between low and high tube voltages, slow kV switching that acquires sequential projections at different energies, dual source systems, and dual layer detectors that simultaneously capture DE projections in a single scan. The first three methods inevitably introduce spatial and temporal mismatches. A dual layer detector performs dual energy imaging with perfect spatial and temporal registration and therefore would be ideal for this task. We previously reported a simple phantom study of automatic projection-domain metal correction using a dual layer detector[3]. In this work, we extend the method to address more realistic problems, including performing a needle biopsy of a liver phantom, and imaging a cadaver head in the presence of dental fillings and other metal.

II. Method

A. Prototype dual-layer flat panel detector

A prototype dual layer detector with an active area of 432×432 mm^2 was used for this work[4],[5]. The detector comprises two amorphous-silicon (a-Si) panels, each coupled

[1] Research reported in this publication was supported in part by Varex Imaging and by NIH T32CA009695. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. (Corresponding author: Linxi Shi)

Linxi Shi is with the Department of Radiology, Stanford University, Stanford, CA 94305, USA (email: linxit@stanford.edu)

N. Robert Bennett is with the Department of Radiology, Stanford University, Stanford, CA 94305, USA

Amy Shiroma is with Varex Imaging Corporation, San Jose, CA 95134, USA.

Mingshan Sun is with Varex Imaging Corporation, San Jose, CA 95134, USA.

Jin Zhang is with Varex Imaging Corporation, San Jose, CA 95134, USA.

Richard E. Colbeth is with Varex Imaging Corporation, San Jose, CA 95134, USA.

Josh Star-Lack is with Varex Imaging Corporation, San Jose, CA 95134, USA.

Minghui Lu is with Varex Imaging Corporation, San Jose, CA 95134, USA.

Adam S. Wang is with the Department of Radiology, Stanford University, Stanford, CA 94305, USA

with a thallium-doped cesium iodide (CsI:Tl) scintillator. The top layer has a 200 μm scintillator that preferentially absorbs lower energy x-rays, while the bottom layer has a 550 μm scintillator that preferentially absorbs higher energy x-rays. A 1 mm copper filter was placed between the two detector layers to further increase spectral separation (difference in average energy between the layers). The resulting spectral separation for a 120 kV spectrum with 2 mm Al filtration was 24.3 keV. Without the 1 mm Cu filter, the spectral separation would only be 9.5 keV. The dual layer images were spatially registered using an affine transformation to account for slight differences in magnification, rotation, and translation between the layers. The transformation was obtained from a one-time calibration and then used to register the layers in all future scans.

B. Phantom and cadaver head studies

Metal correction was performed for two scenarios. The first was a liver phantom that was used for testing a needle steering device. The phantom included metal wire inside the homogeneous liver and a steel needle passing across the liver surface. The latter might be difficult to identify from the first-pass reconstruction in conventional MAR if part of it is outside the reconstructed FOV. The second scanned object was a cadaver head with two dental implants. A steel tag was attached to its ear that represents out-of-FOV metal.

C. Correction methods

For both objects, CBCT scans were obtained with 500 projections at 120 kV (2mm Al filtration) and 15 mA, 18 ms pulses on a tabletop system. The liver phantom was intentionally offset laterally to introduce additional truncation. We first performed a scatter correction for projections obtained from both layers using the CBCT Software Toolkit (CST 2.0, Varex Imaging). The software estimates the scatter distribution using a fast adaptive scatter kernel superposition (fASKS)[6]. The scatter-corrected projections were then log transformed to line integrals, generating a low energy (top layer) image l_L and a high energy (bottom layer) image l_H. The bone projections, l_b, were generated by weighted subtraction of l_L and l_H, expressed as:

$$l_b = -l_H + \frac{\mu_s^H}{\mu_s^L} l_L , \qquad (1)$$

where μ_s^H, μ_s^L are the effective attenuation coefficients of soft tissue for the high energy and low energy spectra, respectively. The metal was first selected by setting a threshold on l_b. A local entropy filtration of l_b was applied to remove background structures and increased noise due to the weighted subtraction. In particular, the metal projections, Ω_m, were identified as:

$$\Omega_m \in \{(i,j) | l_b(i,j) > 1.4, l_{entro}(i,j) > 5\} \qquad (2)$$

where (i,j) is the pixel index on the detector and l_{entro} is the local entropy image obtained by scanning a 5×5 window across l_b and calculating the entropy within the window for each pixel index.

For each projection view, the Ω_m was followed by a morphological dilation operation to include any missing metal along the edges. The metal regions were then replaced by inwardly interpolating the pixel values on the outer boundary of the regions. The inpainting was used to suppress any metal. If the metal itself is of interest, the metal regions could be reconstructed separately. The typical processing time for a CBCT acquisition of 500 projections was about 45 sec on a standard PC, without run-time optimization. The corrected projections were then reconstructed using FDK reconstruction (CST 2.0, Varex Imaging). Because the metal correction algorithm is strictly projection-based, in principle all processing and reconstruction steps can occur in real time as each projection is acquired, thereby providing the reconstructed volume soon after the last projection is acquired.

III. Results

A. Liver phantom

Figure 1 shows sinograms of the liver phantom at three different detector rows. In the uncorrected images, it is seen that a metal wire is inside the phantom and a biopsy needle goes in and out of the phantom. In our previous work, we found that simply setting a threshold on single-energy projections was difficult to detect the metal towards the edge and outside of the phantom due to its lower total attenuation. In contrast, by applying the proposed dual energy method, both metal objects are fully detected as shown in Fig. 1. Note that the metal correction was performed in each 2D projection, and the sinograms of detector rows shown in Fig. 1 are only for demonstration of successful metal detection.

CBCT reconstructions of the liver phantom were performed, with and without our proposed metal correction. Figure 2 shows the reconstructed images at three axial planes, including where the needle is outside, at the edge of, and inside the phantom. The uncorrected images show strong streaks and beam hardening artifacts caused by metal inside and outside the phantom. The proposed method successfully removed these artifacts. The comparison of soft tissue uniformity is summarized in Table 1, where the average standard deviation of soft tissue (σ_s) for uncorrected and corrected images are 0.37 and 0.21 cm^{-1}, respectively. The higher σ_s in uncorrected images are mainly due to the metal artifacts, which are largely removed by the proposed correction method. The remaining σ_s in the corrected images is limited by residual soft tissue beam hardening and image noise.

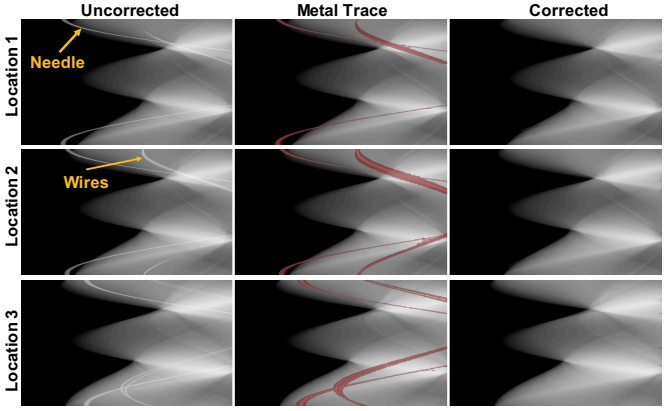


Fig. 1. Left column: sinograms of the liver phantom recorded at 3 different detector rows. Middle column: uncorrected sinograms showing metal wire inside the phantom and a biopsy needle crossing over the phantom surface. The red overlays indicate metal traces detected by the proposed method. Right

column: corrected sinograms obtained by interpolating the metal traces using surrounding values.

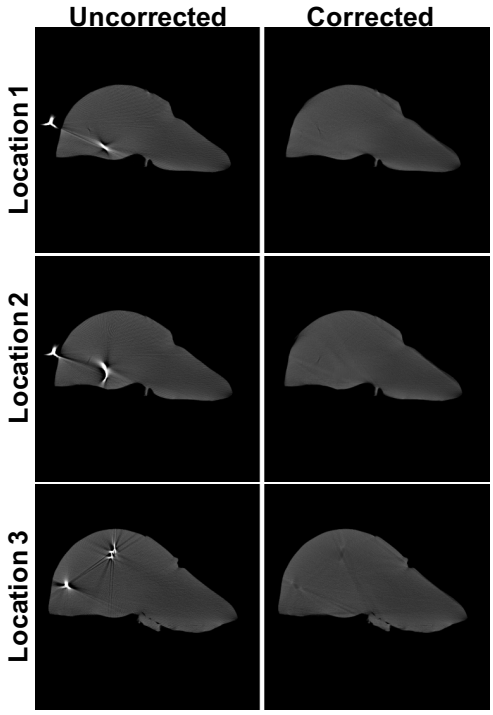


Fig. 2. CBCT reconstruction of the liver phantom at three different axial locations. Left column: uncorrected images. Right column: corrected images. Display window: [0.1 0.6] cm^{-1}.

Table 1. Summary of image uniformity

σ_s (cm^{-1})	Uncorrected	Corrected
Location 1	0.31	0.22
Location 2	0.39	0.21
Location 3	0.41	0.21

B. *Cadaver head*

Figure 3 shows the projections of a cadaver head obtained at four different angles. Two dental implants inside the head and one ear tag outside the head are seen in these acquisition angles. The proposed method is able to detect this metal correctly.

Three axial slices are presented in Fig. 4. Location 1 shows a slice through the lower jaw with a filling on a back tooth. The metal artifact is manifest as streaks and shading intruding into the adjacent soft tissue. The proposed method removes these artifacts and restores the soft tissue uniformity. Location 2 is a slice containing two fillings on both sides of the jaw. With the narrower display window, soft tissue between the fillings is hardly seen, but becomes visible after correction. The metal artifact is largely removed. Some blurring is observed for the tooth next to the left filling. This is mainly due to inpainting of the metal region with linear interpolation. In projections where the metal overlaps the tooth, linear interpolation leads to incorrect line integrals of the tooth, e.g., using surrounding soft tissue to interpolate over dense enamel. Location 3 shows a slice containing metal outside the phantom, which could be truncated if reconstructed with a smaller FOV or scanned with a smaller detector. This could be challenging for image-based correction methods, while the proposed dual energy method is able to correct the metal in the projection domain. The zoomed-in images show the removal of windmill artifacts resulting from out-of-field metal.

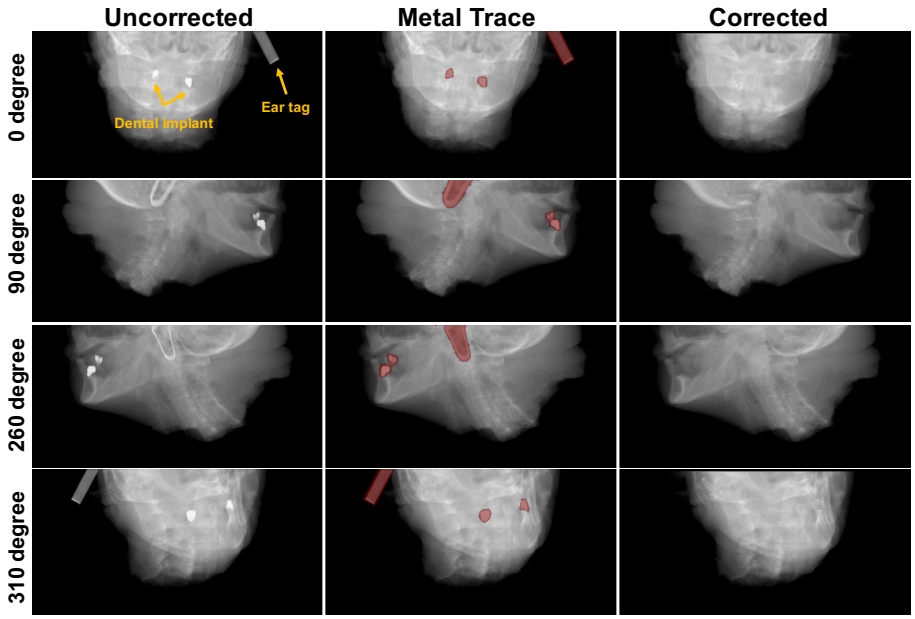


Fig. 3. Left column: log-normalized projections at different acquisition angles. Middle column: uncorrected projections with detected dental implants and ear tag overlaid in red. Right column: corrected projections after local interpolation. Display window: [1 6] unitless.

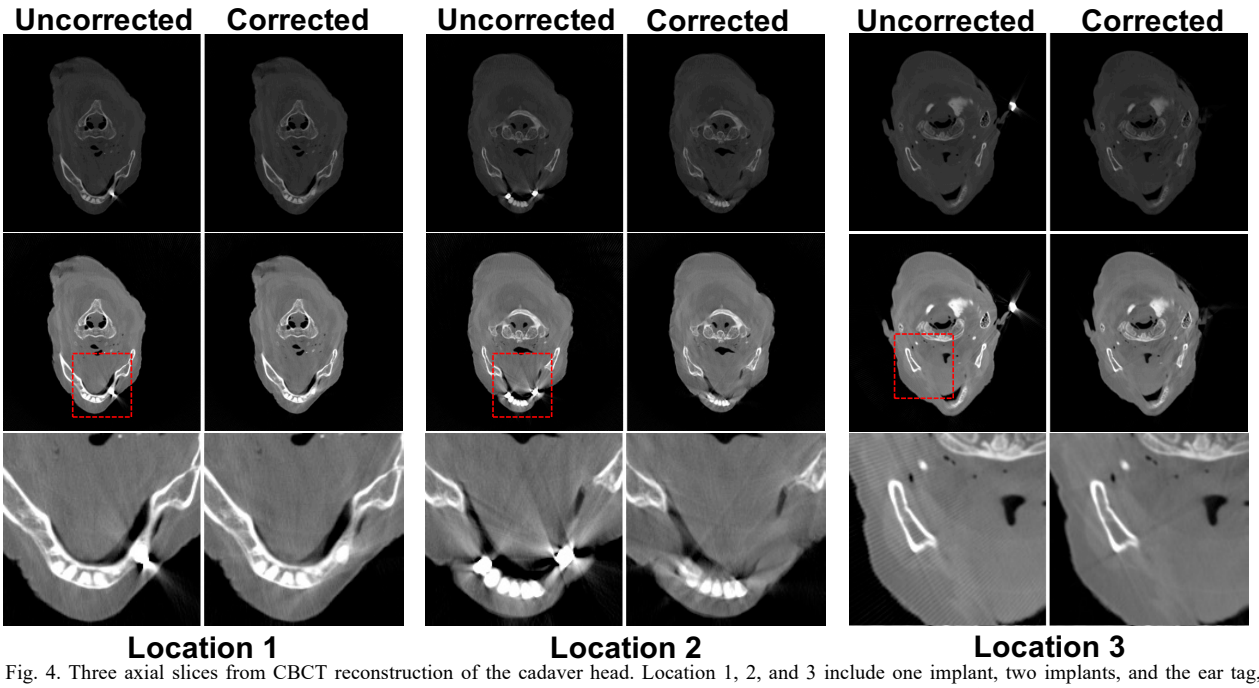


Fig. 4. Three axial slices from CBCT reconstruction of the cadaver head. Location 1, 2, and 3 include one implant, two implants, and the ear tag, respectively. The display window for the top row is [0.1 1] cm^{-1}, and for the middle and bottom rows: [0.1 0.6] cm^{-1}. The bottom row shows zoomed-in views of the red bounding box indicated in the middle row.

IV. Discussion and conclusion

We have proposed a novel metal artifact correction method which is robust to FOV truncation and motion during the acquisition. The method takes advantage of the dual energy capability of a dual layer detector to generate material-specific projections for metal detection in the projection domain. The merit of this approach is that it only requires a single reconstruction, substantially reducing processing time compared to image-based segmentation methods. Our current approach removes metal from the projections and reconstructed image. If the metal is of interest, the approach can be extended to reconstruct the metal separately since it has already been segmented in the projections.

Our method uses linear interpolation for metal region inpainting. A limitation of such a method is that it causes additional streaking artifacts due to mismatch between the surrounding heterogeneous structures used for interpolation and the true anatomy. An example of this was seen in Location 2 in Fig. 2. This will be less problematic for anatomical sites with homogenous backgrounds. The classic image-domain segmentation method, NMAR[7], substantially reduces such effects, but requires extra computation time, including an extra forward-projection and reconstruction. In our ongoing work, we will investigate more advanced inpainting methods beyond linear interpolation to solve this problem.

In this work, DE information was only used to select the metal region presented in the single energy projections for correction. As a result, the DQE was inevitably reduced. In our future work, we will improve DQE by incorporating DE data in the metal corrected images.

References

1. Mouton A, Megherbi N, Van Slambrouck K, Nuyts J, Breckon TP. An experimental survey of metal artefact reduction in computed tomography. J. Xray. Sci. Technol.2013;21(2):193–226.
2. Schäfer D, Haase C, van der Sterren W, Grass M. Reduction of metal artefacts in CBCT caused by needles crossing the FOV border. In: 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine. 2019. page 110720H.
3. Shi L, Bennett NR, Star-Lack J, Lu M, Wang A. Projection-domain metal artifact correction using a dual layer detector. In: Proc. of SPIE Medical Imaging: Physics of Medical Imaging. 2020. page 1131277.
4. Shi L, Bennett NR, Shapiro E, Colbeth RE, Star-Lack J, Lu M, et al. Comparative study of dual energy cone-beam CT using a dual-layer detector and kVp switching for material decomposition. In: Proc. of SPIE Medical Imaging: Physics of Medical Imaging. 2020. page 1131272.
5. Lu M, Wang A, Shapiro E, Shiroma A, Zhang J, Steiger J, et al. Dual energy imaging with a dual-layer flat panel detector. In: Proc. of SPIE Medical Imaging: Physics of Medical Imaging. 2019. page 1094815.
6. Sun M, Star-Lack JM. Improved scatter correction using adaptive scatter kernel superposition. Phys Med Biol 2010;55(22):6695–720.
7. Meyer E, Raupach R, Lell M, Schmidt B, Kachelrieß M. Normalized metal artifact reduction (NMAR) in computed tomography. Med Phys 2010;37(10):5482–93.

Deep learning low dose CT image denoising using synthetic training data

Masakazu Matsuura, Jian Zhou, Hiroki Taguchi, Naruomi Akino and Zhou Yu

***Abstract*—Deep convolutional neural networks (DCNN) have achieved a great success in the accomplishment of low dose CT image denoising. However, unlike the denoising of natural images, collecting specific training data for CT can be challenging due to the lack of high quality training target data and efficient low dose/noisy input simulations. In this paper we propose a novel training method using synthetic CT images. We use independent image and noise training pairs where the noise is estimated directly from single real low dose realizations without using a low dose simulator. The performance of our proposed method has been verified with various clinical and phantom images. The results show that the proposed method is efficient with performance comparable to existing methods.**

I. Introduction

Low dose scans are often preferred in many clinical X-ray CT applications in order to reduce the potential radiation risk. However, the quality of image reconstruction can decrease greatly when the radiation dose becomes low. Recently, DCNN has been successfully developed to denoise low dose CT image. There are mainly two popular training paradigms for the denoising task: the noise-to-clean (N2C) training (see [1][2] for example) and the noise-to-noise (N2N) training [3][4]. In N2C, a low dose CT image is taken as the training input while a matched high dose CT image is often used as the training target. The N2N method does not involve any high dose images, instead two independent low dose images of the same scan are necessary to create the training input-target pairs.

Both methods require obtaining low dose CT images as training input. Directly simulating low dose CT images in image domain can be difficult because the specific reconstruction noise can be signal dependent. Most DCNN denoisers are trained with images reconstructed from simulated low dose scans, thus their efficiency relies heavily on how realistic the low dose simulator can generate independent noisy realizations.

In this paper we attempt to understand what a CT denoising DCNN actually learns and what is important to denoising performance in terms of training datasets. Our method is inspired by the *residual learning* in which an efficient network training actually learns to tell the difference between input (noise + signal) and target (signal). Therefore, we postulate that it is perhaps more critical for a neural network to learn the difference or *the noise feature* rather than the underlying statistical dependency between noise and signal. This motivates us to consider *synthetic* training pairs where the training input is simply formed by superimposing an independent noise image onto the training target image. To mimic more realistic CT images, we estimate the image noise directly from a single real low dose scan without using a low dose simulator. The generated noise samples are guaranteed to be independent. In the following we will present more details about the proposed method including the training procedure, and demonstrate its efficiency through experiments of clinical CT image denoising.

Masakazu Matsuura, Hiroki Taguchi and Naruomi Akino are with Canon Medical Systems Corporation, 1385, Shimoishigami, Otawara-Shi, Tochigi, 324-8550, Japan. Jian Zhou and Zhou Yu are with Canon Medical Research USA, Inc. 706 N. Deerpath Dr. Vernon Hills, IL 60061 USA.
E-mail: masakazu1.matsuura@medical.canon

II. Methods

A. Network training with synthetic CT images

We consider to denoise a low dose FBP reconstruction y which may be expressed as

$$y = \bar{x} + n \tag{1}$$

with $\bar{x}$ being the unknown clean image and n the associated (zero-mean) noise. The goal here is to recover $\bar{x}$ from y. For the standard N2C, we usually collect a set of training pairs $\{(y_i, \bar{x}_i)\}$ and train a convolutional neural network f_θ with parameters defined in θ by minimizing the empirical risk with respect to θ:

$$\sum_i L(f_\theta(y_i),\ \bar{x}_i) \triangleq \sum_i L(f_\theta(\bar{x}_i + n_i),\ x_i) \tag{2}$$

where L denotes the loss function. We can see that the training process attempts to teach the neural network how to separate the noise and signal. When the noise and the signal are independent, we can use the following alternative for training

$$\sum_i L(f_\theta(\bar{x}_i + \epsilon_i),\ \bar{x}_i) \tag{3}$$

where ϵ_i denotes the synthetic noise that is independent of $\bar{x}_i$ (e.g., the noise extracted from different CT image rather than y_i). We refer to $\bar{x}_i + \epsilon_i$ as the synthetic CT image and $\bar{x}_i$ as the background image. This implies that under the assumption of signal-independent noise the use of synthetic image does not change the network learns as in (2). We note that the real noise of reconstructed CT image is not fully signal independent. The typical noise of FBP reconstruction can be intensity dependent. Therefore, in general the network trained with synthetic images is not fully equivalent to the one obtained in (2).

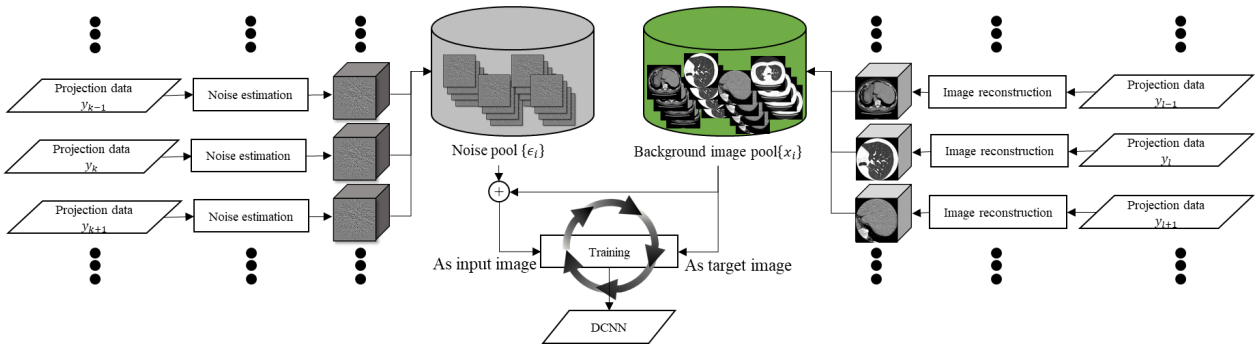


Fig. 1. The proposed denoising DCNN training using synthetic CT images.

TABLE I
THE PROPOSED TRAINING PROCEDURE

Create a background image pool $\{x_i\}$ and a synthetic noise pool $\{\epsilon_i\}$;
Choose a learning rate α, an optimizer and a batch size M.

for iteration $k \leftarrow 1, 2, \cdots$ **do**
  Sample a minibatch $\{x_m\}_{m=1}^{M}$ from the background image pool
  Sample a minibatch $\{\epsilon_m\}_{m=1}^{M}$ from the synthetic noise pool
  Calculate the gradient $g_\theta = \sum_m \nabla_\theta L(f_\theta(x_m + \epsilon_m), \epsilon_m)$
  $\theta \leftarrow \theta + \alpha \cdot \text{Optimizer}(\theta, \nabla g_\theta)$
end for

B. Real image noise estimation

In order for the network to denoise real CT images, the synthetic noise cannot be arbitrarily chosen. We extract the noise from real FBP reconstruction. The method we use is based on the following fact: if there are two independent realizations of a same scan each of which is reconstructed using the same algorithm, the corresponding reconstructions are denoted by $\hat{y}^{(1)}$ and $\hat{y}^{(2)}$, then the calculation below

$$\hat{\epsilon} = \gamma \left(\hat{y}^{(1)} - \hat{y}^{(2)} \right). \tag{4}$$

produces a valid image noise estimation up to a scaling factor γ ($\gamma > 0$). With this in mind, we only need to produce two independent scans. The method we propose is similar to the one used in [4]. More specifically, we split a single scan into two half-data scans that mimic the two independent noisy realizations. This can be done by collecting only the odd projections as one realization and the even projections as the other. The difference of the two reconstructed half-data images can be used as the noise estimation of a full data image. This method is generally effective since the FBP is a linear operator, but it might not be appropriate when the reconstruction is nonlinear (e.g., MBIR) or involves data processing in the angular direction (e.g., the Hilbert transform).

C. Training implementation

The proposed training framework as illustrated in Fig. 1. We maintain two training data pools: the background image pool and the synthetic noise pool. In practice we use small image patches to form the required data pool. To increase data diversity, one can also randomly draw the minibatch samples from each data pool during each backpropagation. The detailed training procedure can be found in Table I.

III. EXPERIMENTAL STUDY

A. Training materials

Our DCNN denoiser adopted the structure of U-net [5] with additional summation between input and output added to form the residual learning. To generate a noise pool, four clinical data scans at doses varied from 5mAs to 50mAs were obtained. All of these scans cover different anatomies including lung, abdomen, pelvis and etc. The proposed method with $\gamma = 1/2$ was applied to obtain samples of image noise.

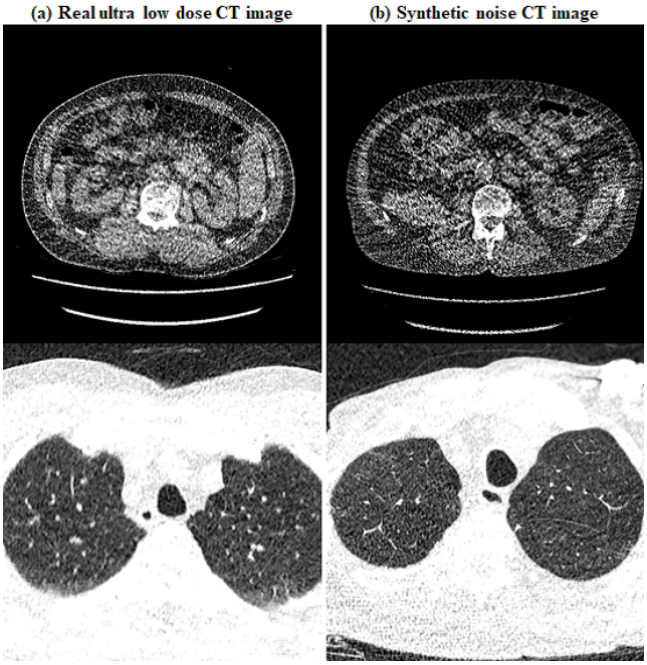


Fig. 2. A comparison between real ultra low dose CT image and synthetic noise CT image. (a) Real low dose CT image was scanned at low dose (Helical, 120kVp, 5mAs, CTDIvol:0.3mSv.). (b) Image generated using synthetic noise. Top row: 400mm rFOV and soft tissue display (WL/WW=40/400). Bottom row: Zoomed lung area from rFOV: 400mm and lung display (WL/WW=-600/1200)

Another independent dataset, which includes three relatively high dose patient scans, was used to generate the background image pool. These data were reconstructed using MBIR with manually tuned regularization parameters in order to obtain relatively high quality target images. We reconstructed images with various sizes of reconstruction field of view (rFOV) from 160 mm to 400 mm. This is similar to the data augmentation technique and helps enhance feature diversity. We extracted 128×128 patches from reconstructed image volumes as a patch to form the required training data pool. The total number of training pairs in the pool is roughly about eleven thousand five hundred. Fig. 2 shows a comparison of a real ultra low dose CT image and a synthetic noise CT image. As can be seen, the synesthetic noise CT image has similar visual appearance as the real ultra low dose image.

We compared our method with the standard N2C method. To prepare training data for N2C, we used the low dose scan simulator proposed in [6] to simulate the dose reduced scan from the existing high dose scans. For each high dose data, three dose reduction levels were chosen: 12%, 5% and 3% (with respect the full dose scan) such that the simulated low dose data roughly matches the one used to create the noise pool. Similarly, these data were reconstructed using FBP with various rFOV sizes. The same MBIR images were used as training targets. Similar amount of image patches of size 128×128 were extracted and used for training. The same U-net was also used. In both trainings, we used the ADAM optimizer [7] with learning rate equal to 10^{-4} where the loss function was the mean absolute error. We set the batch size equal to eight and ran a total number of 150 epochs for effective convergence.

B. Evaluation using simulated data

First we analyze network performance using simulated low dose data which is not included in our training dataset. The result is shown in Fig. 3. Fig. 3 (a) shows the high dose MBIR image as ground truth. Fig. 3 (b) shows FBP low dose image which is reconstructed using simulated low dose data by FBP with a lung kernel. The data is simulated 5% dose from full dose data used in Fig. 3 (a). Fig. 3 (c) and (d) shows the image denoised by the standard N2C and the proposed method with Fig. 3 (b) as input. We evaluate the image quality by calculating the mean square error (MSE) and the structural similarity metric (SSIM). They are calculated between Fig. 3 (a) ground truth and the denoised image Fig. 3 (c) and (d). And they are also calculated between Fig. 3 (a) ground truth and Fig. 3 (b) FBP low dose image as reference. As can be seen, the denoised image using the proposed network is very comparable to the image denoised by the standard N2C network. The noise has been well removed, and the two denoised images have similar MSE and SSIM values, indicting that the image quality difference is very subtle.

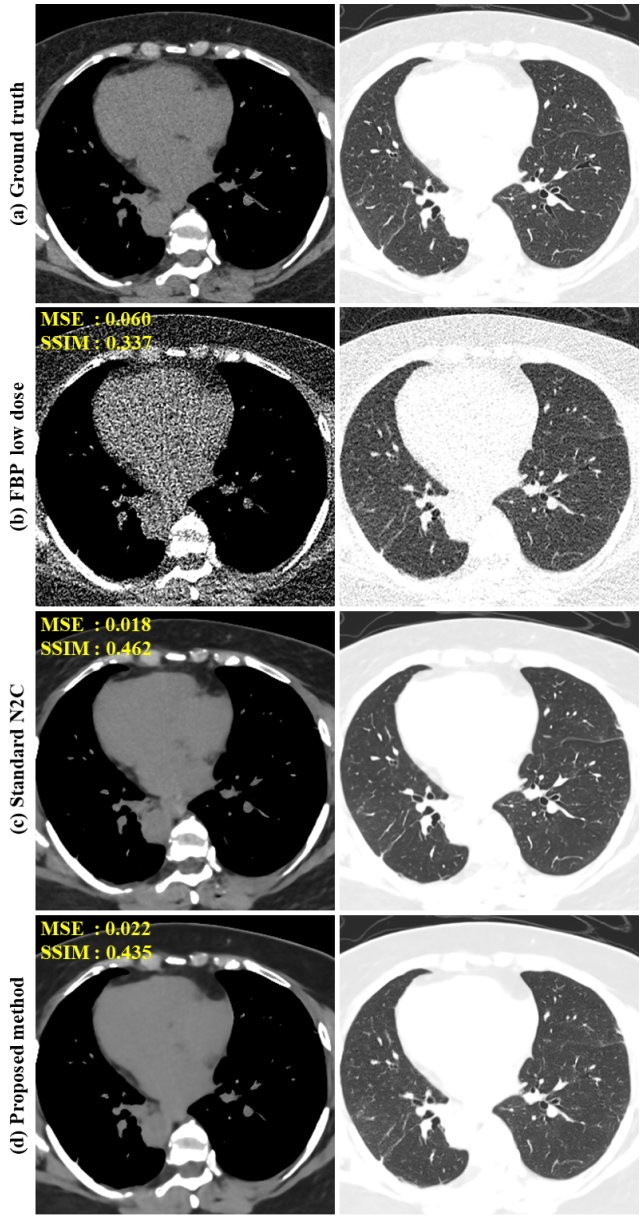


Fig. 3. A comparison of denoising method using simulated data. (a) the high dose MBIR image as ground truth (Helical, 120kVp, 105mAs, CTDIvol:14.5mSv.). (b) shows FBP low dose image which is reconstructed using simulated low dose data by FBP. The data is simulated 5% dose from full dose data used in (a). (c) and (d) show the image denoised with (b) as input. MSE and SSIM are calculated by each method using (a) as reference. Left column: soft tissue display (WL/WW=40/400). Right column: lung display (WL/WW=-600/1200)

C. Evaluation with clinical low dose data

We then tested our neural network on two independent real clinical low dose scans of one patient: the first scan is 5mAs (CTDIvol:0.3mSv, i.e., the ultra low dose case) and the second 50mAs (CTDIvol:3.3mSv). Because there are no high dose reference images, we justify the network performance through visual inspection. Fig. 4 shows the comparison for soft tissue region and lung region. Fig. 4 (a) shows the FBP image with lung kernel. Fig. 4 (b) and (c) show the show the

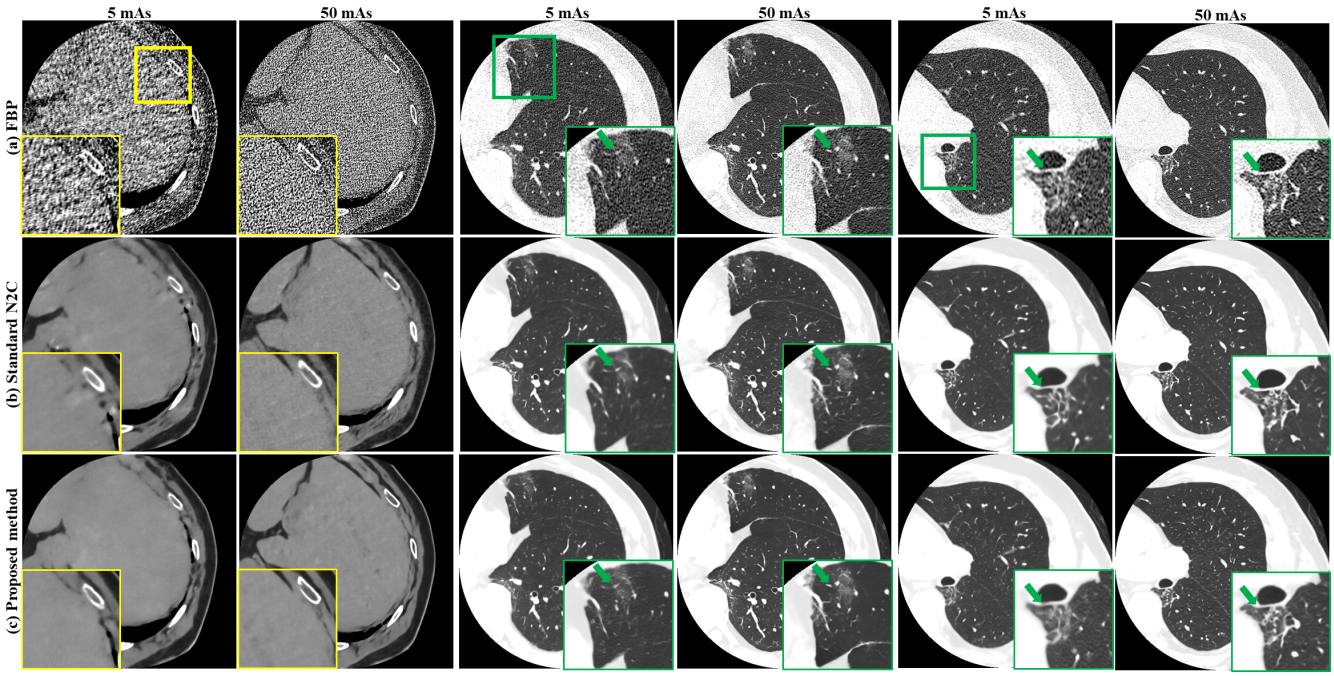


Fig. 4. A comparison of denoising method using two independent scans of one patient (Helical, 120kVp, rFOV:200mm). Ultra low dose case: 5mAs, CTDIvol:0.3mSv. Low dose case: 50mAs, CTDIvol:3.3mSv. Soft tissue display (WL/WW=40/400). Lung display (WL/WW=-600/1200). (a) FBP image with lung kernel. (a) is used as input for (b) and (c) denoinging methods. (b) Standard N2C. (c) Proposed method using synthetic noise training data.

image denoised with Fig. 4 (b) as input. Both networks show excellent denoising performance and the image quality has been well improved. The proposed method is able to produce similar results as the standard N2C method. Taking the lung image as an example, there is no significant difference between the proposed network and the standard N2C. Even for the ultra low dose case, the denoised image with the proposed method still maintains good spatial resolution of ground glass opacity (GGO) and fibrosis, and are quite comparable to the image denoised by the standard N2C method.

IV. Discussion and Conclusion

We have proposed an alternative training method for the specific CT image denoising task based on synthetic CT images. The synthetic CT images are created by adding independent image noise on top of clean CT images. We have estimated the realistic image noise from single clinical scans without using a low dose simulation tool. We have tested our neural network with both simulated and real low dose CT images. The DCNN trained with synthetic data shows comparable denoising performance as standard N2C training. The result justifies our previous postulation that an efficient denoising neural network might be highly dependent on the noise provided in training. To some extent our method implies that even if the background image is replaced with a non-CT image, the trained neural network may still be efficient for noise removal. This will be very interesting and we will investigate it in future work. More detailed evaluations and quantitative analysis should be conducted to further confirm the proposed method as well. The proposed method can be improved by including regularization techniques such as GAN, and this will be explored as well.

References

[1] H. Chen, Y. Zhang, M. K. Kalra, *et al.*, "Low-dose CT with a residual encoder-decoder convolutional neural network (RED-CNN)," *IEEE Trans. on Medical Imaging* **36**, 2524–2535 (2017).

[2] E. Kang, W. Chang, J.Yoo, *et al.*, "Deep convolutional framelet denosing for low-dose CT via wavelet residual network," *IEEE Trans. on Medical Imaging* **37**, 1358–1369 (2018).

[3] N. Yuan, J. Zhou, and J. Qi, "Low-dose ct image denoising without high-dose reference images," *Proc. SPIE 11072, 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine* (2019).

[4] D. Wu, K. Gong, K. Kim, *et al.*, "Consensus neural network for medical imaging denoising with only noisy training samples," *arXiv:1906.03639* (2019).

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolution networks for biomedical image segmentation," *MICCAI* (2015).

[6] Y. Fan, A. A. Zamyatin, and S. Nakanishi, "Noise simulation for low-dose computed tomography," *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)* , 3461–3463 (2012).

[7] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980* (2014).

Deep Learning-Based Cross-Scatter Correction for Clinical CT

Julien Erath, Tim Vöth, Joscha Maier, Eric Fournié, Karl Stierstorfer, Martin Petersilka, and Marc Kachelrieß

***Abstract*—Dual source CT (DSCT) systems use two measurement threads, A and B, offset by about 90 degree. A major problem in DSCT is the detection of scattered radiation from X-ray tube A in the detector of system B and vice versa. To correct for these cross-scatter artifacts we present three different deep learning-based methods. The 2D xDSE receives one projection as input and estimates the corresponding cross-scatter intensity. The 2D SSE_xDSE uses an approximation of the cross-scatter as an additional second input, which can be calculated from a previous reconstruction. Both methods are based on a two-dimensional deep convolutional neural network. The 3D xDSE uses a full set of projections and is based on a three-dimensional deep convolutional neural network. The networks are trained with scatter profiles obtained from our Monte Carlo simulation. All three methods are able to correct most of the scatter artifacts and clearly improve the image quality. 3D xDSE shows the best results closely followed by the 2D SSE xDSE and the 2D xDSE.**

I. Introduction

Dual source CT (DSCT) systems use two source-detector pairs offset by an angle of 90°or 95° depending on the system.

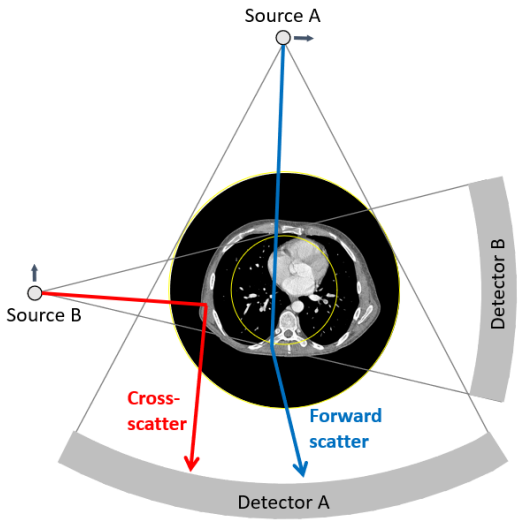


Fig. 1: Scanner geometry of a DSCT system. While forward scatter is radiation coming from the same measurement system, cross-scatter is scattered radiation from x-ray tube A detected in the detector of system B and vice versa.

Julien Erath is with the German Cancer Research Center, Heidelberg, Germany, with Siemens Healthineers, Forchheim, Germany and with the University of Heidelberg, Germany. Eric Fournié, Dr. Martin Petersilka and Dr. Karl Stierstorfer are with Siemens Healthineers, Forchheim, Germany. Tim Vöth, Dr. Joscha Maier and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center, Heidelberg, Germany and with the University of Heidelberg, Germany. Corresponding author: Julien Erath (julien.erath@dkfz.de)

By doubling the angular coverage the temporal resolution can be doubled. This is especially useful in time critical applications like cardiac CT [1]–[3].

A major challenge in DSCT is cross-scattered radiation (scattered radiation from x-ray tube A detected in the detector of system B and vice versa). Figure 1 shows the design of a DSCT system and the resulting cross-scatter. This effect can lead to artifacts and reduction of the contrast-to-noise ratio of the images. Scatter correction algorithms are crucial to correct the effect of the cross-scattered radiation [4].

For the correction of forward scatter, different correction techniques have been presented. Recently Maier et al. have presented the deep scatter estimation (DSE) which uses a deep convolutional neural network (CNN) to estimate forward scatter with the acquired projection data as input [5], [6]. The purpose of this paper is to propose and evaluate different deep learning-based techniques to correct the negative effect of cross-scatter. To do so, three different approaches are presented.

II. Methods and Material

A. Generation of Training Data

Simulated DSCT	
z-collimation at isocenter	70.2 mm
Field of measurement	500 mm
Detector rows	128
Detector columns	1024
Source-isocenter distance	595 mm
Isocenter-detector distance	490.6 mm
Angle between detectors	95°
Detector shape	cylindrical
Anti-scatter grid height	35 mm
Blade thickness	80 µm

TABLE I: Technical specifications of the simulated DSCT.

To obtain cross-scatter profiles we use our in-house Monte Carlo simulation [7]. The geometry of the Monte Carlo simulation is set to the scanner Somatom Force Siemens Healthineers, with some minor changes. For the perfect use of the downconvolutional layers we adapted the size of our detector pixels to 1024×128. The complete technical specifications are shown in table I. We are using cylindrical detectors and an anti-scatter grid. The distribution of the attenuation coefficient, that is required for the simulation of the CT data acquisition is obtained from clinical full body CT scans. The patients are assumed to consist of 4 materials (air, adipose tissue, soft tissue, and bone). One of these materials is assigned to each voxel based on CT value thresholds. In total 19 patients at 14 z-positions are Monte Carlo simulated. 15 patients are used for training, 2 patients are used as validation data and 2 patients

are used as test data. At each z-position the scatter profiles are simulated every 10° for the whole 360°. As for low tube voltages the scatter correction gets more sensitive to errors in the scatter estimation, the simulation is done at 80 kV. Being able to correct scatter artifacts for low tube voltages and therefore to correct images with a high scatter-to-primary ratio these algorithms can easily be transferred to higher tube voltage.

B. Deep Cross-Scatter Estimation (2D SSE_xDSE)

DSE uses a CNN to estimate scatter using a function of the acquired projection data as input [6]. The network is trained to predict the Monte Carlo scatter intensities. DSE has shown great potential for the estimation of forward scatter as it outperforms other scatter estimation techniques in terms of accuracy and it is much faster than the Monte Carlo simulation [5]. First experiments showed that the architecture of the DSE network can also be used to estimate cross-scatter [8]. So far the estimation of cross-scatter was just simulated without anti-scatter grid. Here, a deep cross-scatter estimation (xDSE) is evaluated with anti-scatter grid. The CNN architecture is similar to the U-Net [9], with an encoding path to extract the features from the input projection, which is then used in the decoding path to generate the output. Skip connections are used to pass high frequency spatial information from the encoding path to the decoding path. So far no deep learning-based cross-scatter estimation with anti-scatter grid has been published.

C. Deep Cross-Scatter Estimation with scatter along primary approximation (2D SSE_xDSE)

The idea of the 2D SSE_xDSE (2D single scatter estimation cross deep scatter estimation) algorithm is to give a second input to the neural network which improves the scatter estimation. Therefore a first order cross-scatter approximation is calculated. First an initial reconstruction from uncorrected-projections is computed. Then this reconstructed volume is used to simulate the interactions of X-rays with matter, but with one simplifying assumption that the cross-scatter only occurs along the direct way from the primary source to the detector. A simplified sketch of the calculation of the cross-scatter approximation for parallel-beam can be seen in Figure 2. The cross-scatter approximation for our algorithm is calculated for a cone-beam geometry. The projections $p_{\text{in}}(x, y)$ is the attenuation from the tube to the voxel (x, y):

$$p_{\text{in}}(x, y) = \int_{-\infty}^{x} d\hat{x}\ \mu(\hat{x}, y).$$

The linear attenuation coefficient $\mu(x, y)$, can be interpreted as the probability per path length that a photon interacts at the point (x, y). Similarly, $p_{\text{out}}(x, y)$ is the attenuation from the voxel to the detector along the primary:

$$p_{\text{out}}(x, y) = \int_{y}^{\infty} d\hat{y}\ \mu(x, \hat{y}).$$

The cross-scatter approximation $I_2(x)$ is proportional to:

$$I_2(x) \propto \int_{-\infty}^{\infty} dy\ e^{-p_{\text{in}}(x,y)}\ \mu(x, y)\ e^{-p_{\text{out}}(x,y)}.$$

The 2D SSE_xDSE algorithm obtains the acquired projection and the corresponding cross-scatter approximation as input. With the projection and the approximation as inputs the neural network is trained to learn the corresponding cross-scatter intensities. The neural network architecture is similar to the xDSE and is also based on a two-dimensional CNN.

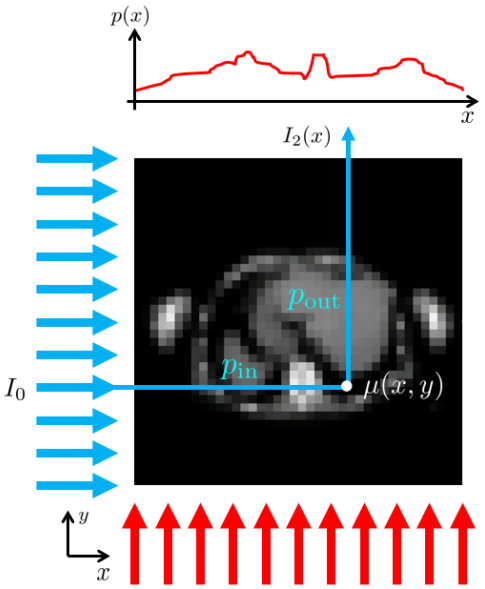


Fig. 2: Sketch of the calculation of the cross-scatter approximation (SSE).

D. 3D Deep Cross-Scatter Estimation (3D xDSE)

As our 2D xDSE uses 2D convolution kernels to estimate the scatter for a single projection, the network is able to leverage context across the rows and the columns of one projection to make predictions about the corresponding scatter, but not across the projections. This may be helpful for the prediction of the cross-scatter. Therefore we have designed a 3D CNN which uses several projections as input in order to estimate the cross-scatter for these projections. The 3D xDSE is using 3D convolution kernels to predict the scatter profiles at different angles. The use of 3D convolution kernels leads to an increased number of parameters and comes with increased computational cost. For this reason the 3D xDSE is trained with down-sampled scatter profiles. Therefore the resolution is adapted from 1024×128 to 512×16 (detector pixels). After obtaining the down-sampled scatter estimate these are up-sampled by bilinear interpolation to the original grid. The neural network uses 24 projections every 10° as input to cover a range of 240°. The range of 240°is chosen to cover all the projections which have an influence on the cross-scatter. The network is trained to predict the cross-scatter intensities for these 240°.

E. Loss-function

The scatter-to-primary ratio (SPR) is defined as the proportion of scattered X-rays to primary X-rays. We denote I_{prim} as the primary signal and I_{scatter} as the cross-scatter signal. The projection data p are given by

$$p = -\ln\left(I_{\text{prim}} + I_{\text{scatter}}\right) \quad = -\ln\left(I_{\text{prim}}\right) - \ln\left(1 + \frac{I_{\text{scatter}}}{I_{\text{prim}}}\right).$$

This equation shows that the influence of the cross-scatter on the image quality is controlled by the SPR [10]. Therefore in particular situations with large attenuation will lead to strong cross-scatter artifacts if not corrected. To consider the SPR for optimizing the cross-scatter estimation the SPMAPE (scatter-to-primary-weighted mean absolute percentage error) is implemented as loss function. While $I_{\text{scatter,DSE}}$ is the scatter estimation by the neural network and $I_{\text{scatter,MC}}$ is our scatter signal obtained by the Monte Carlo simulation. With N as number of detector pixels the SPMAPE is defined as

$$\begin{aligned}\text{SPMAPE} &= \frac{1}{N}\sum\left|\frac{I_{\text{scatter,MC}} - I_{\text{scatter,DSE}}}{I_{\text{scatter,MC}}} \times \frac{I_{\text{scatter,MC}}}{I_{\text{prim}}}\right| \\ &= \frac{1}{N}\sum\left|\frac{I_{\text{scatter,MC}} - I_{\text{scatter,DSE}}}{I_{\text{prim}}}\right|.\end{aligned}$$

F. Scatter-Correction

The total intensity is scatter-corrected by subtracting the scatter estimates $I_{\text{scatter,DSE}}$ in the projection domain. Before subtracting the scatter, the estimated scatter by the neural network is blurred with a Gaussian filter to remove remaining high frequencies. To prevent very small or even negative values, the resulting corrected intensity $I_{\text{corrected}}$ is clipped to 1% of I_{total}:

$$I_{\text{corrected}} = \begin{cases} I_{\text{total}} - I_{\text{scatter,DSE}}, & \text{if } I_{\text{scatter,DSE}} < 0.99 I_{\text{total}} \\ I_{\text{total}} - 0.99 I_{\text{total}}, & \text{otherwise} \end{cases}$$

III. Results

A. Projection Domain

The results of the different validation and training loss (SPMAPE) are shown in table II. In figure 3 the different cross-scatter estimations are plotted for one projection of the two test patients (averaged over all detector slices).

Correction method	Validation loss	Training loss
2D xDSE	0.66	0.43
2D SSE_xDSE	0.35	0.12
3D xDSE	0.21	0.11

TABLE II: Validation loss and Training loss (SPMAPE) for the three different correction methods.

B. Image Domain

To visualize the impact of our deep cross-scatter correction techniques in image domain, the corrected projections are reconstructed with the extended parallel back-projection (EPBP) algorithm [11]. The Monte-Carlo simulated cross-scatter artifacts are added to the projections and the estimates of the neural networks are subtracted to correct these artifacts. In figure 4, it can be seen that all three proposed methods clearly reduce cross-scatter induced artifacts in DSCT images. Almost all cupping artifacts are removed, leading to an improved image quality. For the two test patients, 14 different z-positions have been reconstructed and the corrected images have been compared to the ground truth images. Table III shows the means of the mean absolute error and the corresponding standard deviations of our corrected images. As it can be seen, 3D xDSE shows the best average in correcting the scatter artifacts. With 2D SSE_xDSE, the mean error is also very low, while with 2D xDSE the correction is less accurate.

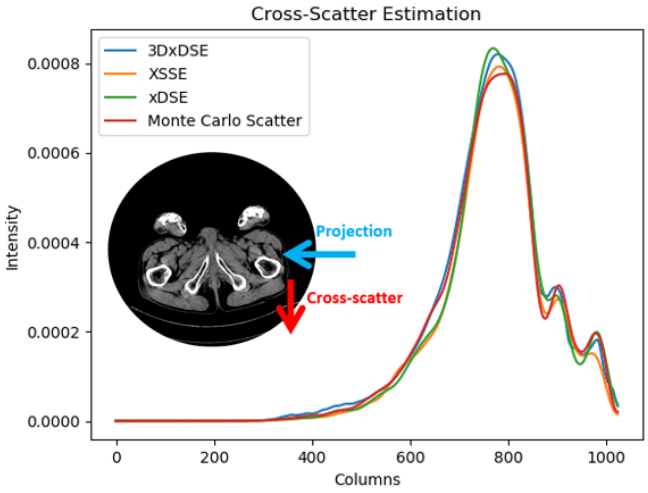


Fig. 3: Scatter estimation of the three different neural networks and the Monte Carlo scatter profile.

Correction method	Mean absolute error (MAE)
Uncorrected	45.66 HU ± 7.49 HU
2D xDSE	9.08 HU ± 2.44 HU
2D SSE_xDSE	4.81 HU ± 0.84 HU
3D xDSE	2.97 HU ± 0.67 HU

TABLE III: Mean absolute error of the reconstructed images with the different cross-scatter correction algorithms.

IV. Discussion

We propose three different deep learning-based methods to correct cross-scatter artifacts for DSCT. The best correction performance was obtained with 2D SSE_xDSE and 3D xDSE. For the 2D SSE_xDSE we calculated a cross-scatter approximation which was used as a second input to the neural network and clearly improves the results compared to the 2D xDSE. The ability to leverage context about different angles leads to improved performance for the 3D xDSE compared to 2D xDSE. Since with the use of an anti-scatter grid, forward scatter doesn't lead to strong image artifacts. Hence these methods are evaluated without the influence of forward scatter. In our following step we will present strategies for the correction of forward and cross-scatter in DSCT. The methods here are evaluated on simulated data. In the next stage we will also analyze if the methods can correct measured intensities.

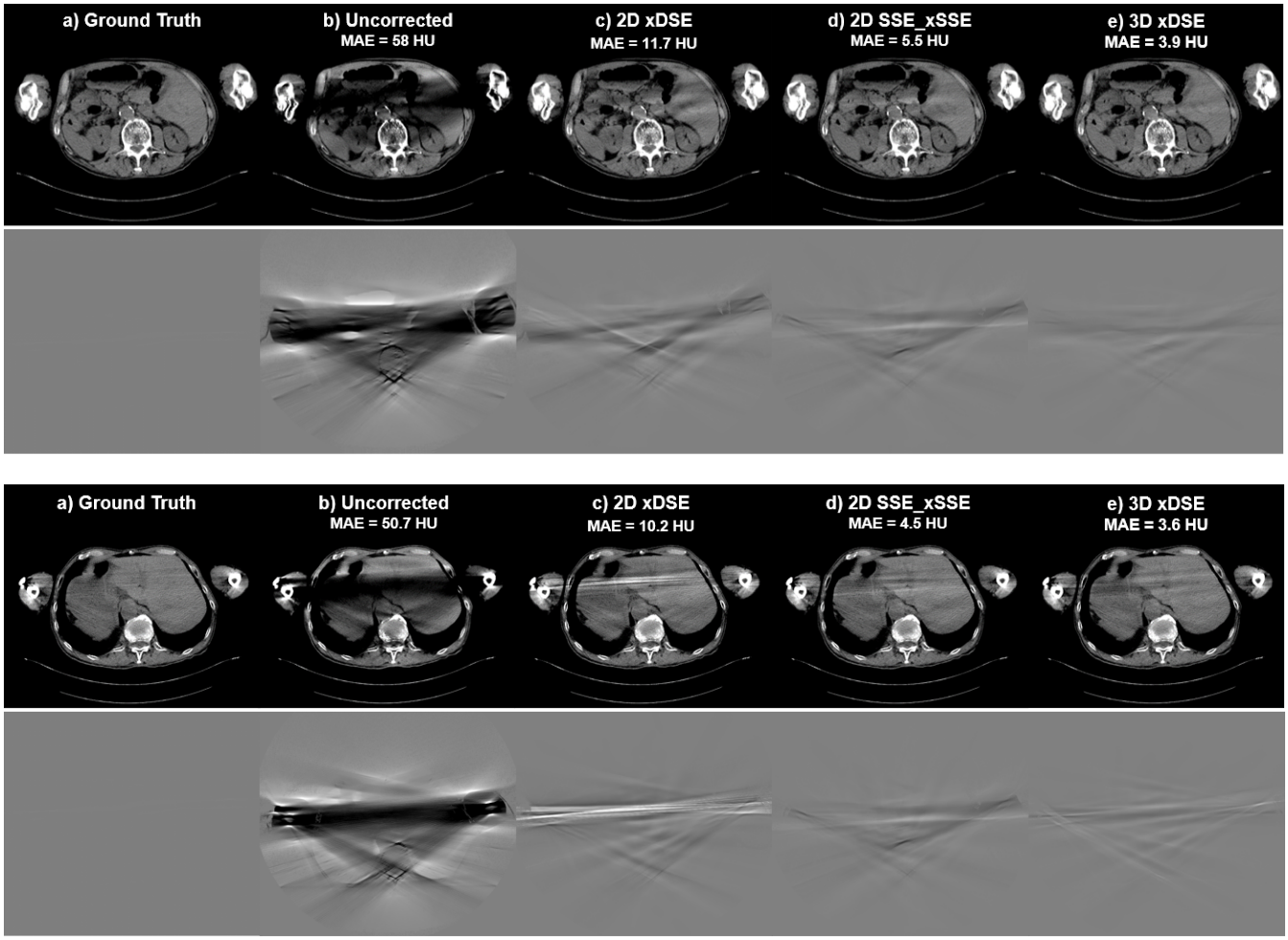


Fig. 4: Qualitative comparison of different cross-scatter correction methods for two different patients. The reconstructed images show a) Ground Truth (GT), b) Uncorrected, c) 2D xDSE, d) 2D SSE_xDSE and e) 3D xDSE ($C = 40$ HU, $W = 300$ HU). Below the reconstructed images the difference to GT is shown ($C = 0$ HU, $W = 300$ HU).

ACKNOWLEDGMENT

The Monte Carlo software RayConStruct-MC was provided by RayConStruct® GmbH, Nürnberg, Germany.

REFERENCES

[1] T. G. Flohr, H. Bruder, K. Stierstorfer, M. Petersilka, B. Schmidt, and C. H. McCollough, "Image reconstruction and image quality evaluation for a dual source CT scanner," *Medical Physics*, vol. 35, no. 12, pp. 5882–5897, 2008.

[2] T. G. Flohr, C. H. McCollough, H. Bruder, M. Petersilka, K. Gruber, C. Süß, M. Grasruck, K. Stierstorfer, B. Krauss, R. Raupach, A. N. Primak, A. Küttner, S. Achenbach, C. Becker, A. Kopp, and B. M. Ohnesorge, "First performance evaluation of a dual–source CT (DSCT) system," vol. 16, no. 2, pp. 256–268, Feb. 2006.

[3] M. Kachelrieß, M. Knaup, and W. A. Kalender, "Multithreaded cardiac ct," *Medical Physics*, vol. 33, no. 7Part1, pp. 2435–2447, 2006.

[4] M. Petersilka, K. Stierstorfer, H. Bruder, and T. Flohr, "Strategies for scatter correction in dual source CT," *Medical Physics*, vol. 37, no. 11, pp. 5971–5992, 2010.

[5] J. Maier, E. Eulig, T. Vöth, M. Knaup, J. Kuntz, S. Sawall, and M. Kachelrieß, "Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation," *Medical Physics*, vol. 46, no. 1, pp. 238–249, 2019.

[6] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep Scatter Estimation (DSE): Accurate Real-Time Scatter Estimation for X-Ray CT Using a Deep Convolutional Neural Network," *Journal of Nondestructive Evaluation*, vol. 37, no. 3, p. 57, Jul 2018.

[7] M. Baer and M. Kachelrieß, "Hybrid scatter correction for CT imaging," vol. 57, pp. 6849–6867, 2012.

[8] T. Vöth, J. Maier, J. Erath, and M. Kachelrieß, "Forward and cross-scatter estimation in dual source CT using the deep scatter estimation (DSE)," in *Medical Imaging 2019: Physics of Medical Imaging*, vol. 10948. Proc. SPIE 10948, 2019.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[10] H. Bruder, K. Stierstorfer, M. Petersilka, C. Wiegand, C. Suess, and T. Flohr, "Correction of cross-scatter in next generation dual source CT (DSCT) scanners," *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, vol. 6913, 04 2008.

[11] M. Kachelrieß, M. Knaup, and W. A. Kalender, "Extended parallel backprojection for standard three-dimensional and phase-correlated four-dimensional axial and spiral cone-beam CT with arbitrary pitch, arbitrary cone-angle, and 100% dose usage," *Medical Physics*, vol. 31, no. 6, pp. 1623–1641, 2004.

Deep Scatter Splines: Learning-Based Medical X-ray Scatter Estimation Using B-splines

Philipp Roser, Annette Birkhold, Alexander Preuhs, Christopher Syben, Norbert Strobel, Markus Korwarschik, Rebecca Fahrig, Andreas Maier

***Abstract*—The idea of replacing hardware by software to compensate for scattered radiation in flat-panel X-ray imaging is well established in the literature. Recently, deep-learning-based image translation approaches, most notably the U-Net, have emerged for scatter estimation. These yield considerable improvements over model-based methods. Such networks, however, involve potential drawbacks that need to be considered. First, they are trained in a data-driven fashion without making use of prior knowledge and X-ray physics. Second, due to their high parameter complexity, the validity of deep neural networks is difficult to assess. To circumvent these issues, we introduce here a surrogate function to model X-ray scatter distributions that can be expressed by few parameters. We could show empirically that cubic B-splines are well-suited to model X-ray scatter in the diagnostic energy regime. Based on these findings, we propose a lean convolutional encoder architecture that extracts local scatter characteristics from X-ray projection images. These characteristics are embedded into a global context using a constrained weighting matrix yielding spline coefficients that model the scatter distribution. In a first simulation study with 17 thorax data sets, we could show that our method and the U-Net-based state of the art reach about the same accuracy. However, we could achieve these comparable outcomes with orders of magnitude fewer parameters while ensuring that not high-frequency information gets manipulated.**

***Index Terms*—Splines, Neural Network, X-ray Scatter.**

I. Introduction

FLAT-PANEL detector cone-beam computed tomography (CBCT) is a powerful technique to enrich fluoroscopically-guided interventions (FGI) with 3-D information of a patient's anatomy. In contrast to conventional spiral computed tomography (CT), flat-panel X-ray acquisitions suffer, due to the larger X-ray field, from severe photon scattering effects. Without scatter compensation, reconstructed 3D images are likely to be impaired by streaking or cupping artifacts [1], [2]. In principal, two compensation strategies are possible: hardware-based and software-based approaches.

Hardware-based solutions describe direct manipulations of the X-ray field to either suppress incident scatter or modulate the primary beam. Typically, today's systems are equipped with anti-scatter grids mounted in front of the detector. They physically block a large portion of incoming scattered photons [3]. By increasing the object-detector-distance, scatter can be reduced as well. More sophisticated, experimental approaches include primary modulation [4] and slit scanning [5]. In general, and especially for interventional imaging, hardware-based solutions may complicate flexible adjustment of imaging settings or introduce artifacts of their own [6]. Therefore, and since hardware-based approaches are often accompanied by supporting algorithms in most cases anyways, purely software-based scatter correction methods are very desirable.

As in many applications, Monte Carlo (MC) simulation is considered the gold standard for scatter estimation [7], [8]. Its high computational complexity, however, still renders the actual application of MC simulation for scatter estimation in the interventional environment impracticable. As a solution, the Boltzmann transport equation can be solved directly using finite differences and discretization [9]. Yet, an acceptable run-time is only achievable by employing dedicated graphics processing units (GPU). This is why simpler, model-based approaches are still most widely used in clinical systems. They aim at estimating scatter from the X-ray projections directly, and are either based on simplified physical or analytical descriptions of X-ray scattering [10], [11] or expressed through convolutional kernels [12]. Unfortunately, model-based approaches are likely to fail if acquired images do not meet the underlying assumptions.

With the advent of convolutional neural networks (CNN) and most notably the U-Net [13] in the context of image-to-image translation, promising advances in the field of X-ray scatter correction were reported [14], [15]. The Deep Scatter Estimation (DSE) framework [14] is, for example, able to estimate X-ray scatter solely based on X-ray projections. In fact, it is on par with MC methods under optimal simulation conditions. However, the application of deep CNNs for image-to-image translation, especially in diagnostic imaging or FGIs, raises questions that are difficult to answer. Besides the overall difficulty in comprehending the operating principle of deep networks, the large number of trainable weights makes the U-Net highly dependent on the associated training data and therefore prone to adversarial effects and over-fitting. In addition, omitting prior knowledge can lead to higher maximum error bounds [16]. Finally, to maintain a fast run-time, U-Net-based scatter estimation is likely to require a dedicated, high-end GPU potentially adding considerable extra cost to the system.

We believe that by including additional constraints on the scatter characteristics within a CNN-based approach, improvements can be achieved. Since in the medical X-ray energy

P. Roser, A. Preuhs, C. Syben, and A. Maier are with the Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany. P. Roser is funded by the Erlangen Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany. A. Maier is principal investigator at the SAOT. A. Birkhold, M. Kowarschik, and R. Fahrig are employees of Siemens Healthcare GmbH, 91301 Forchheim, Germany. N. Strobel is with the Institute of Medical Engineering Schweinfurt, University of Applied Sciences Würzburg-Schweinfurt, 97421 Schweinfurt, Germany.

regime the majority of scattered radiation is low-frequency, there needs to exist a low-dimensional scatter representation. Here, we show that this low-frequency scatter distribution can be approximated by sparse bivariate splines [17] with an accuracy of up to 98 %. To this end, we train a CNN-based encoder to directly infer this representation from X-ray projections. Thanks to the spline approximation, the number of parameters to train could be reduced by orders of magnitude, while maintaining the accuracy of the established U-Net. By using a low-dimensional spline as surrogate, we also ensure low-frequency characteristics of the resulting scatter distribution. As a consequence, the proposed method cannot alter any high-frequency information because it is simply removing a smoothly varying background.

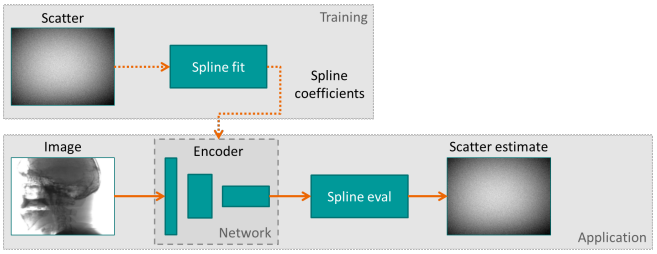


Fig. 1. Overview of the proposed method. Ground truth scatter distributions are approximated using bivariate B-splines. Then a convolutional encoder is trained to directly infer spline coefficients from the projection image. Afterwards, these spline coefficients are used to calculate the actual scatter distribution of the underlying input image.

II. Material and Methods

A. X-ray Scatter Model

In general, three photon interaction mechanisms are relevant for scatter in X-ray projections: (1) Rayleigh scattering, (2) Compton scattering, and (3) arbitrary, subsequent combinations of both, related as multiple scattering. Although each of them is subject to typical photon Poisson noise, the total scatter distribution in X-ray projections has low-frequency characteristics. In the following, we consider the flat-field normalized X-ray projection $\boldsymbol{I} \in \mathbb{R}^{w\times h}$ with width w and height h in pixels. The X-ray projection consists of the primary signal $\boldsymbol{I}_\mathrm{p} \in \mathbb{R}^{w\times h}$ described using the Beer-Lambert law and scatter $\boldsymbol{I}_\mathrm{s} \in \mathbb{R}^{w\times h}$. The scatter signal includes all scattering mechanisms. Thus $\boldsymbol{I}$ at pixel $\boldsymbol{u} = (u_w, u_h)$ is given by

$$\boldsymbol{I}(\boldsymbol{u}) = \underbrace{\int_E I_0(E) e^{-\int_x \mu(x,E)dx} dE}_{\boldsymbol{I}_\mathrm{p}(\boldsymbol{u})} + \boldsymbol{I}_\mathrm{s}(\boldsymbol{u})\,, \qquad (1)$$

with the photon energy E, X-ray spectrum $I_0(E)$, and the linear attenuation coefficient $\mu(x, E)$.

We approximate $\boldsymbol{I}_\mathrm{s}$ at pixel $\boldsymbol{u}$ using a bivariate spline of degree k and order $n = k + 1$. The spline is defined by a $w_c \times h_c$ grid of coefficients $\boldsymbol{C} \in \mathbb{R}^{w_c\times h_c}$ on a fixed knot grid $\boldsymbol{T} \in \mathbb{R}^{(w_c-k+1)\times(h_c-k+1)}$. This gives us the surrogate $\tilde{\boldsymbol{I}}_{\mathrm{s,n}}(\boldsymbol{u})$ at pixel $\boldsymbol{u} = (u_0, u_1)$

$$\tilde{\boldsymbol{I}}_{\mathrm{s,n}}(\boldsymbol{u}) = \sum_{i=i_0}^{n+i_0} \sum_{j=j_0}^{n+j_0} \boldsymbol{C}_{i,j} B_{i,n,\boldsymbol{T}}(u_0) B_{j,n,\boldsymbol{T}}(u_1)\,, \qquad (2)$$

with the basis splines B. Note that the starting indices i_0 and j_0 depend on both the evaluated pixel $\boldsymbol{u}$ and the knot grid $\boldsymbol{T}$. Since $\boldsymbol{T}$ is fixed for any X-ray projection, we fit the associated splines in a least-square sense, following Dierckx [18]. The error residual $\varepsilon_r = ||\boldsymbol{I}_\mathrm{s} - \tilde{\boldsymbol{I}}_\mathrm{s}||_2$ approaches 0 with increasing number of spline coefficients.

B. Proposed Method

Based on our scatter model, we propose to infer scatter spline coefficients directly from the X-ray projection domain. The outline of the presented method is depicted in Fig. 1. From ground truth MC scatter simulations, we find a sparse scatter representation by fitting bivariate splines with respect to an arbitrary but fixed knot grid $\boldsymbol{T}$. The resulting spline coefficients $\boldsymbol{C}$ are then used to train a deep encoder to extract such a sparse scatter representations from an X-ray projection. By evaluating the inferred spline coefficients using the previously defined grid, we are able to obtain the low-frequent portion of scatter distributions for any desired pixel spacing.

The encoder employed consists of blocks of two convolutional layers, each followed by a rectified linear unit (ReLU). The number of blocks is arbitrary and manually adjustable to the input and output dimensions. Between two convolutional blocks, the dimensions of the feature maps are halved by averaging a neighborhood of 2×2 pixels for each feature map. The number of intermediate feature maps is constant for all levels. Before each convolution, replication padding is applied to preserve spatial dimensions. The last block is followed by a weighted sum over all channels to reduce the number of intermediate channels to one. Eventually, the latent variable of width w_l and height h_l is vectorized. This convolutional path $\mathcal{S}_{\boldsymbol{\theta}} : \mathbb{R}^{w\times h} \mapsto \mathbb{R}^{w_l\cdot h_l\times 1}$ extracts latent variables that locally characterize X-ray scattering. However, scattered radiation globally affects image formation. To account for this, we model each spline coefficient $C_{i,j}$ as the weighted, normalized sum of all intermediate latent variables. This global relationship can be expressed by multiplying the latent variable with matrix $\boldsymbol{W} \in \mathbb{R}^{w_c\cdot h_c\times w_l\cdot h_l}$, which is only feasible due to the low dimensionality of the latent space and not directly transferable to the DSE.

To be comparable to the DSE, the network parameters $\boldsymbol{\theta}$ and $\boldsymbol{W}$ are optimized by minimizing the mean absolute percentage error (MAPE) with respect to the ground truth

$$\boldsymbol{\theta}, \boldsymbol{W} = \arg\min_{\boldsymbol{\theta},\boldsymbol{W}} \sum_i^{w_c} \sum_j^{h_c} \frac{||\left[\boldsymbol{W}\mathcal{S}_{\boldsymbol{\theta}}(\boldsymbol{I})\right]_{i,j} - \boldsymbol{C}_{i,j}||_1}{\boldsymbol{C}_{i,j}}, \qquad (3)$$

where $[\cdot]_{i,j}$ is the access operation for a vectorized matrix. Since $\boldsymbol{W}$ models a weighted sum, we impose two constraints: (1) each element of $\boldsymbol{W}$ must be positive $W_{i,j} > 0, \forall i, j$ and (2) each row sums up to unity $\sum_j W_{i,j} = 1, \forall i$.

C. Ground Truth Data Simulation

Acquiring matching pairs of training data using conventional X-ray imaging is tedious and time-consuming. As an

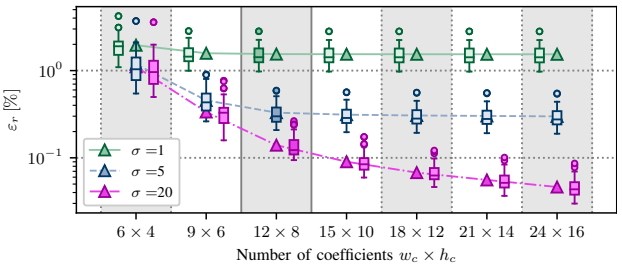


Fig. 2. Absolute average spline approximation error ε_r (triangles) and associated box plots for different numbers of coefficients for strong ($\sigma = 20$), medium ($\sigma = 5$), and no ($\sigma = 1$) Gaussian smoothing.

alternative, we used the X-ray transport code MC-GPU [19] to generate synthetic training data. As inputs for our simulation, openly accessible CT scans from The Cancer Imaging Archive (TCIA) [20] were used. In total, we extracted 17 thorax phantoms from the CT Lymph Nodes torso [21] data sets. To prepare the phantoms for MC simulation, we employed a basic pre-processing pipeline [22] to remove non-patient objects, and assign tissues and densities to each voxel. Based on the simulated scatter distributions, we fit bivariate cubic B-splines.

For each X-ray projection, we simulated 5×10^{10} primary photons sampled from a 100 keV tungsten spectrum. In total, we considered three slightly different iso-center positions $\boldsymbol{o} \in \mathbb{R}^3$, seven different projection angles $\alpha \in \{0, 15, 30, \ldots, 90\}$ in °, and fixed source-to-detector (1300 mm) and source-to-iso-center (800 mm) distances, respectively. The virtual detector is 399 mm × 299 mm large and consists of 1248×928 pixels. To reduce simulation noise, we filter the scatter signals using a Gaussian kernel $\boldsymbol{G}_\sigma$.

III. Results

A. Spline Approximation Capability

Initially, we investigated the quality of splines for medical X-ray scatter approximation. To this end, we fit cubic splines ($k = 3$) with a different number of coefficients $\boldsymbol{C} \in \mathbb{R}^{w_c \times h_c}$ to the scatter signal and juxtaposed the associated residuals ε_r. Figure 2 summarizes the results for $w_c \times h_c \in \{48 \times 32, 24 \times 16, 12 \times 8, 6 \times 4, 3 \times 2\}$ for all considered simulated X-ray scatter signals. Evidently, only marginal improvements can be observed with increasing number of coefficients. The best trade-off between sparsity and a low approximation error was obtained with 12×8 coefficients. Hence, we considered $\boldsymbol{C} \in \mathbb{R}^{12 \times 8}$ for further analysis.

B. Experimental Setup

In accordance to the DSE [14], our encoder network is trained with down-sampled X-ray projections comprising 384×256 pixels and targets 12×8 spline coefficients. We excluded one patient for validation during training (1×21 X-ray projections) and one patient for testing afterwards (2×21 X-ray projections). Hence, the encoder was trained using 273 X-ray projections from 13 patients and a batch size of 16. For data augmentation, we used random horizontal flips and random Poisson noise as these techniques preserve physical plausibility. The network was trained using adaptive moments optimizer with default parameters [23] subject to minimize the MAPE to the target MC scatter simulations. The results were compared to the U-Net with respect to the ground truth MC scatter simulations. We trained the baseline U-Net and our network architecture with different depths d and number of initial feature maps c (see Tab. I).

TABLE I
Number of parameters to train and quantitative test results for all investigated network architectures for different depths d and number of feature maps c. For better assessment, the median absolute percentage error (MedAPE) is also given.

Architecture	d	c	Parameters	MAPE [%]	MedAPE [%]
U-Net	6	8	12.6 M	4.66(409)	3.59
		16	50.3 M	5.06(389)	4.24
	5	8	3.1 M	5.04(425)	3.97
		16	12.6 M	5.41(413)	4.60
Ours	7	8	10.1 k	4.84(401)	3.80
		16	32.7 k	4.53(382)	3.51
	6	8	15.8 k	4.66(422)	3.55
		16	35.0 k	4.67(429)	3.65

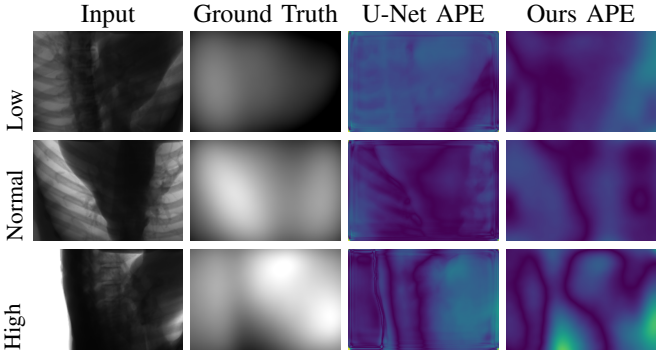


Fig. 3. Qualitative comparison between both approaches for three different views of different intensity ranges (low, normal, high). Input and ground truth are windowed to give a good visual representation. Absolute percentage errors (APE) are in the range of 0 % to 25 % (dark blue to yellow).

C. Comparison of the Network Architectures

Table I shows the pixel-wise MAPE with respective standard deviations and the median absolute percentage error (MedAPE) of the investigated network architectures on the two test patients. Evidently, both network architectures perform similarly for all parameter combinations and median error rates range from 3.55 % to 4.60 %. However, slight trends can be observed. The U-Net configurations with fewer parameters (8 initial feature maps) yield 0.64 % lower median error rates on average compared to their high-parameters counterparts. Such a trend is not seen for our network architecture. Furthermore, it reaches overall lower error rates on average compared to the U-Net. Most strikingly, our method is capable to maintain the U-Net's accuracy while reducing the number of parameters to train by two to three orders of magnitude.

To assess the network architectures from a qualitative point of view, Fig. 3 depicts three example projections, their associ-

ated scatter signals, and the absolute percentage errors (APE) of the predictions by the U-Net ($d = 6, c = 8$) and our encoder ($d = 7, c = 16$) with respect to the ground truth. Overall, the scatter distributions are smooth and almost no salient features are observable, in particular in the proposed approach.

IV. Discussion and Conclusion

We developed a novel method to estimate X-ray scatter from projection images in a data-driven fashion. Leveraging the ability of B-splines to approximate low-frequency signals, we were able to tremendously reduce the number of training parameters to arrive at results comparable to the current state of the art. We presented a proof of concept based on a simulation study involving 17 thorax data sets at one fixed peak tube voltage relevant for medical X-ray diagnostics.

Although these preliminary results appear promising, several limitations of our study need to be discussed. First, as advanced pre-processing has shown potential to improve learning-based scatter estimation by approximately $4\,\%$ [14] as well, future studies need to incorporate different pre-processing techniques in the evaluation. Second, at this point in time we cannot yet make a statement on how 3D reconstruction benefits, as associated experimental results are not yet available. Third, in contrast to the DSE [14], we only considered one anatomic region and one tube peak voltage. Therefore, no assumption on our method's generalizability to other anatomic regions and X-ray energy regimes can be made. It cannot be ruled out that our lean encoder architecture does not have the same capacity as the U-Net.

Besides these potential limitations, our method comes with interesting characteristics that are very desirable in the interventional environment. First, the parameter reduction leads to a lower computational complexity. This allows for (a) accelerated reconstruction, (b) seamless transition between 3D and 2D imaging, and (c) it can be implemented on a wide variety of processing hardware. In addition, it is questionable whether a single network is needed that can adapt to highly varying tube voltages and anatomic regions. Multiple dedicated networks could be used instead. Second, by using a sparse spline as surrogate signal, our method ensures that only low frequencies get altered in the X-ray projection image. Hence, it is very unlikely that potentially diagnostically relevant anatomic features get removed.

In sum, the use of splines as X-ray scatter surrogate signal has the potential to improve learning-based X-ray scatter estimation algorithms, as it requires significantly fewer parameters and much reduced computational complexity.

Disclaimer: The concepts and information presented are based on research and are not commercially available.

References

[1] E.-P. Rührnschopf and K. Klingenbeck, "A general framework and review of scatter correction methods in x-ray cone-beam computerized tomography. part 1: Scatter compensation approaches," *Med Phys*, vol. 38, no. 7, pp. 4296–4311, 2011.

[2] ——, "A general framework and review of scatter correction methods in cone-beam ct. part 2: Scatter estimation approaches," *Med Phys*, vol. 38, no. 9, pp. 5186–5199, 2011.

[3] H.-P. Chan, Y. Higashida, and K. Doi, "Performance of antiscatter grids in diagnostic radiology: Experimental measurements and monte carlo simulation studies," *Med Phys*, vol. 12, no. 4, pp. 449–454, 1985.

[4] B. Bier, M. Berger, A. Maier, M. Kachelrieß, L. Ritschl, K. Müller, J.-H. Choi, and R. Fahrig, "Scatter correction using a primary modulator on a clinical angiography c-arm ct system," *Med Phys*, vol. 44, no. 9, pp. 125–137, 2017.

[5] R. Bhagtani and T. G. Schmidt, "Simulated scatter performance of an inverse-geometry dedicated breast ct system," *Med Phys*, vol. 36, no. 3, pp. 788–796, 2009.

[6] D. M. Gauntt and G. T. Barnes, "Grid line artifact formation: A comprehensive theory," *Med Phys*, vol. 33, no. 6, pp. 1668–1677, 2006.

[7] W. Zbijewski and F. J. Beekman, "Fast scatter estimation for cone-beam x-ray ct by combined monte carlo tracking and richardson-lucy fitting," in *IEEE Nucl Sci Symp Conf Rec*, vol. 5, 2004, pp. 2774–2777.

[8] G. Poludniowski, P. M. Evans, V. N. Hansen, and S. Webb, "An efficient monte carlo-based algorithm for scatter correction in keV cone-beam CT," *Phys Med Biol*, vol. 54, no. 12, pp. 3847–3864, 2009.

[9] A. Wang, A. Maslowski, P. Messmer, M. Lehmann, A. Strzelecki, E. Yu, P. Paysan, M. Brehm, P. Munro, J. Star-Lack, and D. Seghers, "Acuros cts: A fast, linear boltzmann transport equation solver for computed tomography scatter – part ii: System modeling, scatter correction, and optimization," *Med Phys*, vol. 45, no. 5, pp. 1914–1925, 2018.

[10] W. Yao and K. W. Leszczynski, "An analytical approach to estimating the first order scatter in heterogeneous medium. ii. a practical application," *Med Phys*, vol. 36, no. 7, pp. 3157–3167, 2009.

[11] M. Meyer, W. A. Kalender, and Y. Kyriakou, "A fast and pragmatic approach for scatter correction in flat-detector CT using elliptic modeling and iterative optimization," *Phys Med Biol*, vol. 55, no. 1, pp. 99–120, 2009.

[12] B. Ohnesorge, T. Flohr, and K. Klingenbeck-Regn, "Efficient object scatter correction algorithm for third and fourth generation ct scanners," *Eur Radiol*, vol. 9, no. 3, pp. 563–569, Mar 1999.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Med Image Comput Comput Assist Interv*, ser. Lecture notes in computer science, vol. 9351. Springer, 2015, pp. 234–241.

[14] J. Maier, E. Eulig, T. Vöth, M. Knaup, J. Kuntz, S. Sawall, and M. Kachelrieß, "Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation," *Med Phys*, vol. 46, no. 1, pp. 238–249, 2019.

[15] P. Roser, X. Zhong, A. Birkhold, N. Strobel, M. Kowarschik, R. Fahrig, and A. Maier, "Physics-driven learning of x-ray skin dose distribution in interventional procedures," *Med Phys*, vol. 46, no. 10, pp. 4654–4665, 2019.

[16] A. Maier, C. Syben, B. Stimpel, T. Würfl, M. Hoffmann, F. Schebesch, W. Fu, L. Mill, L. Kling, and S. Christiansen, "Learning with known operators reduces maximum error bounds," *Nat Mach Intell*, vol. 1, pp. 373–380, 2019.

[17] I. J. Schoenberg, "Contributions to the problem of approximation of equidistant data by analytic functions," *Quart Appl Math*, vol. 4, pp. 45–99, 1946.

[18] P. Dierckx, *Curve and Surface Fitting with Splines*. New York, NY, USA: Oxford University Press, Inc., 1993.

[19] A. Badal and A. Badano, "Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit," *Med Phys*, vol. 36, no. 11, pp. 4878–4880, 2009.

[20] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository," *J Digit Imaging*, vol. 26, pp. 1045–1057, 07 2013.

[21] H. Roth, L. Le, S. Ari, K. Cherry, J. Hoffman, S. Wang, and R. Summers, "A new 2.5 d representation for lymph node detection in ct. the cancer imaging archive," 2018. [Online]. Available: http://doi.org/10.7937/K9/TCIA.2015.AQIIDCNM

[22] P. Roser, A. Birkhold, A. Preuhs, B. Stimpel, C. Syben, N. Strobel, M. Kowarschik, R. Fahrig, and A. Maier, "Fully-automatic ct data preparation for interventional x-ray skin dose estimation," in *Bildverarbeitung für die Medizin*, 03 2020, pp. 125–130.

[23] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 12 2014.

Artificial intelligence is powerful...

Shared Intelligence is (empowering)

Because everyone in the cancer fight shares the same goal.

We're combining the insights and smarts from a wide range of sources, such as artificial intelligence, machine learning, and data analytics, to build a network of collective knowledge to help you deliver advanced care. We call it Shared Intelligence™. Because when all intelligence is shared, we'll get closer to what we're all fighting for—a world without fear of cancer.

Visit varian.com/victories to see how Shared Intelligence can help elevate your cancer care.

Safety information: Radiation may cause side effects and may not be appropriate for all cancers.

© 2020 Varian Medical Systems, Inc. Varian is a registered trademark, and Shared Intelligence is a trademark of Varian Medical Systems, Inc.

varian

Thursday, August 6

Photon Counting

Breast CT

Poster Session 3

Security and Phase Contrast CT

The Number of Energy Windows for Photon Counting Detectors: Is More Actually More?

Katsuyuki Taguchi, *Senior Member*

***Abstract*— It has been debated whether photon counting detectors (PCDs) with moderate numbers of energy windows, N_E, perform better than PCDs with larger N_E. A larger N_E results in fewer photons in each energy window, which degrades the signal-to-noise ratio of each datum. Unlike energy-integrating detectors, PCDs add very little electronic noise to measured counts; however, there exists electronic noise on the pulse train, to which multiple energy thresholds are applied to count photons. The noise may increase uncertainty of counts within energy windows, however, this effect has not been studied in the context of spectral tasks. The purpose of this study was to investigate the effect of N_E on the quality of the spectral information. We obtained the following three types of PCD data with various N_E (=2–24) and noise levels using a Monte Carlo simulation: (A) The ideal PCD with no electronic noise; (B) realistic PCDs with electronic noise added to the pulse train; and (C) hypothetical PCDs with electronic noise added to each energy window's output, similar to energy-integrating detectors. We evaluated the Cramér–Rao lower bound of estimation for the following two spectral tasks: (a) water–bone material decomposition and (b) K-edge imaging with gold. We will show that, for realistic PCDs, the quality of the spectral information monotonically improves when N_E is increased.**

***Index Terms*—Photon counting detector, electronic noise, number of energy windows.**

I. Introduction

Energy-sensitive photon counting detectors (PCDs) have been developed for x-ray computed tomography [1-7]. One of many design and operating parameters that need to be optimized is the number of energy windows (N_E). There is consensus that it is desirable to count the number of photons of the transmitted x-ray spectrum using two or more energy windows and that PCDs with too few windows may be suboptimal for spectral tasks such as material decomposition [1]. There is no clear consensus, however, as to whether PCDs with larger N_E, such as ten, may perform better or worse than PCDs with a moderate N_E, such as six, if PCDs with any number of energy windows can be built. The debate is not settled mainly because the effect of the following two factors have not been thoroughly investigated in the context of spectral tasks: the presence of electronic noise and the signal-to-noise of each datum.

Energy-integrating detectors, which are currently used for x-ray computed tomography, add electronic noise to each datum due to dark current and the analog-to-digital conversion process. Although PCDs are often said to have no electronic noise, there exists electronic noise on the pulse train, which makes pulse heights uncertain. A simple counting PCD with a single energy threshold can set the threshold just above the noise level, count all of the photons above the threshold, and mitigate the effect of the electronic noise to a negligible level. But a spectral PCD with multiple energy thresholds sets the thresholds within the diagnostic energy range to count photons within multiple energy windows. When pulse heights are uncertain due to the noise on the pulse train, thresholding such pulses near their peaks may result in uncertain counts within each energy window. The effect of the uncertainty (in addition to quantum noise) on spectral tasks such as material decomposition has not been studied yet. Particularly it is not clear if increasing N_E (hence, having fewer photons within each narrower energy window) would have an additional negative effect.

PCDs with a larger N_E have narrower window widths in general, resulting in fewer counts within each energy window. The signal-to-noise ratios of narrower energy windows are therefore worse than those of wider ones. The conventional belief is that using less data with better signal-to-noise ratios is better than using more data with worse signal-to-noise ratios. This may be an erroneous belief resulting from long-lasting experiences with energy-integrating detectors, where independent electronic noise is added to each datum, which thereby splits photons into a larger number of datasets and degrades the overall quality of data. Alternatively, the belief may result from experiences with suboptimal processing methods, for example a polynomial fitting that is sensitive to outliers, applied to PCD data.

Therefore, the aim of this work was to study the quality of PCD data in the context of spectral tasks with different numbers of energy windows, N_E, different types of electronic noise, and different noise levels.

II. Methods and Materials

We obtained the following three types of PCD data with various N_E (=2–24) and noise levels using a Monte Carlo simulation: (A) The ideal PCD with no electronic noise; (B) realistic PCDs with electronic noise added to the pulse train;

Manuscript received on January 12, 2020. The work was not supported by extramural funds.

K. Taguchi is with the Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD 21287 USA (email: ktaguchi@jhmi.edu).

and (C) hypothetical PCDs with electronic noise added to each energy window's output, similar to energy-integrating detectors with electronic noise added to each datum. We generated three sets of 3,000 noise realizations (9,000 noise realizations in total) for each attenuation setting. We used 1 ms acquisition with x-rays at 140 kVp and 500 mA, resulting in ~4,500 photons per pixel per noise realization on average. The N_E was 2, 3, 4, …, 24; one energy threshold was always set at 20 keV and the other thresholds were swept from 25 keV to 135 keV in 5-keV increments. The electronic noise added to the pulse train for the realistic PCD was Gaussian-distributed with a standard deviation of 2.5 keV, 5.0 keV, or 7.5 keV. The electronic noise of 2.5 keV was in close alignment with other Monte Carlo simulations and an actual PCD [8]. The electronic noise added to the windowed data for the hypothetical PCD was Gaussian-distributed with a standard deviation of 0.1%, 0.2%, or 0.5% of the mean counts of the entire spectrum (which corresponded to about 5, 9, or 23 counts per energy window, respectively).

The quality of the information the PCD data contained was evaluated using the Cramér–Rao lower bound (CRLB) for the following two spectral tasks: (a) two-dimensional material decomposition with water and bone; and (b) three-dimensional K-edge imaging with gold, water, and bone. The following processing was similar to those used in Refs. [9, 10]. The changes in the mean values were estimated by adding a small amount of water, bone, or gold. The Fisher information matrix was then computed numerically with two assumptions: (1) measured counts were multivariate Gaussian-distributed and (2) the covariance remained unchanged with an additional small amount of attenuator. We think that Assumption (1) can be justified because mean counts for most of the energy windows were larger than 30 when the minimum CRLB was found for a given N_E. Assumption (2) might have resulted in an over-estimation of CRLB, especially for larger N_E, but it was necessary to make the computation stable. In the end, no adverse effect was observed, as the CRLB for the ideal PCD converged and did not increase as the N_E was increased. The inverse of the Fisher information matrix was then calculated. The CRLB of the variance of each basis material was the corresponding diagonal element of the inverse matrix. An unbiased sample standard deviation of CRLB values over the three sets of 3,000 noise realizations was computed.

III. Results

A. Water–Bone Material Decomposition

The results of the water thickness estimation are presented and discussed in this section. The results of the bone thickness estimation were similar to those of the water thickness estimation and therefore, to save space, are not presented. The normalized CRLB presented hereafter was the lowest CRLB for a given N_E and data type (among many different combinations of threshold energies). The CRLB was normalized by that of the ideal PCD with N_E=24.

Figure 1 plots the CRLB over N_E for the ideal PCD, the realistic PCD with 5.0-keV electronic noise, and the hypothetical PCD with 0.2% electronic noise. The ideal PCD

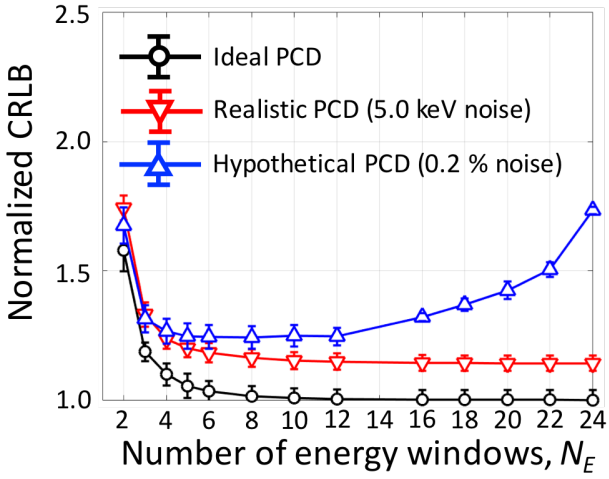


Fig. 1. The normalized CRLB of water thickness estimation for the water–bone material decomposition. Error bars indicate standard deviations.

showed that the CRLB decreases (hence, the spectral information improves) monotonically with increasing N_E. The adverse effect of the electronic noise was comparable for both the realistic and hypothetical PCDs at N_E=2. The effect of the noise was consistent for the realistic PCD regardless of N_E, improving CRLBs monotonically with increasing N_E and the offset from the ideal PCD almost independent of N_E. In contrast, for the hypothetical PCD, the adverse effect of the electronic noise increased with increasing N_E, which eventually overwhelmed the benefit of the improved spectral information

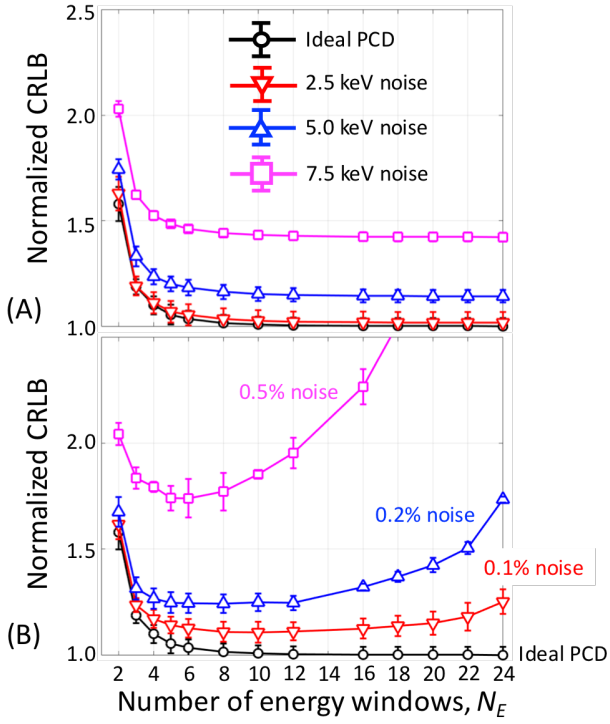


Fig. 2. The normalized CRLB of water thickness estimation for the water–bone material decomposition with the ideal PCD and the realistic PCD with various noise levels (A) *and* the ideal PCD and the hypothetical PCD with various noise levels (B). Error bars indicate standard deviations.

and ended up increasing the CRLB. For N_E=2, 4, 8, 12, 16, 20, and 24, the CRLBs were 1.744, 1.231, 1.183, 1.149, 1.144, 1.142, and 1.142, respectively, for the realistic PCD and 1.676, 1.265, 1.242, 1.245, 1.321, 1.424, and 1.734, respectively, for the hypothetical PCD.

The noise level (5.0 keV) for the data shown in Fig. 1 was higher for the realistic PCD (as the noise level was 2.0–2.5 keV for some PCDs [8]), whereas the noise level (0.2%) may be either higher or lower for the hypothetical PCD depending on the number of true photons, i.e., radiation dose levels. These two data were chosen to clearly demonstrate the two different dependencies on N_E discussed above.

Figure 2A plots CRLBs over N_E for the ideal PCD and the realistic PCD with various noise levels. The realistic PCD with greater noise had a larger CRLB compared to the ideal PCD with the same N_E. Notice, however, that regardless of noise levels, CRLBs monotonically decreased with increasing N_E. This indicates that the quality of realistic PCD data for the water–bone material decomposition task improves with increasing N_E. Figure 2B presents the ideal PCD and the hypothetical PCD with various noise levels. The optimal N_E was smaller when electronic noise was larger: The best CRLB was obtained at N_E=6 with 0.5% noise, N_E=8 with 0.2% noise, and N_E=10 with 0.1% noise.

The minimum N_E required for the two-dimensional water–bone material decomposition is two. The CRLB for N_E=2, 3, and 24 were 1.746±0.202, 1.333±0.205, and 1.145±0.195, respectively. The improvement between N_E=2 and N_E=3 was significant, which corresponded to 68.7% of the improvement between N_E=2 and N_E=24 [i.e., (1.746–1.333)/(1.746–1.145)×100%=68.7%].

B. K-Edge Imaging

The results of the gold thickness estimation are presented in Fig. 3 in the same format as Fig. 1. The results of the water and bone thickness estimations were similar to those presented in Sec. 3.A., and therefore, to save space, are not presented here. All of the PCDs had results that are similar to non-K-edge imaging: The CRLB of both the ideal and the realistic PCDs monotonically improved with increasing N_E, whereas the CRLB achieved the best value with a moderate N_E for the hypothetical PCD. The only difference is, for the realistic PCD, the effect of electronic noise was larger (i.e., the normalized CRLB values were higher) with K-edge imaging than it was with water–bone material decomposition. For example, the CRLB with N_E=24 was 1.282 at 2.5 keV noise and 1.688 at 5.0 keV noise for K-edge imaging in contrast to 1.017 and 1.142, respectively, for water–bone material decomposition. This is because the electronic noise added to the pulse train blurred the K-edge, and hence, degraded the spectral information for K-edge imaging.

Figure 4 plots CRLBs over N_E for the ideal PCD and the realistic PCD with various noise levels (Fig. 4A) and those for the ideal PCD and the hypothetical PCD with various noise levels (Fig. 4B). Similar to Fig. 2, regardless of noise levels, CRLBs monotonically decreased with increasing N_E for the realistic PCD, whereas for the hypothetical PCD, the best

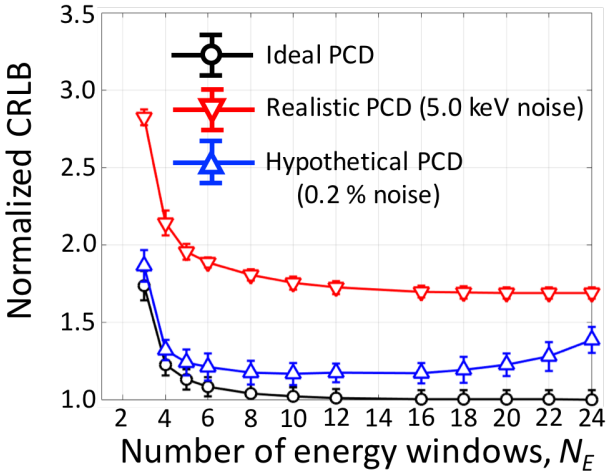


Fig. 3. The normalized CRLB of gold thickness estimation for the K-edge imaging. Error bars indicate standard deviations.

CRLBs were found at N_E=8 with 0.5% noise, N_E=10 with 0.2% noise, and N_E=16 with 0.1% noise.

The minimum N_E required for the three-dimensional K-edge imaging is three. The CRLB for the realistic PCD with N_E=3, 4, and 24 were 2.878±1.312, 2.070±0.925, and 1.658±0.727, respectively. The improvement between N_E=3 and N_E=4 corresponded to 66.2% of the improvement between N_E=3 and N_E=24 [i.e., (2.878–2.070)/(2.878–1.658)×100%=66.2%].

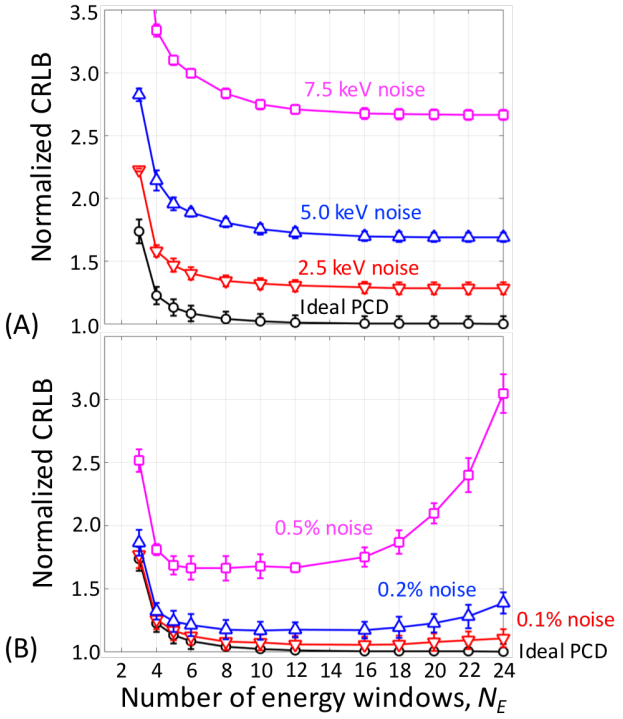


Fig. 4. The normalized CRLB of gold thickness estimation for the K-edge imaging with the ideal PCD and the realistic PCD with various noise levels (A) *and* the ideal PCD and the hypothetical PCD with various noise levels (B). Error bars indicate standard deviations.

IV. Discussion

With the ideal and realistic PCDs, the best performance was achieved when N_E was the largest, as the performance monotonically improved with increasing N_E with no adverse effect. Even though the signal-to-noise ratio of each datum worsened with increased N_E, the total amount of the spectral information contained in all of the energy windows improved. If the performance of PCDs degrades with increased N_E, it must be due to employing a suboptimal algorithm that uses either a wrong statistical model or no statistical model at all.

With the hypothetical PCD, on the other hand, the best performance was found at a moderate N_E and the optimal N_E was smaller when the noise level was higher (or when the dose level is lower). This aligns well with conventional beliefs regarding how the quality of the spectral information changes with N_E. This N_E-dependent performance is hypothetical, unless a PCD adds significant independent noise to each windowed datum.

The spectral information is spread over the energy range differently for the two tasks. The information for the K-edge material is strongest near the K-edge, while lower and higher energies carry much weaker signals for the K-edge materials. In contrast, the information for water and bone (or any non-K-edge materials) is spread over the energy range fairly consistently. Regardless of such a difference, the study has shown that a larger number of N_E provides better performance (i.e., lower CRLBs) for the realistic PCD.

The study focused on physics aspects of the effect of N_E and did not include engineering design factors that are related to N_E, such as power consumption and circuitry densities. When PCDs are designed, an optimal N_E will be determined based on the improved spectral information with larger N_E taking into account the increased engineering challenges that come with larger N_E.

The study has limitations. (1) The effect of pulse pileups was not included to simplify the story. Nonetheless, we believe that the trend should remain true even if pulse pileups are included. (2) Other conditions such as x-ray spectra and pixel sizes would affect the CRLB; however, we believe that the observations and trends would remain the same, and therefore, the conclusions would remain the same. The study did include two distinctly different spectral tasks and they showed very little difference. (3) The study did not include algorithms that use PCD data sub-optimally. The CRLB demonstrated that an unbiased statistical estimator should be able to provide a smaller noise variance with a larger N_E.

V. Conclusions

For both the ideal and realistic PCDs, the quality of the spectral information monotonically improves as N_E is increased, for both non-K-edge imaging and K-edge imaging. PCDs with larger N_E are better than those with smaller N_E. Adding one energy window to the minimum necessary N_E for a given task (i.e., making it N_E=3 for non-K-edge imaging and N_E=4 for K-edge imaging) gained 66.2–68.7% of the improvement N_E=24 provided. For the realistic PCDs the effect of the electronic noise added to the pulse train depends on the spectral task and K-edge imaging was more sensitive to the noise than non-K-edge imaging.

References

[1] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical Physics,* vol. 40, no. 10, p. 100901, 2013. [Online]. Available: http://link.aip.org/link/?MPH/40/100901/1 http://dx.doi.org/10.1118/1.4820371.

[2] K. Taguchi, "Energy-sensitive photon counting detector-based X-ray computed tomography," *Radiological Physics and Technology,* journal article vol. 10, no. 1, pp. 8-22, 2017, doi: 10.1007/s12194-017-0390-9.

[3] C. H. McCollough *et al.*, "Achieving Routine Submillisievert CT Scanning: Report from the Summit on Management of Radiation Dose in CT," *Radiology,* vol. 264, no. 2, pp. 567-580, August 1, 2012 2012, doi: 10.1148/radiol.12112265.

[4] S. Leng *et al.*, "Photon-counting Detector CT: System Design and Clinical Applications of an Emerging Technology," *RadioGraphics,* vol. 39, no. 3, pp. 729-743, 2019, doi: 10.1148/rg.2019180115.

[5] C. H. McCollough, S. Leng, L. Yu, and J. G. Fletcher, "Dual- and Multi-Energy CT: Principles, Technical Approaches, and Clinical Applications," *Radiology,* vol. 276, no. 3, pp. 637-653, 2015, doi: doi:10.1148/radiol.2015142631.

[6] R. Ballabriga *et al.*, "Review of hybrid pixel detector readout ASICs for spectroscopic X-ray imaging," *Journal of Instrumentation,* vol. 11, no. 01, p. P01007, 2016. [Online]. Available: http://stacks.iop.org/1748-0221/11/i=01/a=P01007.

[7] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: Technical Principles and Clinical Prospects," *Radiology,* vol. 289, no. 2, pp. 293-312, 2018, doi: 10.1148/radiol.2018172656.

[8] K. Taguchi, K. Stierstorfer, C. Polster, O. Lee, and S. Kappler, "Spatio-energetic cross-talk in photon counting detectors: Numerical detector model (PcTK) and workflow for CT image quality assessment," *Medical Physics,* vol. 45, no. 5, pp. 1985-1998, 2018, doi: doi:10.1002/mp.12863.

[9] S. S. Hsieh and N. J. Pelc, "A dynamic attenuator improves spectral imaging with energy-discriminating, photon counting detectors," (in eng), *IEEE Trans Med Imaging,* vol. 34, no. 3, pp. 729-39, Mar 2015, doi: 10.1109/tmi.2014.2360381.

[10] S. S. Hsieh and N. J. Pelc, "Improving pulse detection in multibin photon-counting detectors," (in eng), *Journal of medical imaging (Bellingham, Wash.),* vol. 3, no. 2, p. 023505, Apr 2016, doi: 10.1117/1.jmi.3.2.023505.

X-ray Photon-Counting Data Correction via Deep Learning

Mengzhou Li, David S. Rundle, and Ge Wang, *Fellow, IEEE*

***Abstract*—The energy/spectral dimension of X-ray photon-counting detectors (PCDs) potentially brings great benefits such as material decomposition, beam hardening and metal artifact reduction, as well as low-dose CT imaging. However, practical applications of X-ray PCDs are currently limited by several technical issues, particularly by charge splitting (charge sharing, K-shell fluorescence escape or re-absorption) and pulse pile-up effects which distort the energy spectrum and compromise the data quality. Correction of raw PCD measurements with hardware improvement and analytic modeling is rather complicated. Hence, here we propose a deep neural network-based PCD data correction approach which directly maps imperfect data to the ideal data in a supervised learning mode. In this pilot study, we demonstrate the feasibility of the approach with realistic synthetic PCD data and initial experimental data. Specifically, a complete PCD detection model incorporating charge splitting and pulse pile-up effects is constructed for distortion data simulation. The distorted PCD data as well as the ground truth counterparts are fed to our dedicated deep adversarial network. Then, the trained network is used to correct independently generated PCD data for testing. The results show that the trained network successfully recovers the ideal spectrum from the distorted measurement within $\pm 6\%$ error. Major data and image fidelity improvements are clearly observed in both projection and reconstruction domains.**

***Index Terms*—Photon-counting detectors (PCDs), PCD data correction, pulse pile-up, charge sharing, deep learning.**

I. Introduction

X-RAY PCDs are drawing an increasing attention from both industrial and academic sides in the medical imaging field. The unique energy discrimination ability of PCD makes it highly desirable for decomposing materials, solving or alleviating beam hardening and metal artifact problems. Also, the low noise nature and optimal weighting on low-energy and high-energy photons of PCDs increase the image contrast for radiation dose reduction, compared to conventional energy-integrating detectors.

However, in practice performance of current PCDs is far from flawless, and several technical issues still hinder wider applications of PCDs, especially, counting speed limit and spectrum distortion. These problems mainly come from the pulse pileup (PU) and charge splitting (CS) (including charge sharing, K-shell fluorescence escape or re-absorption) effects. The PU effect limits the counting speed, and smaller pixels can help raise the total counting rate but may increase the CS effect as well resulting in a serious distortion of the detected spectrum. Vendors put efforts for increasing the counting rate while keeping the spectral fidelity through hardware improvements, e.g., high speed circuits, active charge summing circuits, while many researchers developed less expensive correction algorithms. The general idea behind many algorithms is to estimate the real spectrum by iteratively maximizing the likelihood for the measurement and the calculated output with a forward PCD detection model. The key to the success relies on the accuracy of the forward model. Monte Carlo simulation could provide accurate results but at a huge computational cost, and many researchers devised fast models for PCD simulation. Taguchi tackled the CS simulation with his Photon Counting Toolkit (PcTK) [1] at reasonable accuracy. Recently, Philips scientists proposed an exact analytical PU prediction model [2] based on Bierme's mathematical results for solving the level crossing problem [3]. Even with that progress in PCD modeling, it is still hard to obtain accurate correction on real PCDs due to the complexity of modeling various physical effects inside the PCDs.

Inspired by the impressive performance of deep learning in many challenging tasks, here we propose a deep learning-based data-driven approach for PCD data correction to circumvent the complex modeling of PCD detection, and directly map the distorted raw data to the ideal data. Previously, Badea tried along a similar direction with a combination of a simple two-layer network and a bilateral filtration denoising method to improve the material decomposition accuracy [4]. Different from his approach, we fully rely on the deep learning technique for spectral distortion correction as well as noise reduction at the same time in an unified framework. Moreover, we focus on the counts correction of PCD data which is a more fundamental and more challenging task.

M. Li and G. Wang are with the Department of Biomedical Engineering, Rensselaer Polytechnic, Troy, NY, 12180 USA (wangg6@rpi.edu).

D. S. Rundle is with JairiNovus Technologies Ltd., 119 Heartland Drive, Butler, PA, USA.

II. Methods

In this pilot study, we first use synthesized distorted PCD data which include CS, PU and Poisson noise for feasibility demonstration. Although the PCD simulation model may not be absolutely perfect, the simulated data should be realistic enough for our purpose.

A. PCD Detection Model

In this work, we only consider the paralyzable-type PCD with non-negative monopolar pulses. The basic idea of our detection distortion pipeline can be readily extended to the general situation. First, we calculate the expected counts in each energy window of every pixel after the CS effect (including the Poisson noise as well), then adding PU distortion pixel-by-pixel. That is, we include both CS and PU effects in our model. The rationale is that the CS effect decides the distribution of charge clouds on the pixel grid, and the PU effect comes from the readout process of the charge clouds. The explicit photon arrivals are random, and then the temporal sequences of charge clouds are formed. Since the timestamps of signals in a pixel are independent from each other and follow a Poisson distribution, the expected counts above an energy threshold after the PU effect is still a level crossing problem. The splitting of the electron cloud of a photon by neighboring pixels generates correlations among pixels, but this does not affect the PU calculation of each individual pixel.

Suppose that the spectrum of X-ray photons reaching the PCD is $\boldsymbol{p}(x, y, E)$, and the detector spectral response function due to the CS effect, which can be simulated with PcTK, is $\mathbf{srf}(x, y, E_{in}, E_{out})$, the noise free signal of $\boldsymbol{p}$ after CS can be formulated as,

$$\boldsymbol{q}(x, y, E) = \int \boldsymbol{p}(x, y, E') \underset{x,y}{\otimes} \mathbf{srf}(x, y, E', E)\mathrm{d}E', \qquad (1)$$

where $\boldsymbol{q} \in \mathbb{R}^{N_x \times N_y \times N_E}$ is the corresponding clean charge splitting signal of $\boldsymbol{p}$; $\otimes_{(x,y)}$ denotes the convolution operation with respect to width and height dimensions. Since we do not study the statistical correlation of noises between pixels, the Poisson noise can be added pixelwise, and the noisy signal after CS becomes $\mathbf{P}(\boldsymbol{q})$, where $\mathbf{P}(\cdot)$ stands for a Poisson distribution.

We modified Roessl's PU estimation model to incorporate the CS effect. The PU distortion is calculated pixel-by-pixel due to its nature

of an intrapixel effect, thus, let us focus on one pixel from $\mathbf{P}(\boldsymbol{q})$ and assume that the height or energy distribution of signal pulses it detected is $TR(U)$ (We do not discriminate energy E and pulse height U in this work, since they are linearly linked together). Instead of in the histogram form, the total response can also be described in a temporal form,

$$X(t) = \sum_{j \in \mathbf{Z}}^{\lambda} U_j g(t - t_j), \tag{2}$$

where U_j represents the signal height of photon j, which is independently and identically distributed following the probability mass function $p(U)$ of the "pulse height spectrum" $TR(U)$; $g(t)$ is the signal shape kernel, and t_j is a timestamp; λ is the total counts of all signals. λ and $p(U)$ are defined as,

$$\lambda = \int_{-\infty}^{\infty} TR(U)dE, \quad p(U) = \frac{TR(U)}{\lambda} \tag{3}$$

Based on Bierme's results on the level crossing problem, if $g(t)$ is a normalized non-negative monopolar pulse, the up-crossing times $N_X(U)$ of $X(t)$ over a pre-defined threshold U can be calculated in the Fourier domain as:

$$\hat{N_X}(u) = \lambda \exp\left[\lambda \int_{-\infty}^{\infty} \left(\hat{P}[ug(t)] - 1\right) dt\right] \frac{\hat{P}(u) - 1}{iu}, \tag{4}$$

where $\hat{P}(u)$ and $\hat{N_X}(u)$ are the inverse and forward Fourier transforms of $p(U)$ and $N_X(U)$ respectively, and i is the imaginary unit.

Let us define the cumulative total response function $\Phi_{TR}(U)$ as

$$\Phi_{TR}(U) = \begin{cases} \int_U^{\infty} TR(U')dU', & U \geq 0 \\ 0, & U < 0 \end{cases} \tag{5}$$

The relationship between $TR(U)$ and $\Phi_{TR}(U)$ in the Fourier spectrum space is as follows:

$$\hat{\Phi_{TR}}(u) = \left[\hat{TR}(u) - \hat{TR}(0)\right] \Big/ iu \tag{6}$$

By substituting Eqs. 3 and 6 into Eq. 4, we obtain

$$\hat{N_X}(u) = \hat{\Phi_{TR}}(u) \exp\left[-\hat{TK}(u)\right] \tag{7}$$

where

$$\hat{TK}(u) = \int_{-\infty}^{-\infty} \left\{\hat{TR}(0) - \hat{TR}[ug(t)]\right\} dt \tag{8}$$

Finally, the recorded counts with the PU distortion can be calculated by transforming $\hat{N_X}(u)$ back to the count space:

$$N_X(U) = \frac{1}{2\pi} \int_{-\infty}^{-\infty} \hat{\Phi_{TR}}(u) \exp\left[-\hat{TK}(u)\right] \exp\left(-iuU\right) du \tag{9}$$

B. Synthetic Data Generation

To generate distorted PCD data, the ideal spectral projections of synthesized material phantoms are first calculated with the X-ray source spectrum simulated using *SpekCalc* [5] and serve as the noisy-free ground truth data. Then spectral distortions as well as Poisson noise are added through our PCD detection model described above.

During the simulation of the X-ray source spectrum, the peak energy and minimum energy were set to $120keV$ and $12keV$ respectively with the bin width being $1keV$. The distance between the phantom and the source was set to 1 meter, and the default source filter consisting of $0.8mm$ Beryllium, $1mm$ Aluminium and $0.11mm$ Copper.

Synthesized 3D Shepp-Logam material phantoms were used to form the population of objects to be scanned, each of which consisted of 5 ellipsoids with random shapes located completely inside a sphere of radius $1.28cm$ centered in a $256 \times 256 \times 256$ cubic. Five material types (soft tissue, adipose tissue, brain grey and white matter, blood and cortical bone) were assigned to the five ellipsoids. A pixel inside the sphere could be covered by multiple ellipsoids. In that case, the material composition of the pixel was set as the equal-volume mixture of the involved ellipsoids' materials. The gaps between the ellipsoids in the sphere were filled with water, while the pixels outside the sphere were treated as air. The spectral linear attenuation coefficient (LAC) data for those representative human tissues and other materials are from the National Institute of Standards and Technology (NIST) database [6]. Totally, 180 spectral projections were generated from each phantom energy bin by energy bin ($1keV$ width) in the parallel beam configuration for convenience and without loss of generality. Those projections will be fed to the distortion simulation followed and also to the network for training as label data later.

The detector spectral response function was simulated using PcTK 3.2 with the following parameters: (1) the effective electron clouds radius r_0 $24\mu m$; (2) the electronic noise parameter σ_e $2.0keV$; (3) the pixel size of the detector d_{pix} $110\mu m$; and (4) the thickness of the sensor d_z $2.0mm$. r_0 and σ_e were selected as default for the best agreement with the Monte Carlo simulation [1], while the d_{pix} and d_z values were selected in reference to the real parameters of the Medipix-3 PCD. Based on the detection model, the noisy signals after CS with 1-keV-width energy windows were generated for the PU calculation. The pulse kernel used for PU calculation was set to $g(t) = e^{-t/\tau}$ for positive t and zero otherwise, with the dead time τ being $10ns$.

Since the number of energy bins of current PCDs is often no more than 8, we used thresholds with a step size of $10keV$ to convert the counts-over-threshold data to the counts-inside-bin data $\boldsymbol{m}$. The thresholds are $20keV$, $30keV$, $40keV$, $50keV$, $60keV$, $70keV$, $80keV$, $90keV$, $100keV$, and $110keV$, which are equally distributed over the whole spectral range.

C. Correction Network

1) Network structure: A Wasserstein Generative Adversarial Network (WGAN) has been designed for PCD data correction. The generator undertakes the data correction task, and is designed for inversion of the forward simulation model. First, a pixelwise fully-connected sub-network (dePUnet) is used to correct the intrapixel distortion of PU, and then a convolutional sub-network (deCSnet) is used for the interpixel cross talk correction of CS as well as noise reduction. A discriminator adopted from WGAN-VGG helps the generator training via the competition between the two during the training. Two losses usesd are the WGAN loss for adversarial training and the correction loss measured as the mean square error (MSE) of the absolute count difference as well as the mean absolute error (MAE) of the relative difference to enforce pixelwise accuracy in counts. The overall architecture of the network is shown in Fig. 1.

2) Network training: Totally, 10 PCD datasets ($180 \times 256 \times 367 \times 9$) were randomly extracted into patches of size $16 \times 16 \times 9$ for network training. The patch training strategy minimizes the GPU memory occupation, and forces the network to focus on a small scope since the cross-talk only happens between neighboring pixels. The extracted 1,841,400 pairs of $\boldsymbol{p}$ and $\boldsymbol{m}$ patches were shuffled to remove the connections between pairs, and normalized with a constant 400,000 before feeding to the network, and 92,070 pairs were used for validation. In addition, 5 PCD datasets were generated for testing only.

We used the Adam optimization method ($\alpha = 1.0 \times 10^{-4}$, $\beta_1 = 0.9$ and $\beta_2 = 0.999$) to train the network, with the mini-batch size 1024. The value of λ for adversarial loss was set to 10, and the

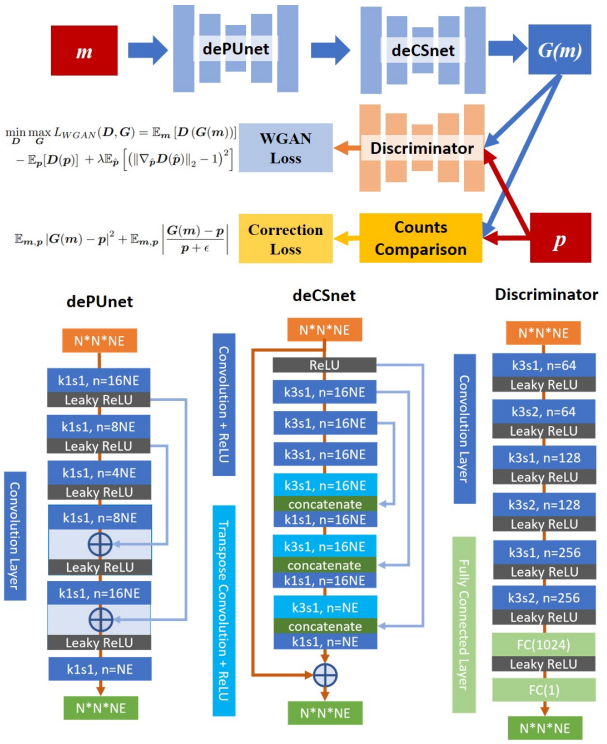


Fig. 1. Correction network structure, where n, k and s stand for the number of kernels, kernel size and convolutional stride, respectively, $\hat{\boldsymbol{p}}$ is uniformly sampled along the straight line between paired $\boldsymbol{G}(\boldsymbol{m})$ and $\boldsymbol{p}$, the penalty coefficient λ is a weighting constant, and ϵ is a small constant set to 1×10^{-4} to stabilize the ratio.

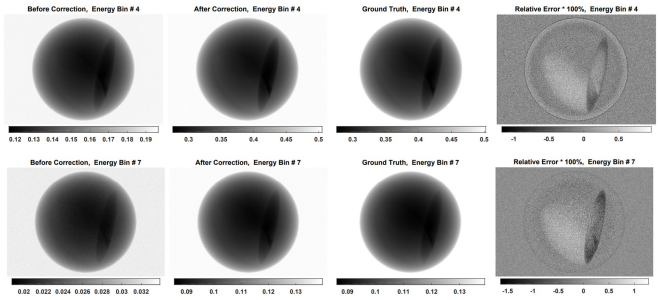


Fig. 2. Projections before and after correction in energy bins 4 ($50-59keV$) and 7 ($80-89keV$). Counts are normalized with 400,000.

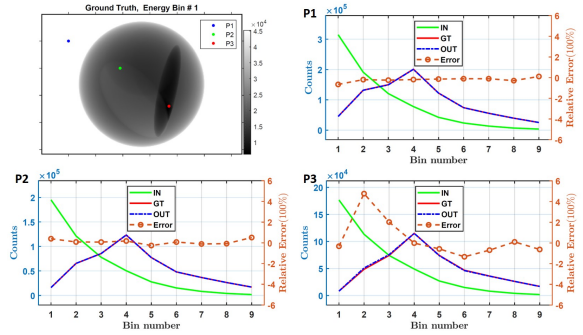


Fig. 3. Energy profiles for three points before and after correction, where IN denotes the input profile before correction, GT denotes the ground truth profile, OUT represents the output of the correction network, and Error is the relative error defined as (OUT-GT)/GT.

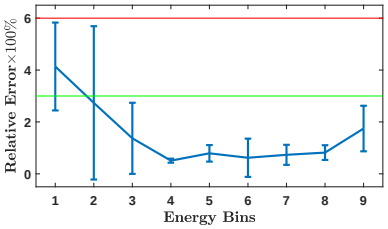


Fig. 4. Mean 99% percentile relative error curve with 3 folds of std. errorbar.

weight balancing the correction loss and the adversarial loss was set to 1000. The network converged after 40 epochs.

D. Real PCD Data Correction Framework

Two problems need to be addressed before applying the proposed method for real PCD data correction, which are (1) the unknown real X-ray source spectrum and (2) the spatial non-uniformity of detector pixels due to charge trapping. The source spectrum can be estimated with indirect measurements [7], but the spatial non-uniformity of PCD may disqualify the data for the convolutional model and must be addressed with calibration [8]. After that, a general work flow can be used to obtain training data as follows. First, we generate a sufficient number of material phantoms with 3D printing and liquid tissue surrogate techniques [9] to acquire real raw data. Then, with the known geometry and material composition the ground truth spectral projections can be calculated with the LAC data from NIST. Finally, the ground truth projections and the distorted PCD measurements are registered and fed to the network for training, validation and testing. Therefore, the trained network can be applied for the network-based feed-forward correction of the PCD data acquired under similar scanning settings, which will be our future work.

III. Results

The trained network on synthetic data was applied on the testing dataset for correction effectiveness evaluation. To demonstrate the correction effects, the spectral projections before and after correction were both visually and quantitatively compared.

A. Comparison in Projections

A qualitative comparison of representative images in two energy channels before and after correction is shown in Fig. 2. The scale ranges are well aligned to the ground truth after correction and obvious noise reduction is also evidenced by clearer backgrounds. The correction results are almost the same with the ground truth, and with negligible relative errors. Figure 3 shows correction results of three representative points experiencing different attenuation levels. The accurate overlapping of corrected spectra and the ground truths demonstrates the success of the proposed method. From the relative error curve, we can find that the correction errors are within $\pm1\%$ for median or less attenuation cases in all energy bins, and for the heavy attenuation case, the errors are within $\pm5\%$ for all bins and less than $\pm3\%$ for bins energy higher than $40keV$.

To further investigate the stability of the correction, 5 testing datasets each with 180 projection views were used to estimate the error range. The 99% percentile relative error in each view for each energy bin was obtained, then the mean value and standard deviation (std.) were calculated out of these data, and the results are plotted in Fig. 4. Suppose the errors follow normal distributions, the results suggest that 99% pixels are corrected to the ground truth within 6% relative error for all energy bins at a 99.7% confidence, and with errors smaller than 3% for bins 3 to 9 ($40keV$ to $110keV$).

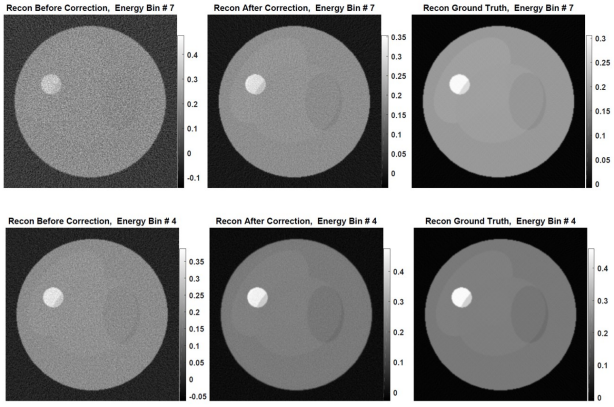


Fig. 5. Reconstructions of data in bin 4 ($50-59keV$) and bin 7 ($80-89keV$) before and after correction. Units in cm^{-1}.

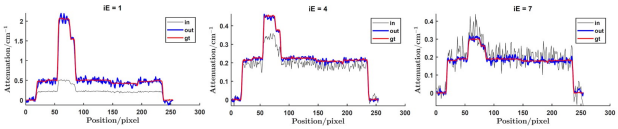


Fig. 6. Horizontal attenuation profiles (105th row, through the white spot in Fig. 5) in the reconstructions, where *in* denotes uncorrected profiles, *out* represents profiles after correction, and *gt* stands for ground truth profiles.

TABLE I
QUANTITATIVE METRICS ON THE RECONSTRUCTION SLICE SHOWN IN FIGURE 5

GT as ref.	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8	Bin 9
Uncorrected (SSIM)	0.4669	0.7679	0.9114	0.9636	0.9691	0.9562	0.9172	0.8265	0.6633
Corrected (SSIM)	**0.9835**	**0.9896**	**0.9943**	**0.9977**	**0.9931**	**0.9976**	**0.9936**	**0.9890**	**0.9698**
Uncorrected (PSNR)	19.99	21.58	23.47	25.43	25.34	22.96	19.01	14.19	9.09
Corrected (PSNR)	**33.63**	**34.20**	**35.10**	**37.41**	**31.76**	**35.75**	**31.04**	**28.36**	**23.73**
Improvement (100%)	68.2	58.5	49.6	47.1	25.3	55.7	63.3	99.9	161.1

B. Contribution to Reconstruction

Figure 5 casts light on the correction effects on the reconstruction, and shows that the correction significantly improves the contrast and reduces noises evidenced by the reappearance of the structures which are immersed in noises before correction. The attenuation profiles shown in Fig. 6 of the reconstructions further suggest that accurate LACs are obtained after the projection data correction. Quantitative assessments with structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) metrics were also performed on the reconstruction slices, and the results are shown in Table I. From the table, SSIM and PSNR scores are significantly improved after correction, by $0.024 \sim 0.517$ for SSIM and by $25.3\% \sim 161.1\%$ for PSNR. Those excellent testing results strongly suggest the promising performance of deep learning application in PCD data correction.

C. Initial Real Result

The spectrum estimation described in II-D was conducted on a MARS scanner in our lab. The source was operated at $120kVp$ and $73\mu A$ with a $1.96mm$ Aluminum filter at the exit, and a $4mm$ Titanium filter was added to simulate human body attenuation. The spectrum was reconstructed using the expectation maximization (EM) method from transmission data of Aluminum blocks with different thicknesses. The measurements were captured with the Medipix-3 CZT PCD in the arbitration channel with active charge summing. The initial guess for the EM method is a generic tungsten anode X-ray tube spectrum. The estimated spectrum is shown in Fig. 7,

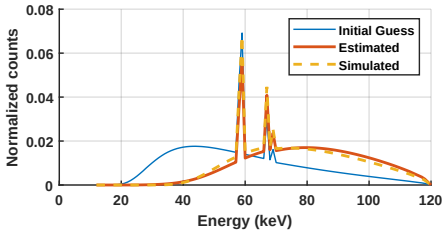


Fig. 7. X-ray spectrum estimated from the transmission data of aluminum blocks using the EM method.

along with the spectrum simulated using *SpekCalc*. The reconstructed spectrum shows a good agreement with the simulation, demonstrating the estimation validity. We are working along this direction in order to apply the proposed deep learning method for real PCD data correction.

IV. CONCLUSION

In conclusion, we have proposed a deep learning-based approach that promises to correct raw PCD data to the ideal data with spectral fidelity. As demonstrated with realistically synthesized PCD data based on our proposed detection model incorporating both PU and CS effects, the testing results suggest that the proposed approach achieves effective noise reduction and accurate spectral correction in both the projection and reconstruction domains. The spectrum shape can be faithfully recovered from significantly distorted measurement within $\pm 6\%$ relative errors in all bins at a 99.7% confidence. The SSIM and PSNR scores between the reconstruction and the ground truth can be improved greatly. In the near future, we plan to collect real PCD spectral projection data, implement the workflow outlined above, and evaluate the performance.

ACKNOWLEDGMENT

The authors would like to thank Drs. Norbert J. Pelc, Paul F. Fitzgerald and Mats Persson for valuable discussions and suggestions.

REFERENCES

[1] K. Taguchi, K. Stierstorfer, C. Polster, O. Lee, and S. Kappler, "Spatio-energetic cross-talk in photon counting detectors: Numerical detector model (pc tk) and workflow for ct image quality assessment," *Medical physics*, vol. 45, no. 5, pp. 1985–1998, 2018.

[2] E. Roessl, H. Daerr, and R. Proksa, "A fourier approach to pulse pile-up in photon-counting x-ray detectors," *Medical physics*, vol. 43, no. 3, pp. 1295–1298, 2016.

[3] H. Biermé and A. Desolneux, "A fourier approach for the level crossings of shot noise processes with jumps," *Journal of Applied Probability*, vol. 49, no. 1, pp. 100–113, 2012.

[4] M. Touch, D. P. Clark, W. Barber, and C. T. Badea, "A neural network-based method for spectral distortion correction in photon counting x-ray ct," *Physics in Medicine & Biology*, vol. 61, no. 16, p. 6132, 2016.

[5] G. Poludniowski, G. Landry, F. DeBlois, P. Evans, and F. Verhaegen, "Spekcalc: a program to calculate photon spectra from tungsten anode x-ray tubes," *Physics in Medicine & Biology*, vol. 54, no. 19, p. N433, 2009.

[6] I. ICRU, "Tissue substitutes in radiation dosimetry and measurement," *International Commission on Radiation Units and Measurements*, 1989.

[7] X. Duan, J. Wang, L. Yu, S. Leng, and C. H. McCollough, "Ct scanner x-ray spectrum estimation from transmission measurements," *Medical physics*, vol. 38, no. 2, pp. 993–997, 2011.

[8] M. Getzin, M. Li, D. S. Rundle, A. P. Butler, and G. Wang, "Non-uniformity correction for mars photon-counting detectors," in *15th Fully3D*, vol. 11072. SPIE, 2019, p. 110722W.

[9] H. Wang, M. Li, M. Getzin, B. L. Arduini, N. Covert, D. Vashishth, and P. Fitzgerald, "Deformable and reconfigurable mouse phantoms with 3d printing and lts technologies," in *Developments in X-Ray Tomography XII*, vol. 11113. SPIE, 2019, p. 111131S.

First Steps towards SiPM-based Scintillation Detectors for Photon-counting CT

Stefan J. van der Sar, Stefan E. Brunner, and Dennis R. Schaart

***Abstract* — Scintillation detectors may become an alternative for CdTe/CZT based photon-counting X-ray detectors with the possible advantages of better temporal stability and cost-effectiveness. In order to experimentally show the potential of this approach, a fast scintillation crystal (LuAP:Ce) was coupled to a fast, R&D stage silicon photomultiplier (SiPM). The detector generated pulses that lasted less than 100 ns in response to 60 keV gamma photons, such that incident rates exceeding 10 Mcps/pixels result in pulse trains with an acceptable level of pile-up. Moreover, a pulse shape model was developed and compared to the measurements. It was used to predict incident rate capability and low-rate energy resolution as a function of various scintillators, SiPM properties, and pixel sizes. The model showed that a 333 µm × 333 µm pixel detector of this particular R&D stage SiPM coupled to $LaBr_3$:Ce crystals with optimized light collection efficiency is expected to handle an incident rate of 100 Mcps/mm^2 with acceptable pile-up and to achieve a low-rate energy resolution about 10% at 60 keV. The results are a first indication that it may be feasible to develop SiPM-based scintillation detectors that can compete with CdTe/CZT based detectors for photon-counting CT.**

I. Introduction

PHOTON-COUNTING Computed Tomography (PCCT) is widely considered to be the next breakthrough in CT hardware. During the past ten years, various vision and review papers, such as [1] and [2], were published, which highlighted the potential benefits of replacing conventional energy-integrating detectors by photon-counting detectors. Examples of such benefits are improved image quality (contrast-to-noise ratio) and/or a reduction of contrast agent and radiation dose, truly simultaneous acquisition of dual-energy data with better spectral separation, and multi-energy data acquisition, which may enable K-edge imaging, for example. Since a typical incident rate on a CT detector is in the order of hundreds of Mcps/mm^2 and a reasonable energy resolution is required, semiconductor materials with a high density and a high atomic number, such as CdTe and CdZnTe (CZT), are generally considered to be the most promising detectors. A few prototype scanners based on such detectors have been installed to investigate the clinical benefits of the technology in practice [2], [3]. [4]. Current challenges seem to be related to the costs of the system and the temporal stability of the detection material. For these reasons, we are interested in investigating the feasibility of a different detection concept, i.e. a scintillation detector with silicon photomultiplier (SiPM) read-out as a potentially cost-effective and stable alternative. In this contribution, we study the raw pulses from such a detector, i.e. without assuming a particular way of pulse processing and counting. Two fundamental properties of the raw pulses are of interest: the pulse duration and the energy information. The former indicates what incident rates result in pulse trains with an acceptable level of pile-up, whereas the latter gives an idea about the achievable energy resolution. We show sub 100 ns pulses measured with a 1 mm × 1 mm SiPM-based scintillation detector, which implies that reasonable rates can be dealt with after pixel miniaturization. Moreover, we develop a model that predicts the raw pulse shapes for given detector properties, compare its predictions to measurements, and use it to evaluate directions for further development of this detector concept.

II. Methods and Materials

In a simplified description of the detector, the deposited X-ray energy is first converted into optical photons in a scintillator. Different scintillators vary in density, atomic number, light yield (#optical photons/keV), and decay time, which characterizes the exponential distribution of photon emission times. The optical photons are in turn converted into a current pulse by a photo-sensor. It is important that the wavelengths of the photons match the spectral sensitivity of the sensor [5]. The SiPM is a solid-state photo-sensor that consists of an array of single photon avalanche diodes (SPADs), which are small photodiodes biased above breakdown voltage. Consequently, they have a very high internal gain (~10^6) and produce a measurable current pulse in response to a single optical photon. When many SPADs detect an optical photon, e.g. after an X-ray interaction in the scintillator, the pulses add up to form the total output pulse. This pulse can be fed into a trans-impedance amplifier that converts it into a voltage pulse while maintaining the pulse shape. Because of the high internal gain, the use of charge-sensitive amplifiers and pulse shapers is unnecessary. The high internal gain (cf. photodiodes) and the possibility to miniaturize a pixel to less than 1 mm × 1 mm (cf.

This paper was submitted on 19 January 2020. The presented work was supported in part by Broadcom, Inc.

Stefan J. van der Sar is with Delft University of Technology, 2629JB, Delft, the Netherlands (e-mail: s.j.vandersar@tudelft.nl).

Stefan E. Brunner is with Broadcom, Inc., 93049, Regensburg, Germany (e-mail: stefan.brunner@broadcom.com).

Dennis R. Schaart is with Delft University of Technology, 2629JB, Delft, the Netherlands (e-mail: d.r.schaart@tudelft.nl).

photomultiplier tubes) currently make SiPMs the most promising photo-sensor for this application. An overview of SiPM working principles and technology can be found in [6].

A. Pulse shape model

The above-described pulse formation process was simulated with an in-house developed Matlab code. First, a deposited X-ray energy of 60 keV was converted into a number of optical photons based on the light yield. This number was multiplied by the efficiency of the light transport (light collection efficiency, LCE) and the photo-detection efficiency (PDE) of the SiPM. The outcome was used as the mean value of a Poisson distribution, from which a total of 2500 samples were drawn to simulate as many statistically fluctuating events.

For each event, an arrival time was assigned to all optical photons by sampling from an exponential distribution with a mean equal to the decay time of the scintillator [5]. Samples from a uniform distribution were drawn to assign the detected photon to a particular SPAD in the array. The SPAD response was modelled as a double exponential decay with a rise time of 1 ns, i.e. $0.5\left[\exp(-(t-1)/1)+\exp(-(t-1)/\tau_r)\right]$ with t the time since the detection of an optical photon and τ_r the recharge time constant of the SiPM, both in ns. The total output pulse was constructed by adding up the individual SPAD responses at their respective photon arrival times.

Two SiPM-specific effects were included in the model. Firstly, after detection of an optical photon, a SPAD has to recover to its initial state. During recovery, the SPAD can detect another photon, but the amplitude of its response is reduced by a factor $1-\exp(\Delta t/\tau_r)$ with Δt the time between the moments of detection of both photons. This leads to a reduced sensor response, also called saturation, in case many optical photons hit the SiPM in a short period of time. Secondly, SiPMs suffer from correlated noise in the form of optical crosstalk. This was implemented as the geometric chain process described by Vinogradov [7]. Dark pulses were neglected, because their rate is typically low compared to the required pulse duration.

For all statistically varying pulses, the time window t_{99} that contains 99% of the pulse integral was determined. The mean t_{99} was calculated and regarded as the pulse duration. Its inverse, r_{99}, was considered to be an incident rate that would give rise to pulse trains with an acceptable level of pulse pile-up. Given the exponential distribution of inter-arrival times [5], 'an acceptable level' means a 63% chance that a pulse arrives within t_{99} after the previous one. The value of the pulse integral itself is the fundamental measure of the energy information in the pulse. A low-rate energy resolution R was calculated by dividing the full width at half maximum $FWHM$ of the pulse integral distribution by the mean pulse integral PI_0 [5].

B. Selection of scintillators and SiPM

Three scintillators were selected for this study: LYSO:Ce, $LaBr_3$:Ce, and LuAP:Ce. Their properties are listed in Table 1. LYSO:Ce is often used in clinical PET scanners and could also be useful for PCCT. However, PCCT imposes more severe requirements on the rate capability and energy resolution than PET, so the faster and brighter $LaBr_3$:Ce was included, as well. A drawback of $LaBr_3$:Ce is its hygroscopic nature. Therefore, LuAP:Ce was selected as a non-hygroscopic, but comparably fast alternative. It furthermore has a very high density, but its disadvantage is a relatively low light yield. The performance of these scintillators coupled with three different LCE values to SiPMs of various dimensions, all with a SPAD pitch of 15 µm and biased at 7 V over breakdown, was investigated. The properties of these SiPMs are shown in Table II in bold. For comparison, properties of the same SiPMs at 3 V and properties of a typical Broadcom NUV-HD SiPM with 30 µm SPAD pitch are also listed in Table II.

TABLE I: OVERVIEW OF SCINTILLATOR PROPERTIES

Scintillator	$LaBr_3$:Ce [8]	LuAP:Ce [9]	LYSO:Ce [8]
Light yield (ph/keV)	63	15	33
Wavelength (nm)	380	375	420
Decay time (ns)	16	17	36
Density (g cm^{-3})	5.1	8.3	7.1
CT thickness (mm)	3.5	1.0	1.5
Hygroscopic	Yes	No	No

TABLE II: OVERVIEW OF SIPM PROPERTIES

SiPM	Over voltage 3 V	**7 V**		Over voltage 3 V	7 V
SPAD pitch (µm)	**15**			30	
Crosstalk probability	0.03	**0.10**		0.09	0.29
Recharge time constant (ns)	12	**10**		55	50
Photo detection efficiency	0.20	**0.30**	At relevant wavelengths	0.41	0.48

C. Experiments

Experiments were done with a 1 mm × 1 mm R&D stage SiPM from Broadcom, Inc., which is not commercially available. However, the SiPM is based on the company's NUV-HD technology and only differs in size and SPAD pitch from commercial products. Its properties are described in bold in Table II. The sensor was coupled to a 0.9 mm × 0.9 mm × 1.0 mm as-cut LuAP:Ce crystal (Hilger Crystals) using UV curing glue and covered in Teflon powder to increase the LCE. The detector was exposed to 60 keV gamma photons from a 35 MBq Am-241 source. Pulses were processed by Broadcom's AFBR-S4E001 preamplifier board and fed into an oscilloscope (Teledyne LeCroy HDO9404). Ten 100 ms long traces were recorded at a sampling rate of 1 GS/s and a bandwidth of 200 MHz. An algorithm selected all pulses that crossed a threshold set to half the mean pulse height (i.e. 30 keV, a commonly applied threshold in medical PCCT [4]). The selected pulses were used to determine r_{99} and the energy resolution.

III. RESULTS AND DISCUSSION

In part A of this section, the experimental results are presented and discussed, whereas part B is about the model predictions. Part C discusses the predictive power of the model by comparing its predictions with the measurements.

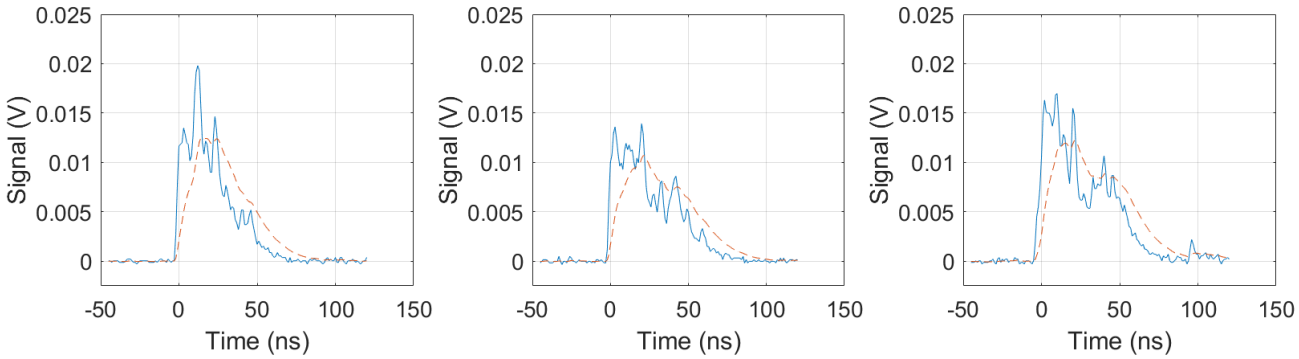


Fig. 1. Three examples of measured pulses. The pulses return to baseline within 100 ns, but contain many spikes, which are signs of the arrival of individual photons. They are visible due to a small number of detected photons. This can be improved by increasing the LCE or by applying a (digital) low-pass filter, such as a first order RC filter with a time constant of 10 ns (dashed curves), which leads to limited pulse elongation.

A. *Experiments*

Three examples of pulses measured with the LuAP:Ce detector are shown in Fig. 1. It can be seen that the pulses return to the baseline within 100 ns. Hence, an r_{99} of more than 10 Mcps/pixel, e.g. 11.11 Mcps/pixel, can be achieved. It follows that a pixel size of 333 µm × 333 µm (9 pixels/mm^2) would suffice to achieve an r_{99} of 100 Mcps/mm^2. As explained in section II A, about 63% of pulses piles up on the previous pulse at this rate. However, this leads to very small pulse distortion, if a pulse is superimposed on the end of the exponentially decaying tail of the previous pulse. Indeed, the median of the exponential distribution of inter-arrival times for a rate of 11.11 Mcps/pixel says that 50% of the pulses arrives within 62.4 ns after the previous one. Fig. 1 shows that the remaining 50% arriving after 62.4 ns will not or hardly be distorted by pile-up. We note that this analysis may be different for Gaussian shaped pulses as produced by CdTe / CZT-based detectors.

The pulses in Fig. 1 feature spikes, which are signs of the arrival of individual photons. Their visibility indicates that a small number of photons was detected. A model by Van Dam et al. [10] was used to estimate this number and yielded on average 42 detected photons per pulse, which comes down to an LCE of 0.15. Reasons for this low value may include a misalignment of crystal and sensor, insufficient compression of the Teflon powder, and light losses via the glue and the 0.2 mm thick glass layer that covers the SiPM. One could optimize these factors or apply a (digital) filter, e.g. a first-order RC filter with RC=10 ns, to reduce the spikes without considerable pulse elongation as shown in Fig. 1 [5]. Nevertheless, the small number of detected photons yielded a low-rate energy resolution of only 43.5% at 60 keV in this experiment.

B. *Model predictions*

The model was used to predict the energy resolution at 60 keV for higher LCE values and for other scintillators that have a higher light yield than LuAP:Ce. The results are shown in Table III for a large pixel size of 1000 µm × 1000 µm. An energy resolution below 10% can only be achieved with the highest light yield scintillator, $LaBr_3$:Ce, and high LCE values. Values around 12% may be achieved with LYSO:Ce, whereas LuAP:Ce can achieve 17% at best. For smaller pixel sizes, SiPM saturation may start to degrade the energy resolution to some extent. However, we noticed that this subtle effect was not correctly incorporated in the current version of the model. We work on an update of the model and hope to present the results at the conference. Further improvements in the energy resolution may be achieved by increasing the PDE, but Table II indicates that SiPMs with higher PDE are generally slower and have a larger SPAD pitch, which reduces the incident rate capability and increases saturation. Such trade-offs can be optimized using the model presented in this work.

The calculated values of r_{99} are shown in Fig. 2. It turned out that r_{99} hardly depends on LCE. Furthermore, $LaBr_3$:Ce and LuAP:Ce have almost the same decay times, so only one curve was drawn for both of them for better visualization. The data show that an r_{99} of 100 Mcps/mm^2 can be achieved with 333 µm × 333 µm pixels of $LaBr_3$:Ce or LuAP:Ce and with 225 µm × 225 µm pixels of LYSO:Ce.

TABLE III: PREDICTED ENERGY RESOLUTIONS AT 60 KEV (1 X 1 MM^2 PIXELS)

Scintillator	LuAP:Ce	LYSO:Ce	$LaBr_3$:Ce
LCE = 0.25	30%	20%	15%
LCE = 0.50	21%	15%	10%
LCE = 0.75	17%	12%	8.3%

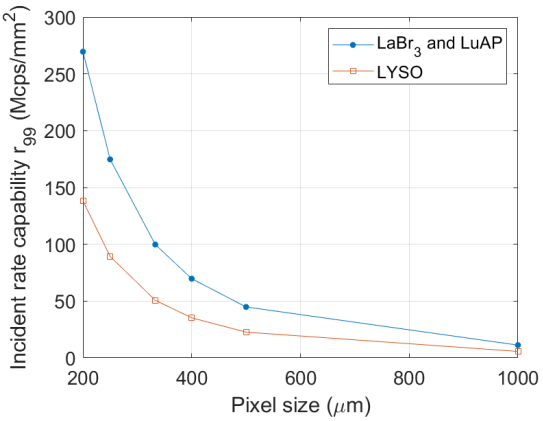


Fig. 2. Incident rate capability r_{99} as a function of pixel size for three different scintillators coupled to the SiPM. Since $LaBr_3$ and LuAP yielded similar results, only one curve was drawn for better visualization.

C. Comparison and model limitations

Here, we briefly discuss the predictive power of the model by comparing the measured and simulated pulse shapes and energy resolutions. To this end, the model was run under the experimental conditions (LuAP:Ce, LCE of 0.15, pixel size of 1 mm × 1 mm). Fig. 3 shows examples of simulated pulse shapes. Compared to the measured pulses in Fig. 1, similar rise time, spikes, and return to baseline within 100 ns can be observed, but without noise. Moreover, the model predicts an energy resolution of 38.5% at 60 keV, whereas the measured value is 43.5%. Such differences may be due to a non-proportional light yield as explained in [11]. Especially LYSO:Ce suffers from this effect, so the energy resolutions presented in Table III for this scintillator will be considerably worse in practice. On the other hand, $LaBr_3$:Ce and LuAP:Ce are known as proportional scintillators [11]. Overlap of K- and L-escape peaks with the photo peak can also degrade the energy resolution [5]. This becomes more important as the pixel size decreases. Both effects can be studied by Monte Carlo particle tracking. Any remaining mismatch may be due to slight inhomogeneity in scintillator properties and light transport.

D. Another important consideration

The pixels of a scintillation detector must be optically isolated to prevent light spread to other pixels. The use of conventional reflective foils may result in unacceptably large dead areas and associated dose efficiency losses for miniaturized pixels, so innovative techniques must be developed. Possibilities that are under investigation by others include 3D printed reflectors, crystals with a columnar microstructure [12], and laser-induced optical barriers [13].

IV. Conclusions

We have experimentally shown that a LuAP:Ce crystal coupled to an R&D stage SiPM generates pulses that return to baseline within 100 ns, such that the detector should generate pulse trains with acceptable pile-up for incident rates of more than 10 Mcps/pixel. The measured low-rate energy resolution of 43.5% at 60 keV is still suboptimal due to a low light collection efficiency (LCE). However, the pulse shape model introduced in this work predicts that incident rates of 100 Mcps/mm^2 will lead to acceptable pile-up levels and that a low-rate energy resolution of about 10% at 60 keV can be achieved, if the detector is made of $LaBr_3$:Ce, the LCE is optimized, and the pixel size is miniaturized to 333 µm × 333 µm. These numbers suggest that SiPM-based scintillation detectors may be a competitive alternative to CdTe/CZT based detectors for PCCT. Therefore, the above-mentioned improvements will be pursued in future experimental work. Furthermore, the effects of pile-up and various counting methods on the observed count rate and energy resolution for the characteristic pulse shape generated by an SiPM-based scintillation detector will be studied in more detail, both by simulation and experiment.

References

[1] K. Taguchi *et al.*, "Vision 20/20: Single photon counting X-ray detectors in medical imaging," *Med. Phys.*, vol. 40, no. 10, 100901, Oct. 2013.

[2] M.J. Willemink *et al.*, "Photon-counting CT: Technical principles and clinical prospects," *Radiology*, vol. 289, no. 2, pp. 293-312, Nov. 2018.

[3] S. Kappler *et al.*, "Photon counting CT at elevated X-ray tube currents: Contrast stability, image noise and multi-energy performance", *Proc. SPIE*, vol. 9033, 90331C, Mar. 2014.

[4] R. Steadman *et al.*, "ChromAIX2: A large area, high count-rate, energy-resolving photon counting ASIC for a spectral CT prototype, *Nucl. Instr. Meth. Phys. Res. A, vol.* 862, pp. 18-24, Aug. 2017.

[5] G.F. Knoll, *Radiation detection and measurement*, 4th ed., Hoboken, NJ, United States of America: John Wiley & Sons, 2010, ch. 2, 4, 8, 9, 10, 16, 17.

[6] C. Piemonte *et al*., "Overview of the main parameters and technology of modern silicon photomultipliers", *Nucl. Instr. Meth. Phys. Res. A.*, vol. 926, pp. 2-15, Nov. 2018.

[7] S. Vinogradov, "Analytical models of probability distribution and excess noise factor of solid state photomultiplier signal with crosstalk," *Nucl. Instr. Meth. Phys. Res. A, vol.* 69, pp. 247-251, Sep. 2011.

[8] Saint Gobain Crystals, 'Crystal Scintillation Materials', 2019. Online available: https://www.crystals.saint-gobain.com/products/crystal-scintillation-materials. Accessed: 7 Jan. 2020.

[9] Hilger Crystals, private communication, Feb. 2019.

[10] H.T. van Dam *et al.*, "A comprehensive model of the response of silicon photomultipliers," *IEEE Trans. Nucl. Sci*., vol. 57, no. 4, pp. 2254-2266, Sep. 2010.

[11] I.V. Khodyuk *et al.,* "Trends and patterns of scintillator nonproportionality," *IEEE Trans. Nucl. Sci.*, vol. 59, no. 60, pp. 3320-3331, Dec. 2012.

[12] H.B. Bandhari *et al*., "Large-area crystalline microcolumnar $LaBr_3$:Ce for high-resolution gamma ray imaging," *IEEE Trans. Nucl. Sci*., vol. 60, no. 1, pp. 3-8, Feb. 2013.

[13] L. Bläckberg *et al.*, "Exploring light confinement in laser-processed LYSO:Ce for photon counting CT application," *Phys. Med. Biol.*, vol. 64, no. 9, 095020 , Apr. 2019.

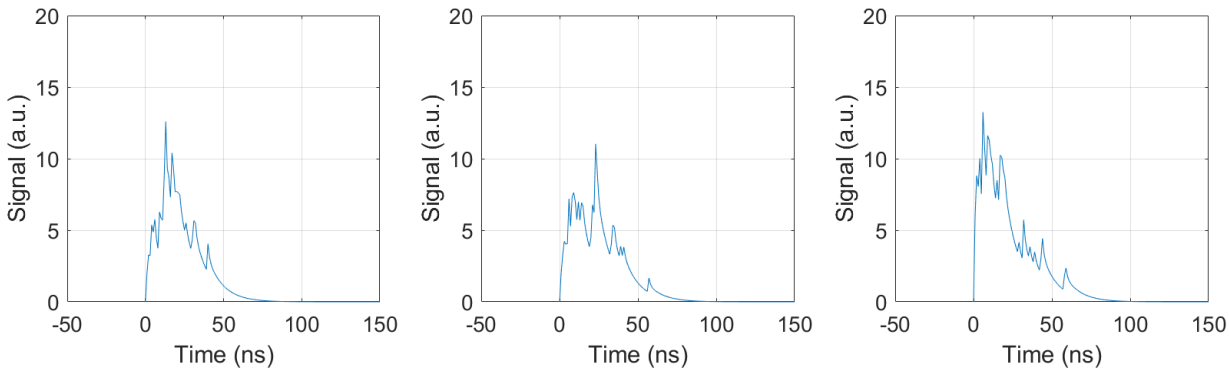


Fig. 3. Three examples of simulated pulses shapes obtained with simulations settings comparable to the experimental conditions.

Sparse-view Cone-beam Breast CT via Conditional Generative Adversarial Network Constrained by Image Edges

Yunhuan Yang, Chengyijue Fang, and Lei Zhu*

***Abstract*—Improving over mammography and tomosynthesis, cone-beam breast CT (CBBCT) provides complete 3D information with high spatial resolution for detection of calcifications and small lesions. However, increased radiation dose and scan time hinder clinical applications of CBBCT, especially in early cancer screening. We propose sparse-view CBBCT, i.e. with a small number of projections, to overcome the above bottlenecks. A deep learning method – conditional generative adversarial network (cGAN) constrained by image edges (ECGAN) – is proposed to suppress artifacts on sparse-view CBBCT images reconstructed by filtered backprojection (FBP). We use cGAN as the basic frame to avoid over-smoothing in some degree, with a modified U-net as the generator. To further preserve subtle structures and micro calcifications which are particularly important for breast cancer screening and diagnosis, edge images of CBBCT are extracted and added in cGAN to guide the learning. The proposed algorithm has been evaluated on 20 clinical datasets of CBBCT. ECGAN substantially improves the image qualities of sparse-view CBBCT, with a performance superior to those of total-variation (TV) based iterative reconstruction and U-net based post-processing. On one CBBCT case with the projection number reduced from 300 to 100, ECGAN enhances peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) on FBP reconstruction from 24.26 and 0.812 to 37.78 and 0.963, respectively. These results indicate that ECGAN successfully reduces radiation dose and scan time of CBBCT by a factor of at least 3 with only small image degradations.**

***Index Terms*—cone beam breast CT, deep learning, sparse-view reconstruction, breast cancer screening, conditional GAN**

I. Introduction

Recently cone beam breast CT (CBBCT) receives heated attention in early breast cancer screening. Compared with the two dominant methods in current clinical practice, including mammography and tomosynthsis, CBBCT requires no compression of breast and significantly improves image qualities with complete 3D information, high spatial resolution and better patient comfort. Nevertheless, due to increased data acquisition, radiation dose and long scan time remain as two of last hurdles toward wide adoption of CBBCT in hospitals and clinics. In this work, we reduce projections of CBBCT to lower radiation dose and shorten scan time by a factor of 3-4, and propose a new deep learning algorithm on sparse-view CBBCT to improve the image qualities to a level matching that of full-view CBBCT.

One barrier limiting the clinical applications of CBBCT is the concern of radiation dose. Women breasts are highly sensitive to radiation. American FDA sets 6 mGy [1] as the dose limit of mammography for each examination in cancer screening. An average glandular dose of 4-12.8 mGy on CBBCT has been reported in the existing literature [2]. Dose reduction on current CBBCT is therefore suggested to make exams safe under the 6 mGy limit. Compared with mammography and tomosynthesis, another drawback of CBBCT is the increased scan time from less than 6 s (mammography)[3] or 3.7 s (tomosynthesis)[4] to 10s[2], which raises the possibility of motion artifacts. One viable solution to the above two problems is sparse-view CBBCT, i.e. CBBCT with a reduced projection number.

Sparse-view CT has been extensively studied in the past decade[5-8]. It is well known that conventional filtered-backprojection (FBP) reconstruction generates view-aliasing artifacts on sparse-view CT. One popular solution is iterative reconstruction, including regularization based algorithms [8, 9] or compressed-sensing based optimization[6, 10]. Recently, researchers have proposed a series of deep-learning algorithms for sparse-view CT, including signal recoveries in both projection[11] and image domain[12-14], or deep CT reconstruction[15] .

Despite superior image qualities of published results, we find that direct implementations of existing algorithms of iterative reconstruction or deep learning on sparse-view CBBCT result in significant loss of small anatomical structures, as shown later in the section of results. The main reason is that full-view CBBCT images contain ample details of glandular structures with low contrast (<150 HU) and high spatial resolution (<0.5 mm). Preservation of these details is challenging but particularly important for early breast cancer screening.

In this work, we propose a new algorithm of conditional GAN (cGAN) [16]constrained by image edges (ECGAN) as a post-processing procedure on sparse-view CBBCT images reconstructed by the standard FBP. cGAN is used as the basic frame, with a modified U-net[17] as the generator. The reason for choosing the cGAN is that the adversarial loss can avoid over-smoothing caused by MSE loss or other similar loss in some degree. Batch normalization is included in U-net to improve training. Convolutional down-sampling and deconvolution replace the max-pooling and up-sampling in the original U-net for better preservation of anatomical details of CBBCT images. Furthermore, to retain small calcifications

Research reported in this publication was supported by the National Natural Science Foundation of China (No. 81671681), the Ministry of Science and Technology of China Key Research and Development Projects (Grant No. 2016YFC0101400) and by the Fundamental Research Funds for the Central Universities (No. WK2030040089.

Yunhuan Yang and Chengyijue Fang are with the Department of Engineering and Applied Physics, University of Science and Technology of China, Hefei, China (e-mails: yyhyyh@mail.ustc.edu.cn, fcyj2013@mail.ustc.edu.cn).

*Corresponding author, Lei Zhu, is with the Department of Engineering and Applied Physics, University of Science and Technology of China, Hefei, China (e-mail: leizhusg@ustc.edu.cn).

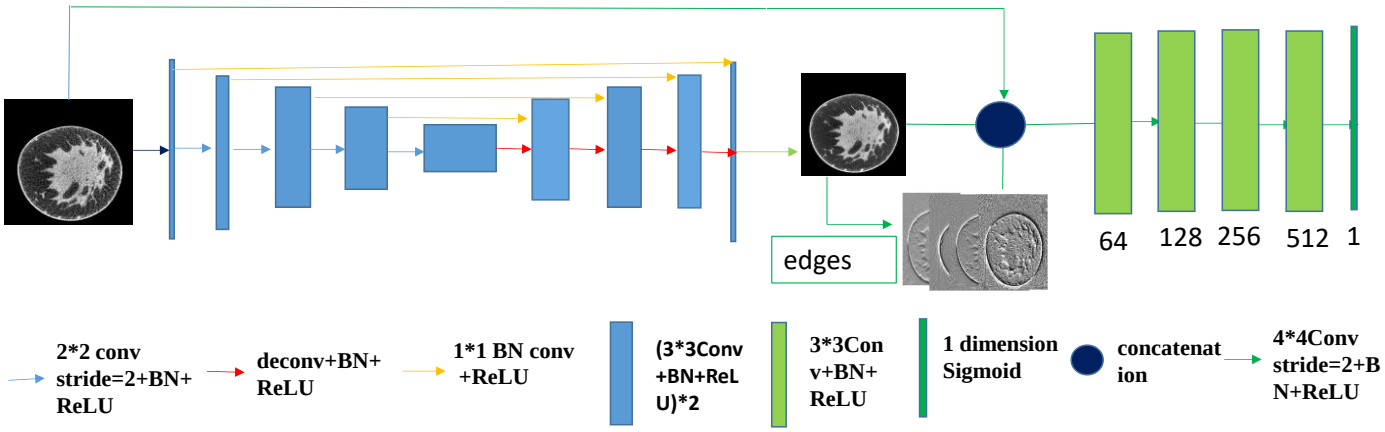


Figure 1 The architecture of the proposed ECGAN.

and potential lesions in a low-contrast background, edges of CBBCT images are feed to the generator and discriminator. The proposed ECGAN has been evaluated using 20 clinical datasets, with focuses on performance comparsions on sparse-view CBBCT with those of total-variation (TV) based iterative reconstruction and U-net based post-processing. Then we further investigate the efficacy of the adversarial loss and edge images.

II. Method

Sparse-view CBBCT via FBP reconstruction suffers severe view-aliasing artifacts. We aim to design a neural network to remove artifacts of sparse-view CBBCT with minimal loss of image details. Let X represent the sparse-view CBBCT images with artifacts and Y be ideal CBBCT images. Our goal is to find a function f to map X to Y: Y = f(X) via neural network. We find the proposed ECGAN, shown in Figure 1,effectively obtains f with the X as input, corresponding Y as a label. The traditional U-net with a loss function of mean-square-error (MSE) causes over-smoothing on resultant images. Therefore we choose cGAN as the framework while keeping the modified U-net as its generator. The adversarial loss plays a supervisor role of forcing the processed image as close as possible to the ground truth and avoiding over-smoothing.

cGAN is a framework for estimating generative models via adversarial process, in which two models was simultaneously trained: a generator network G that maps the sparse-view images to ideal images $\hat{Y}$, and discriminator network D that estimates the probability that an image come from the label data rather than G. To ensure the one-to-one correspondence, the $\hat{Y}$ the X should be combined to be input discriminator.$G^* = \arg min_G max_D \mathrm{L}_{\mathrm{cGAN}}(G, D)$.

$$\mathcal{L}_{\mathrm{cGAN}} = E_{x,y}[logD(x,y)] + E_x[\log(1 - D(x, G(\mathrm{x}))]$$

It is beneficial to mix the GAN objective with a traditional loss, such as L2 distance.

$$\mathcal{L}_2(\mathrm{G}) = \mathrm{E}_{\mathrm{x,y}}[||\mathrm{y} - \mathrm{G}(\mathrm{x})||_2]$$

$$G^* = \arg min_G max_D \mathcal{L}_{\mathrm{cGAN}}(G, D) + \lambda \mathcal{L}_{\mathrm{L1}}(G)$$

We choose the U-net as the generator, which is widely used in processing of biomedical images whose structure including skip connection and encoder-decoder is very useful to deliver the anatomical details. To further improve the ability, the rough max-pooling layers are replaced with intelligent convolution filters with stride equaling two which try to learn the down-pooling process. We replace the up-sampling layers with deconvolution to avoid potential blur. For converging quickly and preventing gradient vanishing or diverging, we add batch normalization.

As for the discriminator, we expect it to model high-frequency information which requires restricting our attention to local patches. The discriminator tries to judge if N*N patch in an image is real or fake. Different from traditional GAN, the probability outputted by discriminator is not a vector but a matrix. The discriminator architecture is plain as follows

$$\mathrm{C64} - \mathrm{C128} - \mathrm{C256} - \mathrm{C512}$$

where C represents convolution, batch normalization and ReLU activation function. The numbers mean the amount of the feature maps. At the last layer, the data are mapped to 1-dimension then feed to a Sigmoid function.

To further retain small calcifications and potential lesions in the CBBCT low-contrast background, the edge images of the processed images $\hat{Y}$ and ground truth Y are calculated. The edge information of the images is gained from convolutional layer which is the Sobel operator SF applied in the horizontal, vertical and the two diagonal directions. The four edge images are feed to the G and D. In generator they add a L2 term:

$$\mathcal{L}_{\mathrm{E}}(\mathrm{G}) = \mathrm{E}_{\mathrm{x,y}}[\|y - G(x)\|_2 + \lambda_1 \|SF(y) - SF(G(x))\|_2]$$

In discriminator they are feed as four channels:

$$\mathcal{L}_{\mathrm{gcGAN}} = E_{x,y}[logD(x,y,SF(y)] + E_x[\log(1 - D(x, G(\mathrm{x}), \mathrm{SF}(\mathrm{G}(\mathrm{x}))]$$

In conclusion, our objective function can be written as below:

$$G^* = \arg \min_{\mathrm{G}} \max_{\mathrm{D}} \mathcal{L}_{\mathrm{gcGAN}}(G, D) + \lambda_g \mathcal{L}_{\mathrm{E}}(G)$$

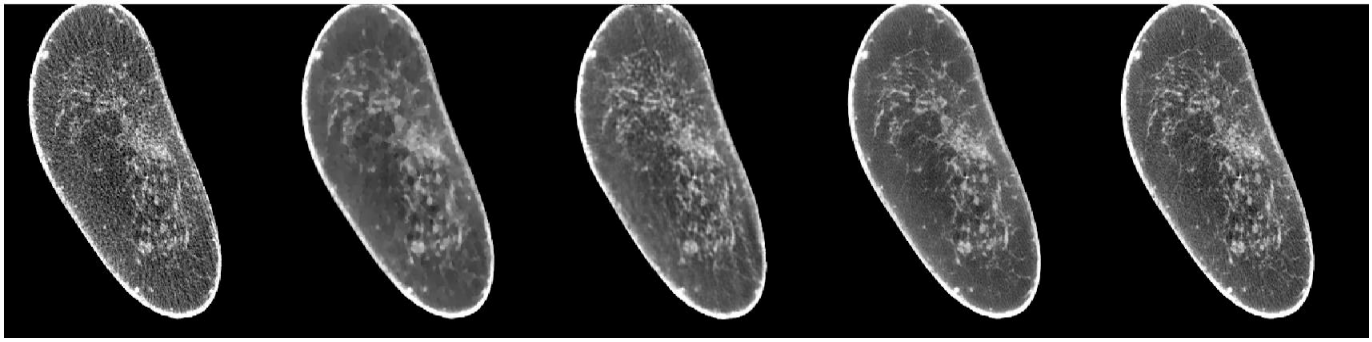


Figure 2 One representative case of CBBCT images using different algorithms. From left to right are results of: sparse-view CBBCT, sparse-view CBBCT with TV-based iterative reconstruction, sparse-view CBBCT with U-net, sparse-view CBBCT with the proposed ECGAN, and full-view CBBCT. Sparse-view and full-view CBBCT are FBP reconstructions using 100 and 300 projections, respectively. Display window: [0.2,0.3] cm^{-1}.

III. Results

A. Experiment details

In evaluation studies, we use 20 datasets of clinical CBBCT scans. Each dataset contains 300 projections over 360 degrees and one FBP 3D reconstruction (full-view CBBCT). Sparse-view CBBCT is generated by FBP reconstruction on downsampled 100 or 75 projections, which corresponds to a reduction of radiation dose and scan time by a factor of 3 or 4. A total of 5900 2D image pairs of full-view and sparse-view CBBCT is used in our studies. 4720 images are used to train ECGAN and the rest is used to test. To optimize the network, we follow the guidance from Ref. [18], i.e., one gradient step for D, then one step on G and train to maximize $D(x, G(\mathrm{x}), \mathrm{SF}(\mathrm{G}(\mathrm{x}))$. The Adam solver is used with a learning rate of 0.0001. We empirically set the batch size to 2, λ_g and λ_1 to 60. The network is implemented in Pytorch on a single Tesla K80 GPU.

B. Results and comparison study

To investigate the efficacy of the proposed ECGAN on sparse-view CBBCT, we compare the algorithm performance with those of TV-based iterative reconstruction [6] and U-net based deep learning image processing [17]. Figure 2 shows the results of sparse-view CBBCT using 100 projections. It is seen that ECGAN outperforms the other two competing algorithms on preservation of small image details with less artifacts. The conclusion by visual inspection is consistent with quantitative analysis listed in Table 1. Peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM)[19] are taken as quantitative metrics. Full-view CBBCT is used as the reference in SSIM calculations. Sparse-view CBBCT with the proposed ECGAN obtains the highest scores on both PSNR and SSIM. Although not shown in the paper, the algorithm performances are similar for sparse-view CBBCT with 75 projections.

To fully demonstrate the advantages of the proposed ECGAN, we further evaluate the effect of the adversarial loss and edge information via training three different networks, including modified U-net with MSE as loss function (MUnet-MSE), cGAN with modified U-net as generator and combined with L2 distance (MCGAN), and the proposed ECGAN. The representative results are shown in Figure 3. It is seen that all three networks reduce the artifacts of sparse-view CBBCT greatly even with 75 projections. MCGAN and ECGAN outperform MUnet-MSE by avoiding image over-smoothing, which indicates the adversarial learning indeed preserves high frequency information. The results of ECGAN contain most subtle structures which are blurred or completely missing in the results of MUnet-MSE and MCGAN. The above fact is better observed in the comparison of zoom-in images shown in Figure 4. ECGAN obtains the highest contrast of the suspicious micro calcification indicated by the rectangle in Figure 4 on sparse-view CBBCT images.

Table 1 PSNR and SSIM of images shown in Figure 2, with full-view CBBCT as the reference.

	PSNR	SSIM
Sparse-view CBBCT	24.26	0.812
Sparse-view CBBCT with TV	34.26	0.925
Sparse-view CBBCT with Unet	33.55	0.911
Sparse-view CBBCT with ECGAN	**37.78**	**0.963**

IV. Conclusions and discussion

In this study, we propose a deep learning method –ECGAN to improve the quality of sparse-view CBBCT. The network is specially designed for avoiding the over-smoothing problem and retaining subtle structure. Therefore adversarial loss combined with edge image is adopted. Both visual inspections and quantitative analysis on clinical CBBCT images show that ECGAN outperforms TV-based iterative reconstruction and Unet-based image processing with better image qualities of sparse-view CBBCT. Evaluation studies are specially designed to validate the effects of two key components of ECGAN. Our research effectively reduces the required imaging dose and scan time of CBBCT, and therefore facilitates the clinical adoptions of CBBCT for early breast cancer screening and diagnosis.

In future work, we will combine ECGAN with deep reconstruction to further improve the image qualities of sparse-view CBBCT. The proposed algorithms will be finally implemented for 3D imaging.

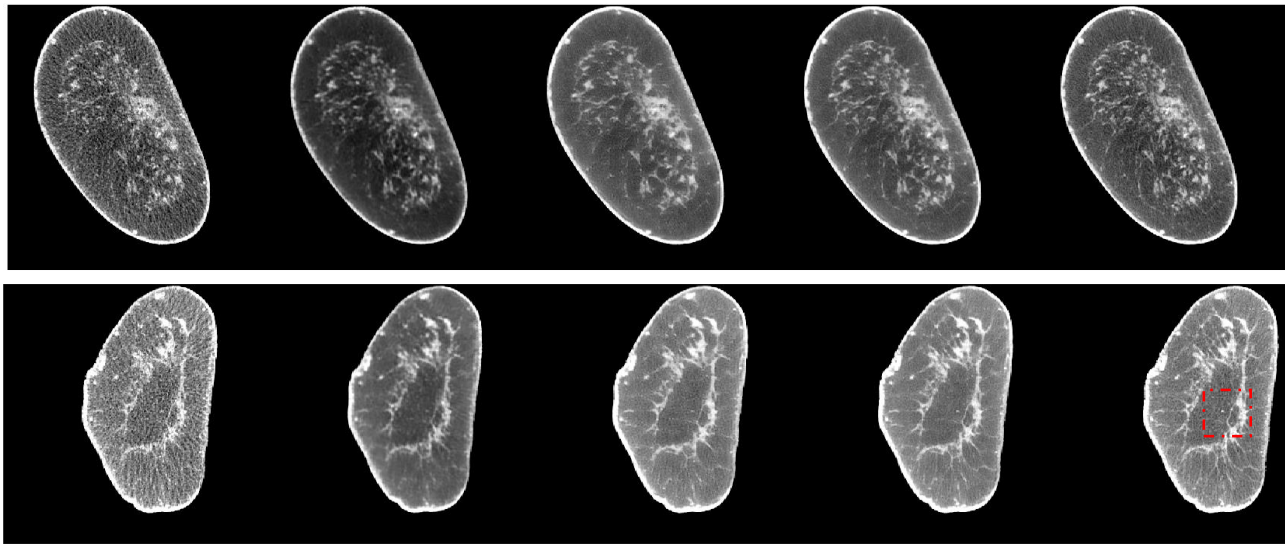


Figure 3 CBBCT images reconstructed from 100 (upper row) and 75 (bottom row) projections and processed by different algorithms, except the last column, which are FBP reconstructions using 300 projections. The first four columns are sparse-view CBBCT results using (from left to right) no post-processing, MUnet-MSE, MCGAN, and ECGAN, respectively. Display window: [0.2,0.3] cm^{-1}.

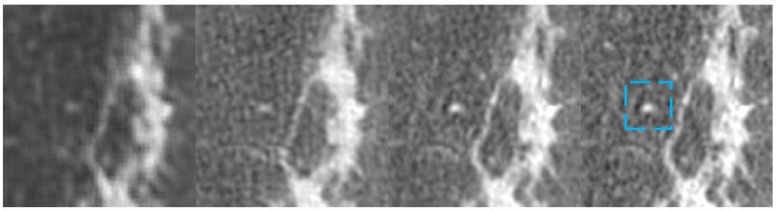


Figure 4 Zoom-in images of the region marked by the red box in Figure 3. From left to right are the sparse-view CBBCT (with 75 projections) results using MUnet-MSE, MCGAN, the proposed ECGAN and the full-view CBBCT image (with 300 projections). The contrast of the white point in the blue box are 0.025, 0.039, 0.063, and 0.081 cm^{-1}. Display window: [0.2,0.3] cm^{-1}.

REFERENCES

1. Tornai, M.P., et al. Design and development of a fully 3D dedicated x-ray computed mammotomography system. in Medical Imaging 2005*: Physics of Medical Imaging*. 2005. International Society for Optics and Photonics.
2. O'Connell, A., et al., Cone-beam CT for breast imaging: Radiation dose, breast coverage, and image quality. *American Journal of Roentgenology*, 2010. **195**(2): p. 496-509.
3. Tesic, M., M.F. Piccaro, and B. Munier, Full field digital mammography scanner. *European journal of radiology*, 1999. **31**(1): p. 2-17.
4. Sechopoulos, I., A review of breast tomosynthesis. Part I. The image acquisition process. *Medical physics*, 2013. **40**(1): p. 014301.
5. Lu, Y., J. Zhao, and G. Wang, Few-view image reconstruction with dual dictionaries. *Physics in Medicine & Biology*, 2011. **57**(1): p. 173.
6. Sidky, E.Y. and X. Pan, Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 2008. **53**(17): p. 4777.
7. Bian, J., et al., Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT. *Physics in Medicine & Biology*, 2010. **55**(22): p. 6575.
8. Niu, S., et al., Sparse-view x-ray CT reconstruction via total generalized variation regularization. *Physics in Medicine & Biology*, 2014. **59**(12): p. 2997.
9. Beister, M., D. Kolditz, and W.A. Kalender, Iterative reconstruction methods in X-ray CT. *Physica medica*, 2012. **28**(2): p. 94-108.
10. Sidky, E.Y., C.-M. Kao, and X. Pan, Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-ray Science and Technology*, 2006. **14**(2): p. 119-139.
11. Styner, M.A., et al., View-interpolation of sparsely sampled sinogram using convolutional neural network. 2017. **10133**: p. 1013328.
12. *Chen, H., et al.*, Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging, 2017*. ***36****(12): p. 2524-2535.*
13. *Jin, K.H., et al.*, Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing, 2017*. ***26****(9): p. 4509-4522.*
14. Cong, W., et al. Deep-learning-based breast CT for radiation dose reduction. in *Developments in X-Ray Tomography XII*. 2019. International Society for Optics and Photonics.
15. Li, Y., et al., Learning to reconstruct computed tomography (CT) images directly from sinogram data under a variety of data acquisition conditions. *IEEE transactions on medical imaging, 2019.*
16. Isola, P., et al. Image-to-image translation with conditional adversarial networks. in *Proceedings of the* IEEE conference on computer vision and pattern recognition. 2017.
17. Ronneberger, O., P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *in International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
18. Goodfellow, I., et al. Generative adversarial nets. in *Advances in neural information processing systems*. 2014.
19. Hore, A. and D. Ziou. Image quality metrics: PSNR vs. SSIM. in *2010 20th International Conference on Pattern Recognition*. 2010. IEEE.

Deep Learning-based Initialization of Iterative Reconstruction for Breast Tomosynthesis

Koen Michielsen, Nikita Moriakov, Jonas Teuwen, and Ioannis Sechopoulos

***Abstract*—Reconstruction of digital breast tomosynthesis is a challenging problem due to the limited angle data available in such systems. Due to memory limitations, deep learning-based methods can help improve these reconstructions, but can not (yet) attain sufficiently high resolution. In addition to this practical issue, questions remain on the possibility of such models introducing 'ghost' information from the training data that is not compatible with the projection data. To take advantage of some of the benefits of deep learning-based reconstructions while avoiding these limitations, we propose to use the low resolution deep learning-based reconstruction as an initialization of a regular high resolution iterative method.**

The network was trained using digital phantoms, some based on a mathematical model and some derived from patient dedicated breast CT scans. The output of this network was then used as initialization for 10 000 iterations of MLTR for nine patient based phantoms that were not included in the training. The same nine cases were also reconstructed without any initialization for comparison.

The reconstructions including initialization were found to reach a lower mean squared error than those without, and visual inspection found much improved retrieval of the breast outline and depiction of the skin, confirming that adding the deep learning-based initialization adds valuable information to the reconstruction.

I. Introduction

Digital breast tomosynthesis (DBT) is a limited angle tomography modality dedicated to breast imaging. It improves on digital mammography by providing a pseudo-3D reconstruction with high in-plane resolution, but low resolution in the third dimension. This lower resolution is caused by the incomplete angular sampling, resulting in missing data in the Fourier domain. This also means that the reconstruction problem is under-determined, and that the result of an iterative reconstruction will depend on its starting point. Therefore, a good initialization not only helps increase convergence speed, but, in DBT, also improves the reconstruction itself [1].

Recently, we developed a deep learning-based reconstruction for a simplified 2D version of the DBT limited angle problem [2]. With it, we were able to reconstruct the distribution of the glandular tissue in the breast to an accuracy that allowed for the estimation of the individual patient dose to within 20% and of the total glandular tissue to within 3%. Extending this work to 3D is a significant challenge due to the difficulty of fitting a complex convolutional neural network adapted to the large data sizes in DBT within GPU memory. Due to these constraints, for now it is only possible to perform these reconstructions at low in-plane resolutions.

K. Michielsen, N. Moriakov, J. Teuwen, and I. Sechopoulos are with the dept. of Radiology, Nuclear Medicine, and Anatomy of Radboudumc, Nijmegen, The Netherlands.

J. Teuwen is with the Department of Radiation Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

I. Sechopoulos is with the Dutch Expert Center for Screening (LRCB), Nijmegen, The Netherlands.

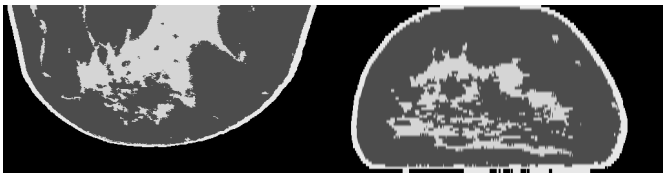


Figure 1. Axial and coronal views of a patient-based phantom. It contains three tissue types: skin at the edge, fibro-glandular structures near the center, and the remainder is adipose tissue.

However, given the promising results, in terms of increased vertical resolution and decreased limited angle artifacts, obtained with our deep learning-based reconstruction, here we examine if it could be introduced as a starting point for an iterative reconstruction method. If the two reconstruction methods are compatible, the iterative algorithm would not change the low resolution information present in the initialization, and only fill in the high resolution details.

II. Materials & Methods

The experimental setup and patient-based phantoms used to train the deep learning network and evaluate the effect of the initialization are described in section II-A, and the metrics used for the evaluation in section II-D. The deep learning reconstruction is presented in section II-B and the iterative reconstruction in section II-C.

A. Phantom & Acquisition

We mixed two sets of breast phantoms to both train the deep learning reconstruction and evaluate the follow-up iterative reconstruction. The first set contained 52 statistically defined anthropomorphic phantoms [3], while the second set consisted of 90 phantoms based on patient scans obtained using a dedicated breast CT system. To generate the phantoms in this second set, the scans were segmented into skin, adipose tissue, and fibro-glandular tissue [4], and then digitally compressed to obtain a phantom that could be used for DBT imaging [5]. An example of the end result is shown in figure 1. We randomly chose 110 samples across the two sets for training and validating the model, with the remaining phantoms reserved for testing. For ease of use, both data sets were converted to an isotropic voxel spacing of $0.25\,\mathrm{mm}$

X-ray projection images were simulated in a geometry based on the Mammomat Inspiration system (Siemens Healthcare, Forchheim, Germany) [6], with a few minor modifications: the detector pixel spacing was set to $0.250\,\mathrm{mm}$, and the detector cover and compression paddle were not included.

A simplified physical model, shown in equation (1), was used in the simulation:

$$\hat{y}_i\left(\vec{\mu}\right) = b_i e^{-\sum_j l_{ij}\mu_j}, \tag{1}$$

with $\vec{\mu}$ being the phantom linear attenuation coefficients, l_{ij} the intersection length between ray i and voxel j, b_i the unattenuated photon count at pixel i, and $\hat{y}$ the resulting projection data. Linear attenuation coefficients for the phantom were calculated for a mono-energetic x-ray beam at $20\,\mathrm{keV}$ with the software of Boone and Chavez [7] using the tissue compositions published by Hammerstein et al. [8]. Blank scan value $\vec{b}$ was set to 16 000 photons to obtain the noise equivalent to a relatively high dose scan.

B. Deep Learning Reconstruction

The algorithm used to compute the low resolution 3D DBT reconstructions, referred to as DBToR-X for now, is a modified, memory-optimized version of the DBToR algorithm that was originally developed for 2D slice-wise DBT reconstruction [2], and was adapted for the purpose of 3D reconstruction with the realistic projection geometry described in section II-A.

The DBToR algorithm takes the raw sinogram and the measured compressed breast thickness as inputs and performs the reconstruction by a neural network, consisting of primal and dual reconstruction blocks [9]. The primal and dual reconstruction blocks are given by small Convolutional Neural Networks (CNNs) that operate in primal (reconstruction) and dual (projection) spaces, respectively. Primal and dual blocks are connected to each other by linear projection and backprojection operators, which implement the corresponding projection geometry. This results in an 'unrolled iterative scheme', where the neural network starts with zero as the initial guess and iteratively improves the reconstruction by applying CNNs in primal and dual spaces. Since all operations are differentiable, the network can be trained end-to-end to minimize the L^2 reconstruction loss on the training set. The inclusion of the compressed breast thickness information in the algorithm significantly improves the reconstruction quality. The DBToR-X version of the algorithm was implemented in PyTorch, and uses the Operator Discretization Library (ODL) with a custom projection geometry to match the DBT acquisition setup. The model was trained to minimize L^2 loss of the reconstruction on the training set, but despite the included memory optimizations, it was not possible to perform this training at full resolution, and output was restricted to a voxel size of $0.75\,\mathrm{mm}\times0.75\,\mathrm{mm}\times1.0\,\mathrm{mm}$.

Nine of the patient-based phantoms in the test set were randomly selected and the matching DBToR-X reconstructions were calculated. These were first up-sampled to full resolution without interpolation and then used as initial guess for the MLTR reconstruction described in the next section.

C. Maximum Likelihood Reconstruction

To evaluate the usefulness of the low resolution deep learning reconstruction as initialization of a high resolution iterative method we continued the reconstruction with the MLTR algorithm [10]. This method maximizes the log-likelihood cost function $\mathcal{L}$ in equation (2), with y_i the measured sinogram and $\hat{y}_i$ the forward model from equation (1). This is done by iteratively updating reconstruction volume $\vec{\mu}^{(n)}$.

$$\mathcal{L}(\vec{\mu}) = \sum_i y_i \ln \hat{y}_i - \hat{y}_i \tag{2}$$

The update step at each iteration is calculated by taking the second order Taylor expansion of $\mathcal{L}$ at that point and further approximating this result by a separable quadratic surrogate function for which the optimum can then be reached in a single update, which is shown in equation (3).

$$\mu_j^{(n+1)} = \mu_j^{(n)} + \frac{\sum_i l_{ij}\left(\hat{y}_i^{(n)} - y_i\right)}{\sum_i l_{ij}\hat{y}_i^{(n)}\sum_k l_{ik}} \tag{3}$$

D. Evaluation

Nine of the patient-based phantoms were selected for reconstruction with 10 000 iterations, once without any initialization (all zero values) and once starting from the deep learning-based reconstruction, which was first up-sampled to the full resolution ($0.25\,\mathrm{mm}\times0.25\,\mathrm{mm}$). Except for the initialization, there was no difference between the reconstructions. Intermediary results were saved after iterations 1 through 10, every 10^{th} iteration until iteration 100, every 100^{th} iteration until iteration 1 000, and then every 500^{th} iteration.

To evaluate the reconstructions, we calculated the log-likelihood cost function and the mean squared error (MSE) for all intermediary results and visually inspected a small number of cases.

III. Results

Figure 2 shows axial and coronal views of reconstructions with and without the deep learning-based initialization after 1, 10, 100, 1 000, and 10 000 iterations. These reconstructions are based on the phantom shown in figure 1. In the reconstruction without initialization, attenuation in the anterior and superior regions remains underestimated and the skin is not visualized even after 10 000 iterations, while the skin is clearly depicted from the start, and remains clearly visible, in the deep learning-initialized reconstruction.

Figures 3 and 4 plot the evolution of the log-likelihood cost function and MSE, respectively, for the same phantom, which are representative of the behavior seen for these two metrics with all 9 tested phantoms. The mean squared error for all 9 phantoms and both reconstructions reaches a minimum between iterations 10 and 100 and then starts increasing again. Meanwhile, the minimum mean squared error of the reconstructions with initialization is always lower than that of the reconstructions without initialization. After 50 iterations, the average MSE was $177 \times 10^{-6}\,\mathrm{mm}^{-2}$ for the reconstructions

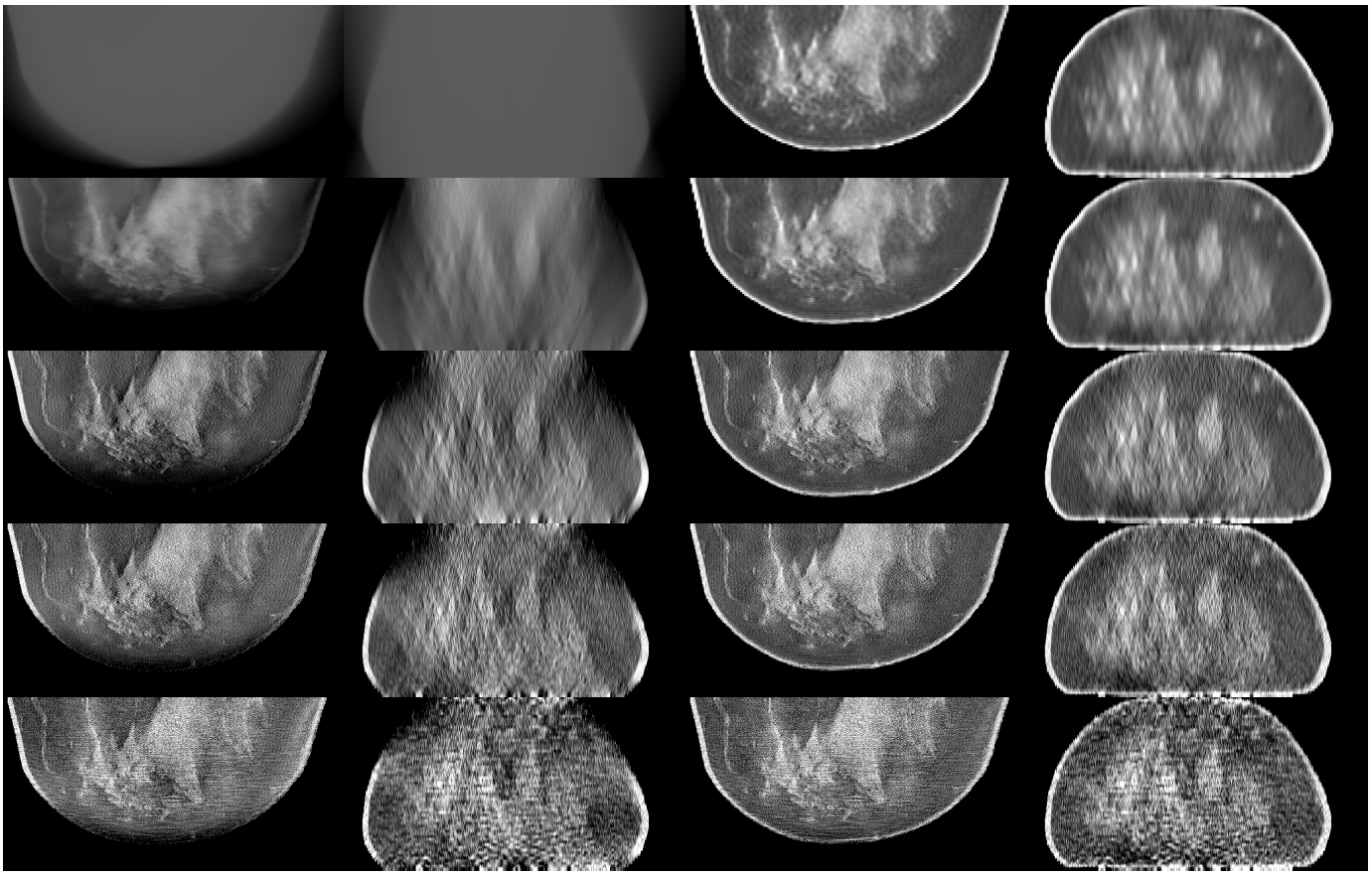


Figure 2. Axial and coronal reconstruction views after iterations 1, 10, 100, 1 000, and 10 000 (top to bottom). Reconstructions without initialization are shown on the left hand side, and reconstructions with the deep learning-based reconstruction on the right hand side. The window is set between $0.04\,\mathrm{mm}^{-1}$ and $0.09\,\mathrm{mm}^{-1}$, except for the first iteration without initialization (top left) where the window is set between $0.00\,\mathrm{mm}^{-1}$ and $0.04\,\mathrm{mm}^{-1}$.

without initialization and $101 \times 10^{-6}\,\mathrm{mm}^{-2}$ for those with initialization.

Log-likelihood was found to be monotonously increasing for all cases, but was not found to be consistently higher for reconstructions with or without the initialization. Specifically, in four cases it was higher for the reconstructions with initialization while for five cases it was higher for the reconstructions without initialization.

Figure 5 shows the difference between both reconstructions and the difference with the ground truth after 100 iterations. Compared to the ground truth, the lack of resolution due to the limited angle acquisition is clearly visible in the distribution of fibro-glandular tissue in the coronal slices, for both methods. However, the skin line is well depicted in the initialized reconstruction, with only a minimal outward shift relative to the ground truth. The comparison between the two methods highlights the differences in visualization of the skin and also hints at a more accurate distribution of the fibro-glandular tissue in the initialized reconstruction, with the subtraction indicating higher attenuation in the center and lower attenuation in the peripheral regions compared to the reconstruction without initialization.

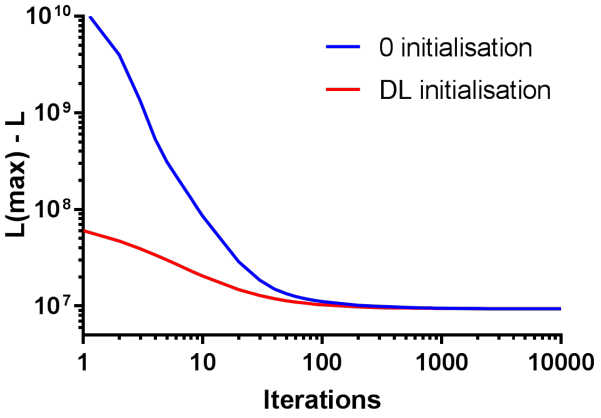


Figure 3. Difference between the maximum log-likelihood and the log-likelihood cost as a function of iterations.

IV. Discussion & Conclusion

Our results clearly show the positive impact gained by including a low resolution deep learning-based reconstruction as initialization of an iterative reconstruction algorithm. The information the initialization adds is compatible with the data, since it was not removed even after 10 000 iterations, but could also not be retrieved from the data by the iterative reconstruction. It is not clear from the metrics in figures 3 and 4 whether or not both reconstructions are converging to the same result, possibly because both methods are starting to over-fit to the

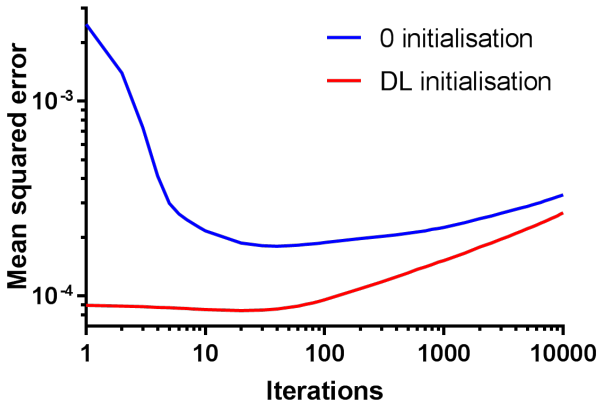


Figure 4. Mean squared error of the reconstructions as a function of iterations.

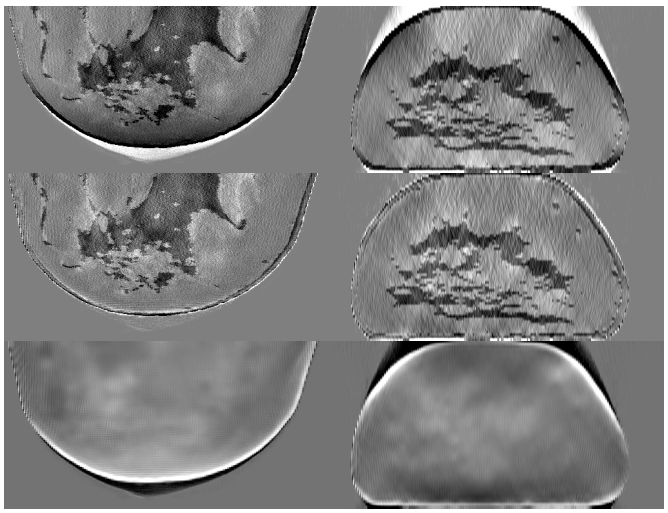


Figure 5. Difference between the reconstruction without initialization and the ground truth (top row), difference between the reconstruction with initialization and the ground truth (middle row), and difference between both reconstructions (bottom row), all after 100 iterations. The window is set between $-0.03\,\mathrm{mm}^{-1}$ and $0.03\,\mathrm{mm}^{-1}$.

noise at that point. At least the visual comparison in figure 2 indicates that the visualization of the skin will not appear in the un-initialized reconstruction, and will not disappear from the initialized reconstruction, and thus both seem to converge to a different solution. For all practical purposes, reconstructions will be stopped well before reaching 10 000 iterations, and the point of convergence is less important than which solution can be found in a relatively short time. With that in mind, the deep learning based reconstruction substantially improves the result.

At this point in time the main reason to combine DBToR-X with MLTR is that the deep learning based method cannot reach the high resolution needed for breast imaging, due to memory constraints in the training process, and thus the regular iterative method is needed to fill in the high resolution details. However, we believe this approach remains valid even if DBToR-X could process reconstructions at full resolution. Because it behaves as a reconstruction combined with some anatomical prior knowledge, it could place structures that are not compatible with the data, especially since it is not clear how much weight is put on the prior knowledge relative to the data fidelity. Combination with a regular iterative method would ensure that such incompatible structures would be removed from the image.

Some modifications would be needed to extend the method to make it applicable to patient data. DBToR-X would need pre-corrected projection data as input, but, this should not be problematic, since filtered backprojection has the same requirement. The main remaining question is, then, how relevant the implicit anatomical prior information learned from the (patient-based) digital phantoms would be, and if it would still help reconstruction quality.

To conclude, we found that the reconstructions that included initialization reached a lower mean squared error than those without, and visual inspection found considerably improved retrieval of the breast outline and depiction of the skin, confirming that adding the deep learning-based initialization adds valuable information to the reconstruction.

References

[1] K. Michielsen and J. Nuyts, "Multigrid reconstruction with block-iterative updates for breast tomosynthesis," *Med. Phys.*, vol. 42, no. 11, pp. 6537–6548, Nov. 2015.

[2] N. Moriakov, K. Michielsen, J. Adler, R. Mann, I. Sechopoulos, and J. Teuwen, "Deep learning framework for digital breast tomosynthesis reconstruction," in *Medical Imaging 2019: Physics of Medical Imaging*, vol. 10948. International Society for Optics and Photonics, Mar. 2019, p. 1094804.

[3] B. A. Lau, I. Reiser, R. M. Nishikawa, and P. R. Bakic, "A statistically defined anthropomorphic software breast phantom," *Med. Phys.*, vol. 39, no. 6, pp. 3375–3385, Jun. 2012.

[4] M. Caballo, J. M. Boone, R. Mann, and I. Sechopoulos, "An unsupervised automatic segmentation algorithm for breast tissue classification of dedicated breast computed tomography images," *Med. Phys.*, vol. 45, no. 6, pp. 2542–2559, Jun. 2018.

[5] C. Fedon, C. Rabin, M. Caballo, O. Diaz, E. García, A. Rodríguez-Ruiz, G. A. González-Sprinberg, and I. Sechopoulos, "Monte Carlo study on optimal breast voxel resolution for dosimetry estimates in digital breast tomosynthesis," *Phys. Med. Biol.*, vol. 64, no. 1, p. 015003, Dec. 2018.

[6] I. Sechopoulos, "A review of breast tomosynthesis. Part I. The image acquisition process," *Med. Phys.*, vol. 40, no. 1, p. 014301, 2013.

[7] J. M. Boone and A. E. Chavez, "Comparison of x-ray cross sections for diagnostic and therapeutic medical physics," *Med. Phys.*, vol. 23, no. 12, pp. 1997–2005, Dec. 1996.

[8] G. R. Hammerstein, D. W. Miller, D. R. White, M. E. Masterson, H. Q. Woodard, and J. S. Laughlin, "Absorbed radiation dose in mammography," *Radiology*, vol. 130, no. 2, pp. 485–491, Feb. 1979.

[9] J. Adler and O. Oktem, "Learned Primal-Dual Reconstruction," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1322–1332, Jun. 2018.

[10] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans, "Iterative reconstruction for helical CT: a simulation study," *Phys. Med. Biol.*, vol. 43, no. 4, pp. 729–737, Apr. 1998.

A dataset of patient-derived digital breast phantoms for in-silico computed tomography dedicated to the breast

Antonio Sarno, Giovanni Mettivier, Francesca di Franco, Antonio Varallo, Andrew M. Hernandez, Kristina Bliznakova, John M. Boone and Paolo Russo

***Abstract* — This work presents the cohort of patient-derived digital breast phantoms as framework for the realization of a platform for virtual clinical trials in 2D and 3D x-ray breast imaging. 92 digital breast phantoms were derived by means of voxel classification of 92 3D breast images acquired at UC Davis (California, USA) via a computed tomography scanner dedicated to the breast (BCT). Each of the image voxel was classified as skin tissue, adipose tissue, glandular tissue or air on the basis of its content and geometrical assumptions. Digital models of tumors developed and presented in previous studies were added to the generated digital breast phantoms for the production of affected breast phantoms. The generated cohort of digital phantoms was characterized on the basis of glandular fraction by mass (mean = 11.9%, min = 1.0%, max = 58.8%), equivalent diameter at the breast center of mass (mean = 106 mm, min = 57 mm, max = 136 mm) and ratio between the equivalent radius at the center of a mass and breast length (mean = 1.6, min = 0.9 mm, max = 3.2). These digital breast phantoms will be adopted for virtual clinical tests via a Monte Carlo software, which reproduces in-silico BCT examinations. Future development of the cohort will include a computed compression of the digital breast phantoms in order to be submitted to in-silico digital mammography or digital breast tomosynthesis exams for comparison to BCT.**

***Index Terms*— Breast CT, Digital breast phantoms, virtual clinical trials, in-silico clinical trials.**

I. Introduction

COMPUTED TOMOGRAPHY dedicated to the breast (BCT) [1] is an x-ray based imaging technique which permits to produce 3D images of the breast by avoiding the breast compression and the related patient discomfort present in digital mammography (DM) and digital breast tomosynthesis (DBT) examinations. BCT has been proposed to overcome the limit of planar DM imaging, where the superimposition of the breast tissues in the projected mammograms may hinder a correct diagnosis. BCT has been approved by the regulatory board (FDA) in USA in conjunction with DM for diagnosis, but it is to be demonstrated that it outperforms both DM and DBT in breast cancer screening and diagnosis. This requires to setup clinical trials showing the advantages of the use of a fully tomographic 3D technique with respect to a 2D imaging technique (DM) and a partial tomographic 3D technique as DBT, at comparable dose levels. As a substitute of clinical trials on a patient population, virtual clinical trials (VCTs) have been proposed in x-ray breast imaging [2,3] for a consistent reduction of the time and the costs related to these studies, with the delivery of no radiation dose to any patient. VCTs can be also adopted for scanner performance assessment via computational phantoms. In this context, VCTs represent in-silico reproductions of DM, DBT or BCT clinical examinations, which rely on digital models of the patient organ and on a computational method for the replication of the exams.

The digital breast phantoms today adopted for the in-silico reproduction of x-ray based breast imaging techniques are mainly meant to replicate, to some degree, the tissue background (the *anatomical noise*). Massive lesions or calcifications can then be introduced with suitable digital insertion techniques. To this purpose, these anthropomorphic or *patient-like* phantoms can be realized based on anatomical assumptions, or be anyway capable of producing images with features similar to real images [2–5]. The sole cohort proposed in ref. [6] was developed from 3D breast images, and then reflected the real glandular tissue distribution.

In this work, carried out in framework of the INFN (Italian National Institute of Nuclear Physics) project AGATA [7] recently started after initial studies [8–10], we present a cohort of *patient-derived* digital breast phantoms suitable for 3D and 2D VCTs in x-ray breast imaging. AGATA will produce a platform for BCT as well as for DM and DBT in-silico studies based on Monte Carlo (MC) simulations. The digital breast phantoms presented in this work were directly derived from high-resolution 3D clinical BCT images acquired at UC Davis (California, USA) and hence they reflect, to a high degree, the breast anatomy (tissue distribution and silhouette) of the real

Manuscript received January 19, 2020.

A. Sarno, G. Mettivier, A. Varallo and P. Russo are with Universita' di Napoli Federico II, Dipartimento di Fisica "Ettore Pancini", Napoli, I-80126 Italy (telephone: +39-081-676339, e-mail: sarno@na.infn.it).

A. Sarno, G. Mettivier and P. Russo are also with INFN Sez. Napoli, Napoli, Italy.

F. di Franco is with with INFN Sez. Napoli, , I-80126, Napoli, Italy.

K. Bliznakova is with Medical University of Varna, Varna, Bulgaria.

A. M. Hernandez and J. M. Boone are with Department of Radiology, University of California Davis Medical Center, Sacramento, CA, USA.

organ in patients. The added digital tumor models, produced within previous studies [11], were derived both from segmentation of clinical breast images and by means of mathematical models. The produced computational breast phantoms will be adopted as input breast models to the MC software for the production of simulated BCT, DBT and DM images. They will also represent input digital models for manufacturing patient-like physical breast phantoms by means of 3D printing techniques [12,13].

II. Materials And Methods

A. 3D breast images cohort

3D clinical breast images were acquired at UC Davis, by means of the Albion BCT clinical scanner [14]. From the available clinical images, 92 breast images of unaffected breasts were selected for this work. These were acquired with W/Cu 80 kV spectrum. The reconstructed 3D images presented a pixel pitch in the coronal slices ranging between 0.1938 mm and 0.4274 mm in 512×512 pixels coronal images and the slice thickness was in the range 0.1907 – 0.2375 mm.

B. Voxel classification

Digital breast phantoms were derived from 3D clinical breast images by means of voxel classification process. Each of the image voxel was classified on the basis of its content in air, glandular tissue, adipose tissue or skin tissue, that are the four main materials represented in the images. The segmentation algorithm [10,15–17] classifies the skin voxel in each of the coronal slice by means of a local adaptive threshold algorithm. The glandular tissue was classified in each of the coronal slice by means of threshold-based algorithm coupled with processes of erosion and hole-filling. Finally, the voxels within the skin border which were not classified as glandular tissue were classified as containing adipose tissue. In order to guarantee the seamless structure of the glandular trees, processes of erosion and hole filling were implemented in the direction perpendicular to the coronal planes. Figure 1 shows an axial slice of one of the patient breast of the adopted cohort before (fig. 1a) and after (fig. 1b) the voxel classification. The voxel content in the original image (fig. 1a) is expressed in Hounsfield Unit (HU); the voxel of the produced digital breast phantom (fig. 1b) can contain one out of four values representing one of the four main classified components.

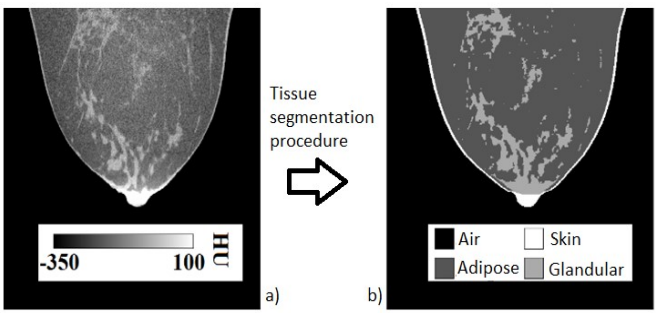


Fig. 1. Axial slice of a 3D image breast acquired at UC Davis (left) and corresponding slice in the classified digital breast phantom (right).

C. Lesion insertion

In order to produce digital breast phantoms with tumor inclusions, we employed the digital models of breast cancer lesions developed within the MAXIMA EU Twin project [11]. In the MAXIMA project a total of 138 3D digital models of breast tumors were developed: 50 models were derived by means of segmentation of lesions in DBT breast images, 8 from CT images of post-mortem specimens and the remaining 80 models were produced by means of a mathematical algorithm. In this work the simulated lesions were added to the unaffected digital breast phantoms by substituting the tumor model within the breast volume with a fifth representing value. The inserted tumor lesion was supposed either to be entirely contained within the glandular tissue or also spreading in the surrounding adipose tissue regions.

D. Monte Carlo software for in-silico BCT simulations

An MC software for the in-silico BCT exams reproduction was developed. It was based on Geant4 simulation toolkit version 10.3 and the low energy physics list Option4 [10, 16,17]. It calculates the x-ray breast image projections and the glandular dose maps of a selected digital phantom on the basis of the input exam specifications (geometry, spectrum and detector dimension and pixel pitch). The MC code simulates coherent and incoherent scatter and photoelectric x-ray interactions within the breast volume. For each Compton scatter or photoelectric interaction, the software scores the released energy as well as the location of the interaction within the glandular tissue for the glandular dose map evaluation. Since, for a fixed beam energy, the detector signal was shown to linearly depend on the exposure at the detector surface [18], the distribution of the incident air kerma on the detector was used for deriving the image projection. The Monte Carlo code was previously validated versus literature data, dosimetric measurements as well as imaging data [10].

III. Results

A. 3D patient-derived digital breast phantoms

Figure 2 shows the whisker plot representing the distribution of the glandular fraction by mass of the 92 digital breast phantoms produced in this work. The corresponding mean glandular fraction (by mass) was 11.9%, ranging between 1.0% and 58.8% on the whole dataset. For the same produced digital breast phantom cohort, the resulting mean equivalent diameter at the breast center of mass - shown to permit more accurate dose estimates than conventionally used diameter at the chest wall [19] - was 106 mm (fig. 3). The minimum and the maximum values of the equivalent diameter at the center of mass resulted 57 mm and 136 mm, respectively.

Figure 4 shows the distribution of the ratio between the equivalent radius (evaluated at the breast center of mass) and the breast length (i.e. chest-wall to nipple distance) of a breast model in the computational dataset. It presents a mean value of 1.6 and minimum and maximum values of 0.9 and 3.2, respectively.

Figure 5 reports sagittal, axial and coronal views as well as a volume rendering of a digital breast phantom with an inserted tumor lesion model. This particular lesion model was

taken from the database previously developed [11], as derived from a CT image of a breast tumour sample.

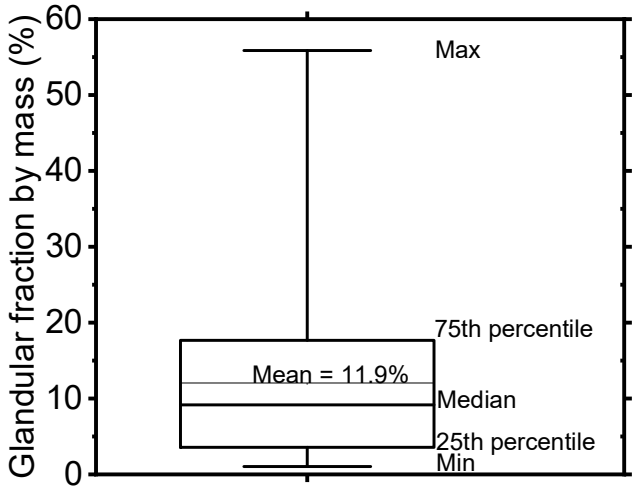


Fig. 2. Distribution of the breast glandular fraction by mass of evaluated in the produced digital breast phantom cohort.

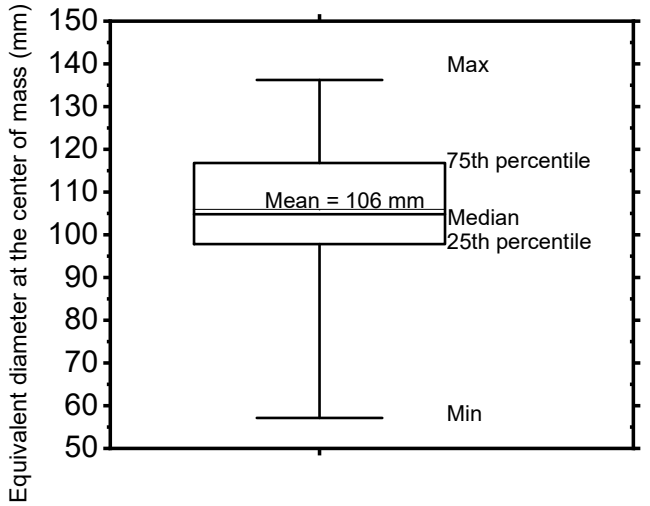


Fig. 3. Distribution of the equivalent diameter evaluated at the breast center of mass in the produced digital breast phantom.

B. First simulated images

Figure 6 compares clinical image (fig. 6a) - sagittal, coronal and axial views and volume rendering - and the corresponding simulated image (fig. 6b) for a selected patient. The simulated image was obtained by means of one of the produced digital phantom and the Monte Carlo software introduced in the sect. II.D. While the clinical image was obtained with a 5 mGy of mean glandular dose (MGD), in this preliminary simulation, the simulated image was obtained with an equivalent MGD of 0.5 mGy corresponding to a total of 1.2×10^{12} photons launched in the simulation.

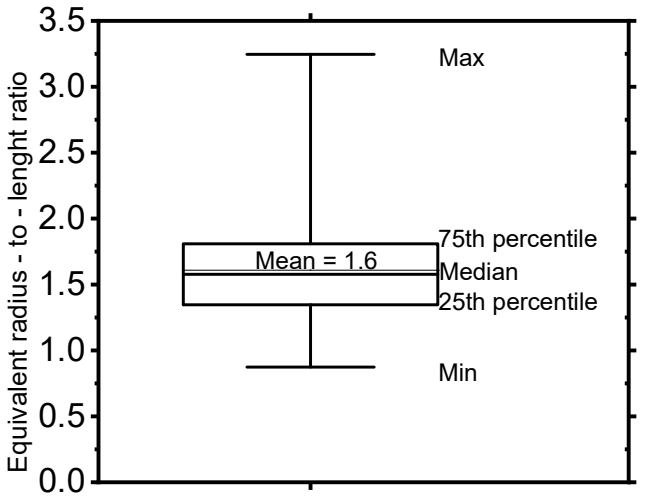


Fig. 4. Distribution of the ratio between the equivalent radius and the length of the digital breast phantoms contained within the produced cohort.

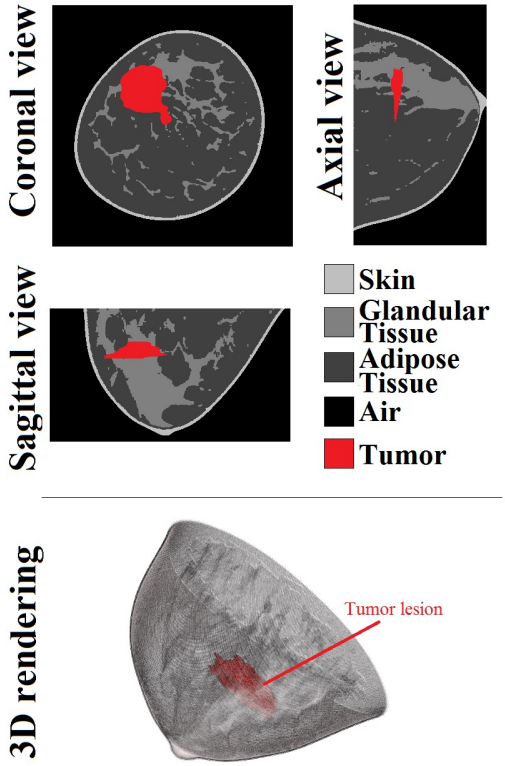


Fig. 5. Coronal, axial and sagittal (top) and 3D rendering (bottom) of a digital breast phantom with a lesion insertion outlined in red. The selected tumor lesion was derived by CT images of post-mortem scan [11].

IV. Discussion and Conclusions

Virtual clinical trials are an alternative strategy for the comparison of imaging performance of BCT versus those of DM and DBT. They permit time and costs sparing and the access to clinical-like data to a broad range of researchers. In addition, VCTs can be adopted for the scanners research and developments in place of tests on physical phantoms.

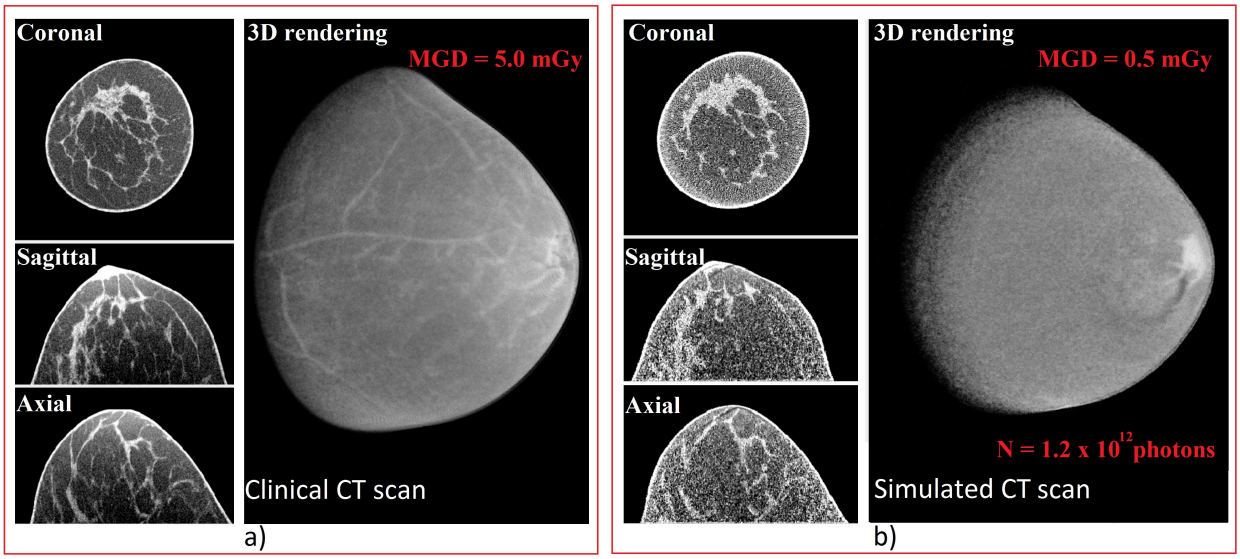


Fig. 6. Coronal, sagittal and axial views and 3D rendering of a a) 3D clinical CT breast image acquired at UC Davis and b) the corresponding simulated 3D images produced by means of one of the digital breast phantoms presented in this work.

A complete platform for VCTs in x-ray breast imaging needs a dataset of digital breast phantoms. These have to mimic in the produced images both a realistic anatomical noise and features and the silhouette of real organs. We here propose the use of digital breast phantoms derived from clinical 3D breast images acquired via a BCT scanner at UC Davis (California, USA). The voxels were classified in air, adipose tissue, glandular tissue and skin and the added digital tumor lesions were taken from a database created in a previous study.

The dataset comprised 92 digital breast phantoms. They presented an average glandular fraction by mass of 11.9% and an average length of the diameter at the breast center of mass of 106 mm. This breast phantom cohort will be used for the in-silico reproduction of BCT exams by means of a Geant4 MC code developed in-house [10]. In addition, each of the produced digital phantoms will be digitally compressed by means of a mathematical models for the adoption in in-silico reproduction of DM and DBT examinations for technique developments and comparisons.

ACKNOWLEDGMENT

This work is funded in part by the AGATA INFN project and MC_INFN project.

REFERENCES

[1] A. Sarno *et al.*, "Dedicated breast computed tomography: basic aspects," *Med. Phys.,* vol. 42, pp. 2786-2804, 2015.
[2] A. Badano et al., "Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial", *JAMA network open,* vol. 1, 2018.
[3] P. R. Bakic et al., "Virtual clinical trial of lesion detection in digital mammography and digital breast tomosynthesis," *Medical Imaging 2018:* Physics of Medical Imaging. vol. 10573, 2018.
[4] J. Boita *et al.,* "Validation of a method to simulate the acquisition of mammographic images with different techniques," *Medical Imaging 2019*: Physics of Medical Imaging. vol. 10948, 2019.
[5] Y. Baneva et al., "Evaluation of a breast software model for 2D and 3D X-ray imaging studies of the breast", *Phys. Med*., vol. 41, 2017.
[6] M. Caballo et al., "An unsupervised automatic segmentation algorithm for breast tissue classification of dedicated breast computed tomography images," *Med. Phys.,* vol. 45, pp. 2542–2559, 2018.
[7] A. Sarno et al., "Advanced Monte Carlo application for in silico clinical trials in x-ray breast imaging. Proc. SPIE, 15th International Workshop on Breast Imaging IWBI 2020.
[8] G. Mettivier *et al*., "The Napoli-Varna-Davis project for virtual clinical trials in X-ray breast imaging," In: *2019 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, in press, 2020.
[9] G. Mettivier et al, "Virtual clinical trials in 3D and 2D breast imaging with digital phantoms derived from clinical breast CT scans," In: Medical Imaging 2020: Physics of Medical Imaging (Vol. 11312, p. 1131259). *International Society for Optics and Photonics*, 2020.
[10] F. di Franco *et al*., "GEANT4 Monte Carlo simulations for Virtual Clinical Trials in X-ray breast imaging: proof of concept," *Phys. Med.,* under review
[11] K. Bliznakova *et al.*, "Development of breast lesions models database," *Phys. Med.*, vol. 64, pp. 293-303, 2019.
[12] D. Ivanov et al., "Suitability of low density materials for 3D printing of physical breast phantoms," *Phys. Med. Biol., vol. 63, no. 17,* 175020*, 2018.*
[13] Esposito G. *et al.*, "Investigation of the refractive index decrement of 3D printing materials for manufacturing breast phantoms for phase contrast imaging", *Phys. Med. Biol., vol. 64, no. 7,* 075008*, 2019.*
[14] J. M. Boone *et al.*, "Dedicated breast CT: radiation dose and image quality evaluation," *Radiology*, vol. 221, pp. 657–667, 2001.
[15] A. Sarno *et al*., "Mettivier G, Tucciariello RM, Bliznakova K, Boone JM, Sechopoulos I et al. Monte Carlo evaluation of glandular dose in cone-beam X-ray computed tomography dedicated to the breast: Homogeneous and heterogeneous breast models," *Med. Phys., vol* 51, pp. 99–107, 2018.
[16] A. Sarno *et al.*, "Dedicated cone-beam breast CT: a patient specific GEANT4 study," III GEANT4 International User Conference – The GEANT4 toolkit at the Physics-Medicine-Biology frontier, October 2018, Bordeaux (Fr)
[17] A. Sarno, "Personalised estimates of mean glandular dose and image quality investigations in cone-beam computed tomography dedicated to the breast," presented at European Conference of Radiology 2019, Vienna (A).
[18] A. Sarno *et al*., "Cone-beam micro computed tomography dedicated to the breast," *Med. Eng. Phys.*, vol. 38, pp. 1449–1457, 2016.
[19] A. M. Hernandez et al., "Updated breast CT dose coefficients (Dg NCT) using patient-derived breast shapes and heterogeneous fibroglandular distributions," *Med. Phys.*, vol. 46, pp. 1455-1466, 2018.

Investigating Deep Learning-based Grating Interferometry Breast Computed Tomography Image Denoising

Stefano van Gogh, Michał Rawlik, Zhentian Wang and Marco Stampanoni

***Abstract*—X-ray grating interferometry is a promising technique that has the potential to bring X-ray phase contrast imaging to the clinic. Breast imaging is predicted to be one of the first application areas due to current shortcomings in dedicated imaging modalities. Our group has therefore developed a grating interferometry breast computed tomography (GI-BCT) prototype that combines grating interferometry with computed tomography. To increase the image quality and to possibly allow for lower dose deposition in the patient, a proper denoising algorithm is desired. However, high and inhomogeneous noise amplitudes across pixels, arising from the complex signal acquisition scheme in grating interferometry, prevent the successful use of classical denoising methods. For this reason, we report about the deployment of a current state-of-the-art deep learning-based denoising method called Noise2Noise to grating interferometry computed tomography data and show that image quality can be improved compared to classical denoisers. In addition, we provide an outlook to methods we are currently implementing in the context of deep learning-based GI-BCT data denoising.**

***Index Terms*—Grating Interferometry, Deep Learning, Image Denoising, Computed Tomography, X-ray imaging, Neural Network**

I. Introduction

X-RAY imaging has been used for over 100 years for biomedical applications. Conventional attenuation-based X-ray imaging, however, has limited soft-tissue contrast due to similar attenuation coefficients of biological materials. For this reason, an ever increasing effort is being made by the research community to exploit X-ray phase contrast, which potentially leads to orders of magnitude higher soft-tissue contrast [1].

X-ray grating interferometry, in particular, is a phase contrast technique that has been studied extensively as it has the prerequisites for clinical compatibility: it has high mechanical robustness, it can be scaled up to large fields of views (FOV) and it requires moderate spatial coherence and monochromaticity [2].

One of the major areas of application of grating interferometry is predicted to be breast imaging, as current technologies, notably digital mammography, breast ultrasound, breast magnetic resonance imaging (MRI), breast tomosynthesis and breast computed tomography (BCT), despite their immense value in clinical practice, all suffer from certain limitations. In fact, none of them is able to provide 3-dimensional volumetric information with sufficiently high density contrast and spatial resolution to visualize clinically relevant imaging biomarkers, such as small soft tissue lesions and their margins, microcalcifications, architectural distortions and tiny soft tissue density differences [3].

To fill this gap in breast cancer imaging, a grating interferometry breast computed tomography (GI-BCT) prototype has been designed and is currently built by our group. To increase the quality of our images and to possibly allow for lower dose deposition in the patient, a proper denoising algorithm is needed. Unfortunately, given the complex noise pattern arising in grating interferometry, and in particular, given the fact that the noise amplitude is different for every pixel, the use of traditional denoising solutions cannot solve the problem.

In recent years, following the impressive results deep learning has achieved in the computer vision domain, people have started to use deep neural networks for biomedical image denoising [4] - [9]. Following this trend, we applied state-of-the-art denoising networks to assess whether image quality can be improved compared to the use of standard denoising solutions.

Given the early stage of the project, the images reported in this work have been obtained by scanning meat samples, specifically pig neck and bacon, rather than human breast samples. Being the focus of this work methodological in nature, this does not have a major impact on the significance of the results.

Section II introduces grating interferometry and our GI-BCT setup, the data types, hardware and the deep learning methods used for GI-BCT data denoising. Section III reports the results and section IV concludes this abstract by summarizing its contributions and providing a brief outlook.

II. Methods

A. Grating interferometry breast computed tomography

Contrarily to attenuation-based imaging, which directly

This work has been supported by the SNF Sinergia Grant Nr. CRSII5_183568 as well as the SwissLOS Lottery Fund of Kanton Aargau.

All authors are with the Paul Scherrer Institute, Villigen AG and ETH Zürich. Correspondence to Stefano van Gogh (e-mail: stefano.van-gogh@psi.ch).

measures the intensity of the transmitted X-ray beam, it is not possible to directly detect the phase, which therefore has to be measured indirectly. Grating interferometry solves this problem through the Talbot effect, by placing three gratings between the X-ray tube and the detector (see [10] for details) and by stepping the (usually) first grating (G1) grating with respect to the second one (G2).

The raw output of this scanning procedure is the so-called phase stepping curve (see [11] for details). During image acquisition we therefore record the per pixel intensity $I_k(m,n)$ as

$$a_0(m,n) + a_1(m,n)\sin\left[2\pi\frac{ik}{N_{ps}} + \phi(m,n)\right] \quad (1)$$

for every detector pixel (m,n), where a_0, a_1, ϕ are the mean, amplitude and phase of the phase stepping curve, N_{ps} the number of phase steps. $k = N_{ps}x_g/p_2$ with x_g being the relative position of G1 to G2 in the main grating direction, and p_2 the pitch of G2.

By combining the phase stepping curves obtained with and without sample, one can retrieve the absorption signal $\bar{T} = a_0^s/a_0^r$, the differential phase (DPC) signal $(\phi^s - \phi^r)/2\pi$, and the dark-field signal $\bar{V} = v^s/v^r$ where r refers to reference and s to sample and v^i is calculated as a_1^i/a_0^i [11].

By rotating the sample, we can acquire enough projections to perform a tomographic reconstruction which leads to reconstructed absorption, phase and dark-field images. Note here that to obtain the phase signal an imaginary filter (Hilbert transform) must be used in the reconstruction [12].

As mentioned in the introduction, the noise properties in grating interferometry are fairly complicated. The uncertainty in the retrieved absorption, differential phase, and dark field contrast depend mainly on detector quantum noise and on vibrations. Here we restrict our analysis to the phase contrast channel, as it is currently the center of our research efforts, and how this is affected by the detector quantum noise. Vibrations correction will be covered elsewhere. As described in [13], the uncertainty in the measured differential phase contrast signal arising from the detector quantum noise is given by:

$$\left(\sigma_{DP,det}\right)^2 = \frac{f_1^r}{2\pi^2(v^r)^2 N_{ps} a_0^r}\left(1 + \frac{f_1^s}{f_1^r \bar{T}\bar{V}^2}\right) \quad (2)$$

where $v^r = a_1^r/a_0^r$ is the flat field visibility. The constants f_1^r and f_1^s describe the relation between photon count and noise spread and are assumed to be equal to 1 based on experimental evidence. As one can appreciate, a low visibility has a strong negative influence on the uncertainty. This is the reason why grating interferometry noise is much more cumbersome to deal with in the case of a large FOV: it is namely very difficult to produce large gratings with precisely controlled pitch, height, duty-cycle and aspect-ratio, and this leads to inhomogeneous visibility maps. In fact, as displayed in Figure 1, the visibility map obtained with our scanner at the moment is very inhomogeneous leading to a non-uniform noise amplitude distribution for different pixels. It is important to note here that, in addition to increased homogeneity, we aim at reaching higher values in our visibility map since it would reduce the uncertainty in the noise. However, this is intrinsically linked to the physics of the system (gratings fabrication) and cannot be artificially improved with deep learning.

We believe that two main aspects contribute to fact that classical state-of-the-art denoising filters such as the Block Matching and 3D Filtering (BM3D) algorithm [14] are not able to satisfactorily denoise our data, and in particular the DPC channel. The first one being the vastly different noise amplitudes across pixels and the second one the very large noise amplitudes that arise in grating interferometry signal retrieval scheme, especially when operating in low-dose regimes.

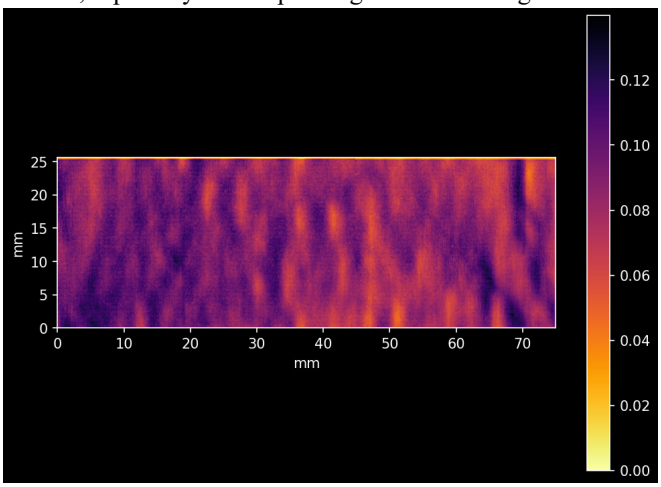


Figure 1: Visibility map (v^r) of part of the FOV (750 × 256 pixels, 100µm pixel size) of our GI-BCT system. It can be seen that it is inhomogeneous which subsequently leads to uneven noise amplitudes across pixels. Note here that the visibility is unitless.

B. *Hardware*

All data has been acquired with our GI-BCT prototype [15] at the Paul Scherrer Institute. All computations involving deep learning architectures have been carried out on a GeForce RTX 2080 Ti NVIDIA GPU.

C. *Deep Learning-based GI-BCT Image Denoising*

Given the early stage of our project, our datasets have been obtained using only a part of the final desired FOV. The flat field data had the following dimensionality: 1890 frames × 5 phase steps × 256 row pixels × 3072 column pixels. The sinogram data had dimensionality: 600 projections × 5 phase steps × 256 row pixels × 3072 column pixels. Each projection, respectively frame, has been obtained with identical exposure time. The flat field data has been generated by averaging the 1890 flat field frames.

The amount of data we used was very limited: we trained on a single scan (pig neck) and tested on another independent scan (bacon). Being this a feasibility study, we believe that the limited amount of data can tell us whether deep learning-based denoising is an approach worth pursuing in the future.

During model training, all hyperparameters have been adjusted based on the training set because no third validation set was available. Very important to note here is that data used for training and testing had to be acquired with the exact same acquisition parameters to avoid differing noise statistics.

Evidently, this is a shortcoming of the technique and future work will try to address this problem by developing more adaptive and robust deep learning algorithms.

Image denoising in the context of computed tomography can be performed at multiple stages in the data processing pipeline. In conventional CT, the data can be denoised either at the projection level or one can filter the final reconstructed volumes. With grating interferometry CT, a third stage of denoising is given, i.e. at the level of the raw phase stepping curves (photon counts for every step).

It is preferable to train the models at the earliest possible point in the signal processing pipeline to avoid uncertainty propagation in the signals. This would suggest to denoise the raw phase stepping curve data. However, we observed that denoising at the retrieved projection level led to better results. This is likely attributable to the fact that at the raw phase stepping curve stage the data has too little structure, i.e. the data is not informative enough, to enable efficient learning.

Finally, given that grating interferometry is a multimodal imaging technique providing absorption, differential phase, and dark field contrast images, neural networks can be trained either on single channels or on multiple channels jointly.

Initial attempts in deep learning-based image denoising used clean target images [5]. However, the availability of such images is rarely given (in our case, no clean target images were available either). In [5] the authors showed that, as long as the noise has zero average, one can train a model F on pairs of noisy images (x_i, x_j) by minimizing the empirical risk

$$\underset{\theta}{\operatorname{argmin}} \sum_i \mathcal{L}(F(\theta, x_i), x_j) \tag{3}$$

where $\mathcal{L}$ is a loss function than compares one noisy image to the transformed version of the other. The results predicted by that model will then converge to the clean version [5]. The authors called this approach "Noise2Noise". If the models need to be trained on sample data, the availability of noisy data pairs can be cumbersome due to dose limitations.

At present, we focused on the Noise2Noise method as in principle it is able to provide best possible denoising performance in the absence of any clean target images, and we will only report about this method here. However, we are also working on the implementation of other popular self-supervised denoising systems such as Probabilistic Noise2Void [8] and Noise2Self [6] which reduce the dose deposition as only single images are required.

As in the original Noise2Noise publication, we used a standard U-net architecture [16] trained on a mean squared error loss, however, we used a different number of filters in different layers: we used ReLU instead of leaky ReLU and included batch normalization layers to improve denoising performance. The data has been normalized channel-wise, to avoid numerical artefacts in the differential phase contrast channel. The model has been trained for 3 epochs with a learning rate of 0.001 for around 15 minutes. All other hyperparameters were identical as in [5]. The model has been implemented in Tensorflow 2.0.

We retrieved the three channels (absorption, differential phase contrast and dark field) by Fourier analysis, and fed them jointly into the model (mini batch data dimensionality: 1 × 256 × 3072 × 3). The network used 2D multichannel convolutions in the detector plane, i.e. no information from neighboring projections was provided during training.

After denoising, the absorption and dark field channels have been reconstructed using standard filtered back-projection, whereas a Hilbert filter has been used to reconstruct the phase tomogram. We analyzed the reconstructed phase image with three standard image quality metrics described below.

D. Image quality metrics

Signal-to-noise ratio (SNR), contrast-to-noise-ratio (CNR) and full-width-half-maximum (FWHM) have been calculated as in [17]. As we aim at optimizing our system for phase-contrast image quality, we present here the performance only for this channel.

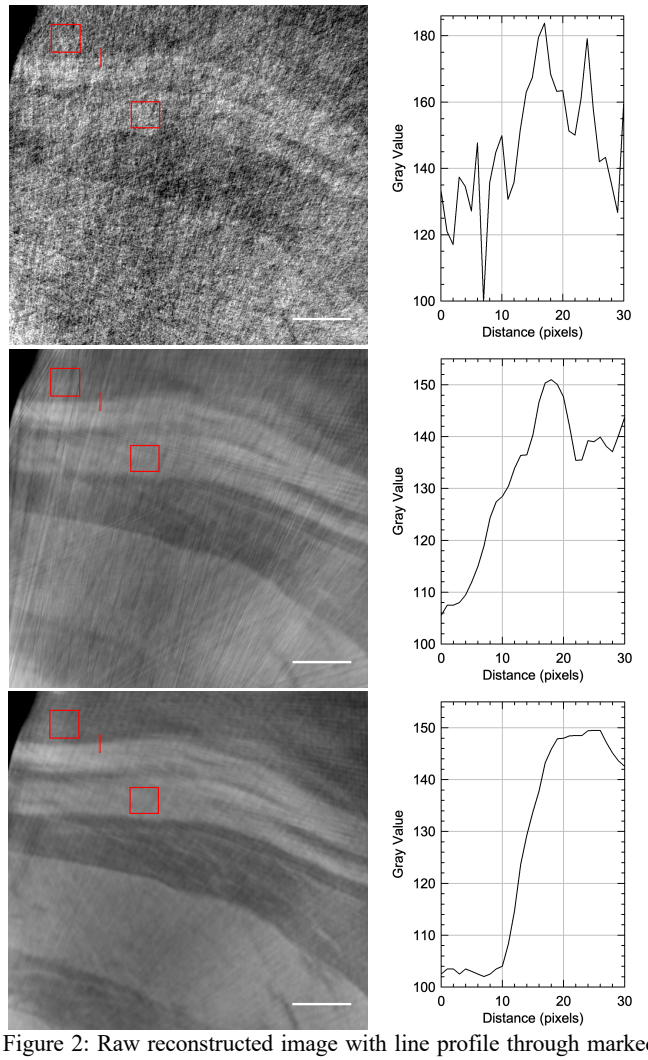


Figure 2: Raw reconstructed image with line profile through marked red line (top), reconstructed image based on projections denoised with the BM3D filter with corresponding line profile (middle), and reconstructed image based on projections denoised with a Noise2Noise-based approach with corresponding line profile (bottom). The sample in the image is a piece of bacon. The scale bar measures 5 mm. SNR and CNR have been calculated based on the two red rectangles. The FWHM has been calculated based on the line profile.

III. Preliminary results

Reconstructions based on raw and denoised projections are shown in Figure 2. The image reconstructed from raw projections had a SNR of 7.38, a CNR of 0.45 and a FWHM of 100 µm. The image reconstructed from projections denoised with the Noise2Noise-based method had a SNR of 29.58, a CNR of 2.41 and a FWHM of 312 µm. SNR and CNR could thus be improved by a factor of 4 and 5, respectively. Note here that although the results suggest a significant loss in resolution, the FWHM of the raw image is misleading owing to the very high noise (see top right panel in Figure 2). For comparison, we report the results of an image reconstructed from projections denoised with the BM3D algorithm [14]: SNR of 21.60, CNR of 1.27 and FWHM of 408 µm. This illustrates that the deep learning-based method is able to produce images with better resolution, as well as higher SNR and CNR compared to the state-of-the-art filter prior the rise of deep learning. Important to mention here is the algorithm speed: 10 seconds for the deep learning denoiser, 72 minutes for the BM3D algorithm (on CPU). Finally, qualitative inspection of the two denoised images supports the quantitative results as it shows that in the Noise2Noise-based denoising small structures can be identified which in the BM3D denoised image are not visible.

IV. Discussion

We could show that Noise2Noise-based denoising is capable of improving image quality, even when trained on a single sample projection data.

Unfortunately, the standard Noise2Noise method as it is currently implemented has a few important shortcomings that have to be addressed. Some authors [9] hypothesized that the model might be implicitly learning a representation (or prior) of the image as seen during the training. Therefore, we will investigate if and how easily the model overfits on the training data. Further, we will examine whether self-supervised methods such as Noise2Void and Noise2Self [4] - [7] can be successfully applied to our data. In all cases, we plan to repeat all experiments with more data and with data with less statistics, i.e. less photon counts, to see how this affects performance.

In a later step, we will further adapt the generic deep learning architecture used in most denoising papers to our particular data type. The main motivations for this are twofold. The first reason is because the noise amplitude is different for every pixel, whereas the second is linked to the fact we want to push the resolution as much as possible because of its fundamental importance for clinical applicability.

In conclusion, this exploratory work highlights the important role deep learning could play in the transition of GI-BCT to the clinic due to its potential to cope with low-dose requirements.

References

[1] S. A. Zhou and A. Brahme, "Development of phase-contrast X-ray imaging techniques and potential medical applications," *Phys. Medica*, vol. 24, no. 3, pp. 129–148, 2008.

[2] T. Weitkamp *et al.*, "X-ray phase imaging with a grating interferometer," *Opt. Express*, vol. 13, no. 16, p. 6296, 2005.

[3] M. E. D'Orsi CJ, Sickles EA, Mendelson EB, "ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System," *Am. Coll. Radiol.*, 2013.

[4] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void - Learning Denoising from Single Noisy Images," pp. 2129–2137, 2018.

[5] J. Lehtinen *et al.*, "Noise2Noise: Learning image restoration without clean data," in *35th International Conference on Machine Learning, ICML 2018*, 2018, vol. 7, no. 3, pp. 4620–4631.

[6] J. Batson and L. Royer, "Noise2Self: Blind Denoising by Self-Supervision," 2019.

[7] J. Xu, Y. Huang, L. Liu, F. Zhu, X. Hou, and L. Shao, "Noisy-As-Clean: Learning Unsupervised Denoising from the Corrupted Image," pp. 1–13, 2019.

[8] A. Krull, T. Vicar, and F. Jug, "Probabilistic Noise2Void: Unsupervised Content-Aware Denoising," 2019.

[9] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep Image Prior," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.

[10] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, "Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources," *Nat. Phys.*, vol. 2, no. 4, pp. 258–261, 2006.

[11] A. Momose, S. Kawamoto, I. Koyama, Y. Hamaishi, K. Takai, and Y. Suzuki, "Demonstration of x-ray Talbot interferometry," *Japanese J. Appl. Physics, Part 2 Lett.*, vol. 42, no. 7 B, 2003.

[12] F. Pfeiffer, O. Bunk, C. Kottler, and C. David, "Tomographic reconstruction of three-dimensional objects from hard X-ray differential phase contrast projection images," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 580, no. 2, pp. 925–928, 2007.

[13] V. Revol, C. Kottler, R. Kaufmann, U. Straumann, and C. Urban, "Noise analysis of grating-based x-ray differential phase contrast imaging," *Rev. Sci. Instrum.*, vol. 81, no. 7, 2010.

[14] K. Dabov, A. Foi, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[15] M. Rawlik, Z. Wang, and M. Stampanoni, "Large-FOV Grating Interferometry Computed Tomography for Breast Imaging," in *5th International Conference on X-ray and Neutron Phase Imaging with Gratings, Abstract Book*, 2019, pp. 130–131.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[17] L. Brombal *et al.*, "Monochromatic breast CT: absorption and phase-retrieved images," vol. 1057320, no. March 2018, p. 71, 2018.

TOmographic MOdel-BAsed Reconstruction (ToMoBAR) software for high resolution synchrotron X-ray tomography

Daniil Kazantsev and Nicola Wadeson

***Abstract*—ToMoBAR is an open-source software which contains a library of model-based regularised iterative reconstruction methods with a plug-and-play capability. With the help of flexible data models and the selection of regularisers, one can achieve a quantifiable image quality for a variety of challenging reconstruction problems. The main aim of ToMoBAR is to solve big real data reconstruction problems in a short period of time with computational efficiency. In this paper, we outline the main functionality of the package in application to simulated data and real tomographic data obtained at Diamond Light Source (UK) synchrotron.**

***Index Terms*—iterative methods, splitting methods, regularisation, open-source software**

I. Introduction

THE challenge of processing big data effectively and efficiently is crucial for many synchrotron facilities which can collect up to several petabytes of data annually. At Diamond Light Source (DLS) synchrotron (UK), the data is collected from over 30 beamlines and integrated facilities. Although the tomographic data is processed using the MPI protocols, it remains important for reconstruction tools to be computationally efficient.

It is known that the image reconstruction of big data (e.g. $(4k)^3$ voxels in size) is a demanding computational task. Many synchrotrons, including DLS, routinely use direct reconstruction methods, such as Direct Fourier Inversion and Filtered Back-projection (FBP) [1]. Notably, there are efficient realisations of direct methods, namely: the Gridrec method [2] (enables fast polar-Cartesian coordinates interpolation for inverse FFT) from the TomoPy package [3] and the GPU-based ray-tracing FBP methods of the ASTRA toolbox [4]. While direct reconstruction enables a fast parameters-free recovery, it performs poorly for undersampled, noisy and erroneous data. Iterative reconstruction (IR) methods, however, can deal much better with problematic measurements by more realistic data modelling. The drawbacks of IR methods are significantly more computationally demanding implementations and the variety of parameters requiring optimisation to achieve a better reconstruction. Fortunately, with the growth of computing capabilities it becomes possible to perform complex IR tasks on CPU/GPU hardware architectures within reasonable time.

There are a number of leading reconstruction software packages which deliver optimised low-level routines, such as forward and back-projection as well as various direct and IR methods (see Table I). In this paper, we primarily focus on IR software for X-ray computed tomography with the exception of STIR for PET/SPECT imaging [6]. In Table I, the standalone low-level open-source packages are presented with core elements implemented in: C, C++, and CUDA languages. Many of these packages have wrappers in Matlab and/or Python providing easy access to efficient core modules. The packages support different reconstruction geometries and have a set of IR algorithms that are either algebraic, based on solving the set of linear equations, (ART, SIRT, CGLS, POCS) or statistical (MLEM, OSEM).

TABLE I
Standalone packages for tomographic reconstruction

Software Version	Stands for	IR methods Geometry	Compilers Interpreters
ASTRA [4] 1.9.9.dev	All Scale Tomographic Reconstruction Antwerp	SART, SIRT, CGLS, MLEM parallel, fan, cone-beam	C++, CUDA Matlab, Python
TIGRE [5] 0.1.8	Tomographic Iterative GPU-based REconstruction	SART, OS-SART, SIRT, CGLS MLEM, TV-POCS parallel, fan, cone-beam	CUDA Matlab, Python
TomoPy [3] 1.7.0	Tomographic Reconstruction in Python	ART, BART, MLEM, OSEM, PML, OSPML, TV-methods parallel	C, CUDA Python
STIR [6] 4.0	Software for Tomographic Image Reconstruction	MLEM, OSEM, OSL, OSSPS specific to PET/SPECT	C++

In Table II, we present a group of packages which can be loosely categorised as high-level tomographic software. Normally, these packages have a dependency on low-level software and are written using interpreted languages. Notably, Python remains a first choice for many high-level packages due to convenient object-oriented elements and open-source nature. Most of the packages in Table II use Python classes to wrap low-level modules of software in Table I.

Additionally the IR methods available with high-level packages are more complex and tend to deal with various challenging aspects of non-linear optimisation. There is a

D. Kazantsev and N. Wadeson are with Diamond Light Source Ltd. Diamond House, Harwell Science and Innovation Campus, Fermi Ave, Didcot OX11 0DE; e-mail: daniil.kazantsev@diamond.ac.uk

TABLE II
HIGH-LEVEL SOFTWARE FOR TOMOGRAPHIC RECONSTRUCTION

Software Version	Stands for	IR methods	Interp.
ODL [7] 0.7.0	Operator Discretization Library	primal-dual ADMM, PDHG CGLS, BFGS with **ASTRA** projectors	Python
SIRF [8] 19.10	Synergistic Image Reconstruction Framework	methods inherited from **STIR, CIL**	Python
CIL [9] 19.10	Core Imaging Library	PDHG, SDPDHG, ADMM, FISTA, GD, CGLS with **ASTRA, STIR** projectors	Python
ToMoBAR [10] 20.01	TOmographic Model-BAsed Reconstruction	FISTA, OSFISTA ADMM with **ASTRA** projectors	Python, Matlab
Savu [11] 2.4	Tomography Reconstruction and Processing Pipeline	methods inherited from **ASTRA, TIGRE, TomoPy**, **ToMoBAR**	Python

substantial effort to implement and employ various splitting algorithms of primal-dual nature to deal with nondifferentiable functionals [12]. Due to the flexibility of dealing with complicated multi-term objective functions, methods such as Alternating Direction of Multipliers (ADMM) [13], Primal-Dual Hybrid Gradient (PDHG) [12], a Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [14], and others are proven to be successful for advanced tomographic reconstruction.

The presented ToMoBAR package also contains the implementation of splitting algorithms, however the main purpose is *to reconstruct large tomographic data efficiently*. In contrast to ODL and CIL packages which aim to solve inverse problems with mathematical rigor, ToMoBAR is driven mostly by computational efficiency and the challenge to minimise various imaging artifacts through nonlinear, and in some cases non-convex, data models [15]. Because of this, for some cases the global solution is not guaranteed, however in practice, for problematic data cases it is usually possible to obtain quantifiable results of a better quality. Apart from various modifications to accelerate convergence and compensate for artifacts, ToMoBAR has also the traditional algorithms with proven convergence which can be used for benchmarking. Additionally, ToMoBAR performs regularisation through the CCPi Regularisation Toolkit [16] which contains more than 10 different CPU/GPU regularisers. With various data models and regularisers one can construct multiple variations of IR algorithms using ToMoBAR.

Being a part of Savu [11] (MPI-based tomographic application for large data), ToMoBAR has an access to multiple CPU/GPU nodes of a computing cluster. With parallel-beam geometry setup at DLS one can iteratively reconstruct $(2\text{k})^3$ voxels volumes using ToMoBAR in under three minutes time.

II. TOMOBAR MODELS AND SOFTWARE DESCRIPTION

ToMoBAR is an open-source Python and Matlab toolbox to perform tomographic data processing and image reconstruction tasks. Inherently it uses the low-level forward and back-projection modules of ASTRA toolbox [4]. The direct and IR methods from ASTRA are also wrapped for ToMoBAR for a quicker access. At the moment, ToMoBAR supports only parallel beam geometry since the main aim is to reconstruct synchrotron tomographic data. It can be generalised, however, to any geometry which ASTRA allows. To install ToMoBAR in Python one can use Conda environment and for Matlab one can simply copy the source code. ToMoBAR has dependencies on ASTRA toolbox [4] and the CCPi Regularisation Toolkit [16].

ToMoBAR v.20.01 consists of two main IR reconstruction algorithms: FISTA and ADMM. While the ADMM algorithm follows a classical implementation [13], FISTA has been modified in a number of ways. With the ordered-subset strategy, FISTA has been accelerated resulting in OSFISTA [17], to give up to 10 times faster convergence than the original method. A number of different data models has been added and due to separability of the objective function one can select a suitable regulariser from the CCPi Regularisation Toolkit [16].

The reconstruction task is to recover the unknown attenuation coefficient distribution (the absorption map) $\mathbf{x} \in \mathbb{R}^N$ using the log-corrected normalised tomographic projection data $\mathbf{b} = -\ln\left(\frac{\mathbf{Y}}{\mathbf{I_0}}\right)$, where $\mathbf{Y} = \{y_j\}_{j=1}^M$ are raw measurements and $\mathbf{I_0}$ is the intensity of the incoming beam.

The following regularised optimisation problem needs to be solved:

$$\min_{\mathbf{x}} f(\mathbf{Ax}, \mathbf{b}) + \beta g(\mathbf{x}),$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the system projection matrix, $f : \mathbb{R}^M \to \mathbb{R}_+$ is continuously differentiable data misfit term, $g : \mathbb{R}^N \to \mathbb{R}_+$ is a convex penalty, and β is the regularisation parameter.

Let $\mathbf{r} = (\mathbf{Ax} - \mathbf{b})$, then for the least-squares (LS) data model we have $f(\mathbf{Ax}, \mathbf{b}) = \mathbf{r}^\top \mathbf{r}$ and for penalised weighted least-squares (PWLS) model $f(\mathbf{Ax}, \mathbf{b}) = \mathbf{r}^\top \mathbf{\Lambda}^{-1} \mathbf{r}$. Here the noise covariance matrix $\mathbf{\Lambda} = \sigma^2 \mathbf{I}$ and one usually select $\sigma_j^2 = y_j$. For PWLS model the weights can be also absorbed as $\mathbf{A} \equiv \sqrt{\mathbf{\Lambda}}\mathbf{A}$ and $\mathbf{b} \equiv \sqrt{\mathbf{\Lambda}}\mathbf{b}$ [15]. Additionally one can consider a more realistic Poisson noise model and minimise Kullback-Leibler divergence: $f(\mathbf{Ax}, \mathbf{b}) = \sum_{j=1}^M (b_j(\log b_j - \log[\mathbf{Ax}]_j))$.

In Table III we provide various data and regularisation models available in OSFISTA method of ToMoBAR v.20.01. Note that apart from the LS and PWLS data models, other models are not mutually exclusive and can be stacked together to form more complex functionals. For instance, one can use the Group-Huber data penalty [18], [15] to minimise ring artifacts together with the Huber data penalty and PWLS forming the following cost function:

$$\min_{\mathbf{s}} \tfrac{1}{2}\|\mathbf{s} - \mathcal{L}^\top \rho(\mathbf{r})\|_{\mathbf{\Lambda}}^2 + \beta g(\mathbf{x}) + \lambda \|\mathbf{s}\|_1.$$

TABLE III
TOMOBAR MODELS WITH OSFISTA ALGORITHM

Data model	$f(\mathbf{Ax}, \mathbf{b}), \mathbf{r} = \mathbf{Ax} - \mathbf{b}$	Regulariser	$g(\mathbf{x})$
LS	$\frac{1}{2}\|\mathbf{r}\|^2_{\mathbf{\Lambda}=\sigma^2\mathbf{I}}$	Rudin-Osher-Fatemi TV	$\|\nabla\mathbf{x}\|_{TV,\epsilon}$
PWLS	$\frac{1}{2}\|\mathbf{r}\|^2_{\mathbf{\Lambda}=\mathbf{Y}}$	primal-dual / fast gradient projection / split-Breg exact TV	$\|\nabla x\|_{TV}$
KL	$\sum_{j=1}^{M}(b_j(\log b_j - \log[\mathbf{Ax}]_j))$	Nonlinear diffusion	$\int \phi(\|\nabla\mathbf{x}\|^2)$
Huber	$\rho(\mathbf{r}) = \sum_{j=1}^{M}\phi(r_j)$ $\phi(r) =$ $\begin{cases} \frac{1}{2}r^2, \|r\| \le \sigma, \\ \sigma(\|r\| - \frac{1}{2}\sigma) \end{cases}$	Total General. Variation	$\alpha_1\|\nabla\mathbf{x} - \mathbf{v}\|$ $+ \alpha_0\|\mathcal{E}(\mathbf{v}\|$
Student's t	$\rho(\mathbf{r}) = \sum_{j=1}^{M}\phi(r_j)$ $\phi(r) =$ $\log(1 + (r/\sigma)^2)$	ROF-LLT	$\alpha_1\|\nabla\boldsymbol{x}\|_{TV,\epsilon}$ $+ \alpha_2\|\nabla^2\boldsymbol{x}\|_{\epsilon_2}$
Group-Huber	$\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{s} - \mathcal{L}^\top\mathbf{r}\|_2^2$ $+\lambda\|\mathbf{s}\|_1,$ $\boldsymbol{\mathcal{L}} =$ $(\mathbf{I}_{m_1} \otimes \mathbf{1}_{m_2})/\sqrt{m_2}$ with $\mathcal{L}^\top\mathcal{L} = \mathbf{I}_{m_1}$	Nonlocal TV	$\|\nabla_\epsilon(\omega)\mathbf{x}\|_{TV}$

Any suitable regulariser $g(\mathbf{x})$ can be chosen from Table III using the CCPi Regularisation Toolkit [16]. The resulting objective can be optimised using a splitting algorithm and in this instance we used FISTA [14].

In Fig. 1, we demonstrate a Python code snippet from ToMoBAR which enables the OSFISTA algorithm with the PWLS data model and TV regularisation.

```
1 from tomobar.methodsIR import RecToolsIR
2 # Set scanning geometry parameters and initiate a class object
3 tomo = RecToolsIR(DetectorsDimH = detHoz,  # Horizontal detector dimension
4                    DetectorsDimV = None,  # Vertical detector dimension (3D data)
5                    CenterRotOffset = None,# Center of Rotation scalar (3D data)
6                    AnglesVec = AnglesRad, # Array of projection angles in radians
7                    ObjSize = N,           # Reconstructed object dimensions (scalar)
8                    datafidelity='PWLS',   # Data fidelity, choose from LS, KL, PWLS
9                    device_projector='gpu')
10 ######################## Creating the data dictionary: ########################
11 _data_ = {'projection_norm_data' : dataNorm, # Normalised projection data
12          'projection_raw_data' : dataRaw,   # Raw projection data
13          'OS_number' : 6                    # Number of ordered subsets
14          }
15 lc = tomo.powermethod(_data_) # calculate Lipschitz constant
16 ## Creating the regularisation dictionary using the CCPi regularisation toolkit: ##
17 _regularisation_ = {'method' : 'PD_TV',        # Selected regularisation method
18                     'regul_param' : 0.000001, # Regularisation parameter
19                     'iterations' :80,         # Number of regularisation iterations
20                     'device_regulariser': 'gpu'}
21 ######################## Creating the algorithm dictionary: ########################
22 _algorithm_ = {'iterations' : 25,           # Number of algorithm iterations
23               'lipschitz_const' : lc}
24 # RUN THE FISTA RECONSTRUCTION METHOD:
25 RecFISTA_os_tv_pwls = tomo.FISTA(_data_, _algorithm_, _regularisation_)
```

Fig. 1. The snippet of a Python code in ToMoBAR enabling the OSFISTA algorithm with the PWLS data-model and Primal-Dual TV [12] penalty term. There are four main components for IR methods in ToMoBAR which need to be set: 1) the data-model and the geometry class (line 3); 2) the data dictionary (line 11); 3) the regularisation dictionary (line 17) and 4) the algorithm dictionary (line 22). One can access the detailed information on parameters by typing *help(RecToolsIR)* in Python command line.

In the next section we demonstrate several numerical examples of how the stacked models of ToMoBAR work synergistically to ensure better reconstruction quality.

III. NUMERICAL EXPERIMENTS

For our experiments with simulated data we use the TomoPhantom software [19], which is a convenient tool to access an extensive library of enumerated 2D-4D phantoms and their analytical Radon transforms. TomoPhantom can also model various imaging artifacts and noise. In this example we use model no. 14 and to the data we add Poisson noise, zingers (dead pixels) and detector offsets. Zingers and offsets can happen due to scattering and detector miscalibration processes, respectively. In the reconstructed images they lead to extensive streaks and ring artifacts [15]. These nonlinear data errors do not fit to the expected linear data model which is used routinely for IR reconstruction. For such cases, one needs to either pre-process (filter) the data or to use a more suitable data model. Since we use the IR methods, one might want to incorporate more realistic data models rather than modifying the raw data which can lead to various corruptions of the reconstructed images.

In Fig. 2 one can see the results of the reconstructed 512×512 phantom using only 256 noisy projections with imaging errors. Expectantly, the FBP reconstruction is noisy and contains multiple errors. The regularised LS OSFISTA-TV algorithm is able to successfully suppress noise and some minor artifacts, however one can clearly see (also noticeable in the image reconstruction error) the streaks and ring artifacts. Applying the Huber data penalty (see Table III) one can suppress the contribution of zingers (streaks) but the ring artifacts overall remain. Clearly one needs a better data model to be able to suppress ring artifacts. One such model has been proposed in [18] and we refer to it as a Group-Huber model [15] (see Table III). In Fig. 2 we employ such model on top of the existing Huber data model. One can see that it is possible to suppress effectively ring artifacts as well as streaks and noise. The resulting RMSE quality metric also shows the smallest error using the Huber-GH-TV model.

In Fig.3 we demonstrate the use of ToMoBAR in application to real data of dendritic growth collected at DLS synchrotron [20]. Here we show how one can use the KL data term together with the GH model and different regularisers. Since the data is seriously undersampled and also low-dose, the FBP reconstruction is very noisy. Applying OSFISTA algorithm with the TV regulariser increases the quality of the image significantly. The ring artifacts, however, remain a serious problem for the subsequent image analysis. Applying the GH data fidelity one can suppress the ring artifacts substantially. One can also argue that the piecewise-constant nature of TV regularisation is not very suitable for the dendritic surfaces. Thus it can be easily substituted to piecewise-smooth regularisation model [16] resulting in a new LK-GH-ROF/LLT reconstruction method.

IV. CONCLUSION

We have presented an open-source ToMoBAR software which is a practical and efficient tool to deal with the challenging data of X-ray tomography. With a variety of data and regularisation models it is possible to construct a suitable objective for a particular problem. We demonstrate the Python

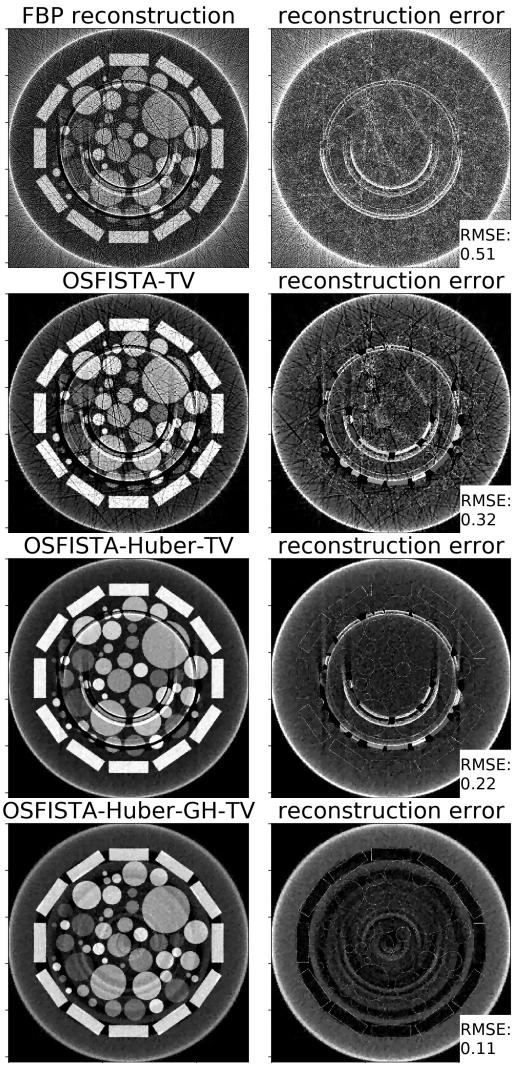


Fig. 2. ToMoBAR reconstructions using the simulated projection data with noise and imaging artifacts. Various data models exploited to show the difference.

syntax of ToMoBAR which makes it easier to construct a desirable cost function.

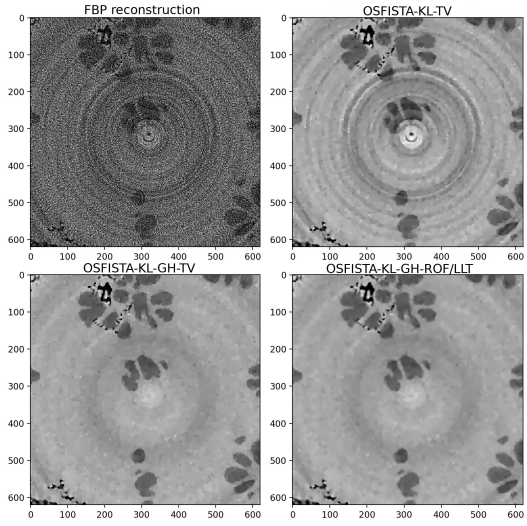


Fig. 3. ToMoBAR reconstructions in application to the real data using KL and GH data models with different regularisers.

REFERENCES

[1] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, N.Y.: IEEE Press, 1988.

[2] F. Marone and M. Stampanoni, *Regridding reconstruction algorithm for real-time tomographic imaging*. Journal of synchrotron radiation, 19(6), pp.1029–1037, 2012.

[3] D. Gürsoy, F. De Carlo, X. Xiao and C. Jacobsen, *TomoPy: a framework for the analysis of synchrotron tomographic data*. Journal of Synchrotron Radiation, 21(5), pp. 1188–1193, 2014.

[4] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J.Batenburg, J. Sijbers, *Fast and flexible X-ray tomography using the ASTRA toolbox*. Optics Express, 24(22), pp. 25129–25147, 2016.

[5] A. Biguri, M. Dosanjh, S. Hancock and M. Soleimani,*TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction*. Biomedical Physics & Engineering Express, 2(5), p.055010, 2016.

[6] K. Thielemans, C. Tsoumpas, S. Mustafovic, T. Beisel, P. Aguiar, N. Dikaios and M. W. Jacobson, *STIR: software for tomographic image reconstruction release 2*. Physics in Medicine & Biology, 57(4), p.867, 2012.

[7] J. Adler, H. Kohr and O. Oktem. *Operator discretization library (ODL)*. Software available from https://github.com/odlgroup/odl, 2017.

[8] E. Ovtchinnikov, R. Brown, C. Kolbitsch, E. Pasca, C. da Costa-Luis, A. G. Gillman, B. A. Thomas, N. Efthimiou, J. Mayer, P. Wadhwa and M. J. Ehrhardt, *Sirf: Synergistic image reconstruction framework*. Computer Physics Communications, p.107087, 2019

[9] E. Ametova, G. Fardell, J. Jorgensen, D. Kazantsev, S. Nagella, E. Pasca, E. Papoutsellis, *Core imaging library for image reconstruction*. Software available at https://www.ccpi.ac.uk/CIL, 2019

[10] D. Kazantsev, *TOmographic MOdel-BAsed Reconstruction software*. Software available at https://github.com/dkazanc/ToMoBAR, 2019.

[11] N. Wadeson and M. Basham, *Savu: A python-based, mpi framework for simultaneous processing of multiple, n-dimensional, large tomography datasets*. arXiv preprint arXiv:1610.08015, 2016

[12] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*. Journal of mathematical imaging and vision 40(1), 2010.

[13] B. Wahlberg, S. Boyd, M. Annergren and Y. Wang, *An ADMM algorithm for a class of total variation regularized estimation problems*. IFAC Proceedings Volumes, 45(16), pp.83–88, 2012.

[14] A. Beck, M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*. SIAM journal on imaging sciences, 2(1), pp.183–202, 2009.

[15] D. Kazantsev, F. Bleichrodt, T. Van Leeuwen, A. Kaestner, P. J. Withers, K. J. Batenburg and P. D. Lee, *A novel tomographic reconstruction method based on the robust Student's t function for suppressing data outliers*. IEEE Transactions on Computational Imaging, 3(4), pp.682–693, 2017.

[16] D. Kazantsev, E. Pasca, M. J. Turner and P. J. Withers, *Ccpi-regularisation toolkit for computed tomographic image reconstruction with proximal splitting algorithms*. SoftwareX, 9, pp.317–323, 2019.

[17] Q. Xu, D. Yang, J. Tan, A. Sawatzky and M. A. Anastasio, *Accelerated fast iterative shrinkage thresholding algorithms for sparsity-regularized cone-beam CT image reconstruction*. Medical physics, 43(4), pp.1849–1872, 2016.

[18] P. Paleo, A. Mirone, *Ring artifacts correction in compressed sensing tomographic reconstruction*. Journal of synchrotron radiation, 22(5), pp.1268–1278, 2015.

[19] D. Kazantsev, V. Pickalov, S. Nagella, E. Pasca, P. J. Withers, *Tomo-Phantom, a software package to generate 2D-4D analytical phantoms for CT image reconstruction algorithm benchmarks*. SoftwareX, 7, pp.150–155, 2018.

[20] E. Guo, S. Shuai, D. Kazantsev, S. Karagadde, A. B. Phillion, T. Jing, W. Li and P. D. Lee, P.D, *The influence of nanoparticles on dendritic grain growth in Mg alloys*. Acta Materialia, 152, pp.127–137, 2018.

A Matched CBCT Projector-Backprojector Based on the Convolution of B-splines

Marion Savanier[1,2], Cyril Riddell[1], Yves Trousset[1], Emilie Chouzenoux[2] and Jean-Christophe Pesquet[2]

***Abstract*—Discretizing tomographic operators is a crucial step in the design of reconstruction algorithms. In particular, for iterative methods, the projector and backprojector are usually assumed to be matched to guarantee convergence. In practice, the reconstruction task is challenging, especially in cone-beam geometry. To this end, in the context of CBCT with a flat panel detector, we propose to rely on an interpolating kernel based on the convolution of B-splines taking into account the sampling of the detector and of the volume. Matched CBCT projector/backprojector based on resampling transforms are designed.**

***Index Terms*—CBCT, discretization, matched pair, projector, backprojector, B-spline, flat-panel.**

I. Introduction

FLAT-panel based C-arm systems are widely used imaging tools for image-guidance in interventional radiology and surgery. In addition, cone-beam computed tomography (CBCT) allows planning, guidance and control of the procedure. C-arm CBCT often suffers from poor sampling rate. The potential of reconstructing undersampled data with non linear iterative algorithms has been demonstrated, for instance to reduce angular and cone-beam undersampling artifacts [1]. One may intentionally decrease the sampling to reduce the X-ray dose. In this context, we are particularly interested in least-squares based optimization criteria regularized with nonlinear terms [2]. Most penalized least-squares algorithms require the use of the forward operator (i.e., projector), and the associated backward operator (i.e., backprojector). The projector must encode the C-arm projective geometry as accurately as possible. Furthermore, an underlying symmetry assumption is made, as the backprojector is assumed to coincide with the algebraic adjoint of the projector. Most forward models assume a geometric voxel shape (e.g. cubic) and its projection onto the detector. When ray-tracing is chosen, the backprojection implementation is not efficient, so that alternative models have been proposed, namely the distance-driven [3] and the separable-footprint [4]. These implementations are generic enough to be used with CT scanners curved detector and C-arm flat panel. In the context of flat panel CBCT, the geometry can be described with projection matrices. These matrices continuously relate any point in space to any point over the detection surface. This naturally leads to modeling the volume as samples in space and the measurements as samples over the detector. Voxels are considered through their centers and no shape is assumed. Very efficient backprojection schemes, using resampling transforms, are derived and used for standard analytical reconstruction, where backprojection is required only. Unfortunately, they cannot be represented as being the adjoint of any of the aforementioned forward models, and thus do not secure the convergence of standard iterative minimization schemes. It is worth pointing out that, in practice, the combination of resampling transforms for backprojection and ray-tracing for forward projection has been used [5], with apparently limited consequences despite the violation of the symmetry condition. The main reason may be that most iterative schemes are run for few iterations only, so that convergence issues may not be detected. Resampling transforms have also been used for the forward projection and here again, symmetry remains an issue [6].

Optimal resizing based on the convolution of B-splines has been extensively studied by Unser et al. [7]. In this paper, we propose to extend this approach to the discretization of resampling homographies, the transforms involved in CBCT with flat panel. We thus derive novel matched pairs of projector/backprojector that allow for precise and efficient forward and backward modeling while keeping symmetry.

In Section II, we recall the decomposition of cone-beam projection/backprojection with projection matrices as a series of homographies and introduce the proposed convolution scheme. Numerical experiments are then provided and discussed in Section III to evaluate the image quality associated with these novel tomographic operators.

II. Proposed method

A. Projector and backprojector pair

Let $\mathbf{f} = (f_l)_{1\leq l\leq L} \in \mathbb{R}^L$ denote the unknown volume attenuation with L voxels, $\mathbf{p} = (p_k)_{1\leq k\leq K} \in \mathbb{R}^K$, the log-transformed of the K detector cell measurements acquired on a C-arm system, and $\mathbf{R} \in \mathbb{R}^{K\times L}$ the forward projection model, whose entries $(R_{k,l})_{1\leq k\leq K, 1\leq l\leq L}$ relate each detector cell k to each voxel l. Reconstruction aims to estimate volume $\mathbf{f}$ from the knowledge of $\mathbf{p}$ and $\mathbf{R}$, based on the linear observation model:

$$\mathbf{p} = \mathbf{R}\mathbf{f} + \mathbf{n}, \tag{1}$$

where $\mathbf{n}$ is an additive noise. Iterative penalized least-squares algorithms minimize the sum of a data fidelity term $\frac{1}{2}\|\mathbf{R}\mathbf{f} - \mathbf{p}\|_2^2$ and a regularizer embedding prior information, such as image smoothness or range constraints. They usually account for the least-squares term through its gradient of the form $\mathbf{B}(\mathbf{R}\mathbf{f} - \mathbf{p})$, where $\mathbf{B} \in \mathbb{R}^{L\times K}$ denotes the backprojection operator. Ideally, we should have $\mathbf{B} = \mathbf{R}^\top$. However, as

(1): GE Healthcare, Buc, France. E-mail: first.last@ge.com
(2): Univ. Paris-Saclay, CentraleSupélec, CVN, Inria, Gif-sur-Yvette, France. E-mail: first.last@centralesupelec.fr
This work was supported by the ANRT CIFRE Convention 2018/1587.

we will discuss hereafter, it may happen that the latter equality does not hold, possibly leading to convergence instabilities and image quality deterioration.

B. Cone-beam geometry

The projection geometry is defined by the relationship between voxel coordinates (x, y, z) and the coordinates of the projected pixels (u, v). In CBCT with a flat-panel detector, the projector is characterized by a set of 3×4 projection matrices. There is one projection matrix per position of the imaging system. For a given projection matrix $\boldsymbol{P}$, the coordinates (u, v) of the projection of (x, y, z) onto the detection plane Π can be written with homogeneous coordinates (su, sv, s), with $s \in \mathbb{R}$, as $\left(su, sv, s\right)^\top = \boldsymbol{P}\left(x, y, z, 1\right)^\top$. The projection operation being separable, its discretization can be decomposed into 1D homographies that relate one free variable in space, e.g. x, to another one over the detector, e.g. u. For instance,

$$\left(su, sv_0, s\right)^\top = \boldsymbol{P}\left(x, y_0, z_0, 1\right)^\top \tag{2}$$

reduces to

$$\left(su, s\right)^\top = \boldsymbol{H}\left(x, 1\right)^\top \tag{3}$$

with $\boldsymbol{H} = (h_{s,t})_{1 \le s,t \le 2}$ a 2×2 homographic matrix. This operation can be analytically recast as the 1D homographic function h:

$$u = h(x) = \frac{h_{1,1}x + h_{1,2}}{h_{2,1}x + h_{2,2}}. \tag{4}$$

Without loss of generality, we shall thus treat the projected data p and original data f as continuous 1D signals. Backprojection resamples f from p and reprojection resamples p from f according to $p(u) = p \circ h(x) = f \circ h^{-1}(u) = f(x)$.

Homography h captures the continuously varying sampling scale by the magnification factor given by derivative $h'(x)$. In our clinical context, the geometry has a large focal distance with respect to the field of view so that h is never singular and the magnification factor varies slowly around $h'(0) = \det \boldsymbol{H}/h_{2,2}^2$. Consequently, the inverse homography h^{-1} behaves in a similar way. The discretization defines index i spanning the subset of detector cell locations $(u_i)_{1 \le i \le I}$ and index j spanning the subset of voxel locations $(x_j)_{1 \le j \le J}$ that are seen through h, giving rise to the discrete signals $\mathbf{p} = (p(u_i))_{1 \le i \le I}$ and $\mathbf{f} = (f(x_j))_{1 \le j \le J}$.

To interpolate, one decomposes signals p and f onto a set of continuous basis functions. A famous family of basis functions is that of B-splines of order $n \in \mathbb{N}$ denoted by $\beta^n : \mathbb{R} \to \mathbb{R}$. The associated centered B-spline of order n and scale $\delta > 0$ is defined by $\beta^n_\delta = \beta^n(\cdot/\delta)$. The simplest B-spline, of order 0, leads to nearest neighbor interpolation, while order 1 corresponds to linear interpolation. For a given h, one defines $\delta_1 = 1$ and $\delta_2 = \delta_1 h'(0)$ as the sampling steps for f and p, respectively.

Backprojection operator $\mathbf{B}$ computes the values $(f_j)_{1 \le j \le J}$ from $(p_i)_{1 \le i \le I}$ of $\mathbf{p}$ using $p(u) = \sum_{i=1}^I p_i \beta^n_{\delta_2}(u - u_i)$ so that $B_{j,i} = \beta^n_{\delta_2}(h(x_j) - u_i)$. Conversely, projection operator $\mathbf{R}$ uses $f(x) = \sum_{j=1}^J f_j \beta^n_{\delta_1}(x - x_j)$ from the components $(f_j)_{1 \le j \in J}$ of $\mathbf{f}$ so that $R_{i,j} = \beta^n_{\delta_1}(h^{-1}(u_i) - x_j)$. To get a matched pair, one would replace $\mathbf{B}$ with $\mathbf{R}^\top$ which gives $f(x_j) = \sum_{i|j \in \Omega_i} p_i \beta^n_{\delta_1}(h^{-1}(u_i) - x_j)$ with Ω_i the set of indices j such that for a given i, $R_{i,j} \neq 0$. This expansion appears as a poor choice for computing f_j in a direct manner because it uses the basis functions $\beta^n_{\delta_1}$ instead of $\beta^n_{\delta_2}$. As a result, detector cells contribute with gaps or redundancies to image pixels, which leads to high frequency artifacts.

C. Convolution-based basis function

We now assume that p and f are decomposed onto sets of B-splines of possibly different orders, $n \in \mathbb{N}$ for p and $m \in \mathbb{N}$ for f.

To overcome the previously described limitation, we propose to use a basis function suitable for both projection and backprojection, the latter requiring to capture $\beta^n_{\delta_1} \circ h^{-1}(u)$. For this purpose, we define the interpolating kernel $\phi^{m,n}_{\delta_1,\Delta_{1,i}}$ made of the normalized convolution of basis functions $\beta^m_{\delta_1}$ and $\beta^n_{\Delta_{1,i}}$ with $\Delta_{1,i} = (h^{-1})'(u_i)\delta_1$. Without loss of generality, we set $\delta_1 = 1$, and $\Delta_{1,i}$ is interpreted as a sampling parameter for the variation of magnification induced by the homography. The projection sampled at locations $(u_i)_{1 \le i \le I}$ now reads:

$$\begin{aligned} p_i = f(h^{-1}(u_i)) = (\mathbf{R}\mathbf{f})_i &= \sum_{j=1}^J f_j \phi^{m,n}_{1,\Delta_{1,i}}(h^{-1}(u_i) - x_j) \\ &= \sum_{j=1}^J \frac{f_j}{\Delta_{1,i}} \left(\beta^n_{\Delta_{1,i}} * \beta^m_1\right)(h^{-1}(u_i) - x_j) \end{aligned} \tag{5}$$

with $R_{i,j} = \frac{1}{\Delta_{1,i}} \left(\beta^n_{\Delta_{1,i}} * \beta^m_1\right)(h^{-1}(u_i) - x_j)$. From the general scaling property, for every $\Delta > 0$, $\left(\beta^n_\Delta * \beta^m_1\right)(u) \equiv \Delta \left(\beta^n_1 * \beta^m_{\frac{1}{\Delta}}\right)(u/\Delta)$. This allows us to deduce the adjoint of matrix $\mathbf{R}$ aiming at computing the backprojection at locations $(x_j)_{1 \le j \le J}$:

$$\begin{aligned} f_j = (\mathbf{R}^\top \mathbf{p})_j &= \sum_{i=1}^I \frac{p_i}{\Delta_{1,i}} \left(\beta^n_{\Delta_{1,i}} * \beta^m_1\right)(h^{-1}(u_i) - x_j) \\ &= \sum_{i=1}^I p_i \left(\beta^n_1 * \beta^m_{\frac{1}{\Delta_{1,i}}}\right)\left(\frac{h^{-1}(u_i) - x_j}{\Delta_{1,i}}\right) \\ &= \sum_{i=1}^I p_i \left(\beta^n_1 * \beta^m_{\frac{1}{\Delta_{1,i}}}\right)\left(u_i - \frac{x_j}{\Delta_{1,i}}\right). \end{aligned} \tag{6}$$

Similarly, we can express $\mathbf{B}$ using the normalized convolution $\frac{1}{\Delta_{2,j}} \left(\beta^m_{\Delta_{2,j}} * \beta^n_1\right)$. The backprojected function at locations $(u_j)_{1 \le j \le J}$ can then be computed as

$$\begin{aligned} f_j = p(h(x_j)) &= \sum_{i=1}^I \frac{p_i}{\Delta_{2,j}} \left(\beta^n_1 * \beta^m_{\Delta_{2,j}}\right)(h(x_j) - u_i) \\ &= \sum_{i=1}^I \frac{p_i}{\Delta_{2,j}} \left(\beta^n_1 * \beta^m_{\Delta_{2,j}}\right)(\Delta_{2,j}x_j - u_i). \end{aligned} \tag{7}$$

The coefficients of matrix $\mathbf{B}$ are thus $B_{j,i} = \frac{1}{\Delta_{2,j}} \left(\beta^n_1 * \beta^m_{\Delta_{2,j}}\right)(h(x_j) - u_i)$. From (7) and (6), we still have $\mathbf{B}^\top \neq \mathbf{R}$, but, up to a normalization factor, the difference now comes from our linear approximation to the homography h, quantified by the set of magnification

factors $(\Delta_{2,j})_{1\leq j\leq J}$ and $(1/\Delta_{1,i})_{1\leq i\leq I}$. The precision of our model relies on two factors, namely the order of the B-spline[1], and the set of magnifications. The former point is well documented and one can easily select the best trade-off between precision and complexity for a given application. Regarding magnification factors, neither the projector nor the backprojector is worthy to be privileged. In (6), $\Delta_{1,i}$ could be replaced by $\Delta_{2,j}$. The largest set of magnifications should be chosen to better capture the changes in sampling rates induced by h and h^{-1}.

D. Implementation

Since detectors are composed of cells, we choose to represent p with zero order B-spline, β_1^0 i.e. $n=0$. We do not restrict the volume to be a set of cubic voxels and we keep the flexibility $m\in\{0,1\}$ for f. Let us set $\Delta>0$. The scaled convolution kernel $\phi_{1,\Delta}^{m,0}$ is such that

$$\phi_{1,\Delta}^{m,0}(\ell)=\frac{1}{\Delta}(\beta_1^m*\beta_\Delta^0)(\ell)=\frac{1}{\Delta}\int_{-\infty}^{+\infty}\beta_1^m(\tau)\beta_\Delta^0(\ell-\tau)d\tau, \tag{8}$$

for $\ell\in\mathbb{R}$. The latter kernel gives the interpolation values at $(x_j)_{1\leq j\leq J}$ and $(u_i)_{1\leq i\leq I}$. In addition, if B-splines of order 1 are chosen for function f (i.e. $m=1$), once the entries of $\mathbf{f}=(f(x_j))_{1\leq j\leq J}$ have been computed with $\mathbf{B}$, a digital post-filter to $\mathbf{f}$ corresponding to the cubic spline interpolation filter must be applied to $\mathbf{f}$ as demonstrated in [7].

Given our choices for n and m, we are interested in two interpolating kernels: $\phi_{1,\Delta}^{0,0}$ and $\phi_{1,\Delta}^{1,0}$. Before deriving their explicit formula from (8), let us recall the formula of the B-splines of order 0 and 1, for every $\ell\in\mathbb{R}$:

$$\beta^0(\ell)=\begin{cases}1 & \text{if } |\ell|<\frac{1}{2}\\ 0 & \text{otherwise,}\end{cases}$$

$$\beta^1(\ell)=\begin{cases}1-|\ell| & \text{if } |\ell|<1\\ 0 & \text{otherwise.}\end{cases}$$

- **Case 1**

$$\phi_{1,\Delta}^{0,0}(\ell)=\begin{cases}\min(1,\Delta)/\Delta & \text{if} \quad |\ell|<a_1\\ \frac{1}{\Delta}(a_2-|\ell|) & \text{if} \quad a_1\leq|\ell|<a_2\\ 0 & \text{if} \quad |\ell|\geq a_2\end{cases}$$

  with $a_1=\frac{|\Delta-1|}{2}$, and $a_2=\frac{\Delta+1}{2}$.

- **Case 2**

$$\phi_{1,\Delta}^{1,0}(\ell)=\begin{cases}c_{k,0}+c_{k,1}|\ell|+c_{k,2}\ell^2 & \text{for } |\ell|\in[a_{k-1},a_k)\\ & \text{and } k\in\{1,2\}\\ 0 & \text{otherwise}\end{cases}$$

  with $a_0=0$, $a_1=|\frac{\Delta}{2}-1|$, $a_2=\frac{\Delta}{2}+1$ and expressions for $(c_{k,0},c_{k,1},c_{k,2})$ are given in Tab. I.

The approach of lowest order is comparable in complexity to linear interpolation and is thus simpler than original distance-driven. Kernels of higher degree increase computation time proportionally to their larger support.

[1]None of the above calculations rely on the fact that the basis functions are B-splines. Thus, one could easily choose another class of functions, noting that B-splines functions enjoy the minimal support for a given approximation order and the maximal approximation order for a given support.

Interval	$c_{k,0}$	$c_{k,1}$	$c_{k,2}$
$\lvert\ell\rvert<a_1$			
if $\Delta\leq 2\wedge\lvert\ell\rvert<\Delta/2$	$1-\Delta/4$	0	$-\frac{1}{\Delta}$
if $\Delta\leq 2\wedge\lvert\ell\rvert\geq\Delta/2$	1	-1	0
if $\Delta>2$	$1/\Delta$	0	0
$a_1\leq\lvert\ell\rvert<a_2$			
if $\lvert\ell\rvert\geq\Delta/2$	$(\Delta+4+4/\Delta)/8$	$-\frac{1}{\Delta}+\frac{1}{2}$	$-\frac{1}{2\Delta}$
if $\lvert\ell\rvert<\Delta/2$	$(-\Delta+4+4/\Delta)/8$	$-\frac{1}{\Delta}-\frac{1}{2}$	$\frac{1}{2\Delta}$

TABLE I: B-spline parameters for case 2

III. Experiments

A. Simulation context

To prove the concept of our approach, we evaluated our projector and backprojector as single modules and within an iterative reconstruction task. We computed simulated noise-free data of geometrical phantoms in CBCT geometry of half cone angle of 20°, as can be found on clinical C-arm systems. The detector bins are sampled on a finer grid than the voxels leading to a magnification of factor $\sqrt{2}$. Linear interpolation was taken as a baseline, for projector/backprojector construction and the distance-driven pair (DD) was added to the comparison. First, we analyzed the performance of matrix $\mathbf{R}$ for the B-splines kernels when projecting a centered uniform cylinder of attenuation 0.1 per voxels of diameter 80 voxels over 360°. The scanned object being invariant by rotation, we quantified the rotation invariance by computing the root mean-square error (RMSE) and the maximum error (MAE) between each profile over the range $[0°,45°]$ and profile at angle 0°. Then, for the same task, we compared the performance of the transpose of matrix $\mathbf{B}$. Finally we evaluated the spatial resolution associated to the different interpolation kernels using the modulation transfer function (MTF) on the reconstruction of a cube from 600 projections over 360°. To avoid the inverse crime, the forward-projection data was calculated with a 3000×20 pixel detector that was rebinned to 750×20 pixels. An FDK reconstruction was first performed to investigate the quality of $\mathbf{R}^\top$ and $\mathbf{B}$ as backprojectors. Then an iterative reconstruction was performed by applying 500 Landweber iterations with $\mathbf{f}^{(0)}$ chosen as the zero vector:

$$\mathbf{f}^{(n+1)}=\mathbf{f}^{(n)}-\tau\mathbf{R}^\top(\mathbf{R}\mathbf{f}^{(n)}-\mathbf{p}),\quad n\in\mathbb{N}. \tag{9}$$

Parameter τ is chosen as $1.9/\|\mathbf{R}\|^2$. No regularization was added as we aim to highlight features that are intrinsic to the model. This experiment gives an insight on the largest singular values of operator $\mathbf{R}$. Since the reconstruction methods investigated here are linear, spatial resolution at a slanted edge (tilting of 5°) is a suitable MTF.

B. Results

For the projection of the uniform cylinder, with $\mathbf{R}$, upon visual inspection of plotted profiles, the results are very similar for all interpolation schemes (plots not shown). For both invariance metrics, the convolution kernel $\phi^{1,0}$ combined with

a digital filter provides slightly better results: RMSE is 0.008 and MAE is 0.044 at angle $0°$ and monotonically increases up to 0.178 and 1.052 at $45°$ while the other schemes perform almost identically (RMSE of 0.04 and MAE of 0.213 at $0°$, RMSE of 0.185 and MAE of 1.09 at $45°$). With $\mathbf{B}^\top$, the profile of the projected cylinder at the angle $45°$ resulting from the transpose of linear interpolation is clearly distinguishable as it is degraded by high-frequency oscillations, which do not appear with any of the other convolutional basis functions (profiles not shown). The RMSE and MAE scores corroborate this observation.

Figure 1 displays the MTF curves obtained for the FDK reconstruction task with $\mathbf{B}$. For this task, $\phi_{\cdot}^{0,0}$ and DD scheme

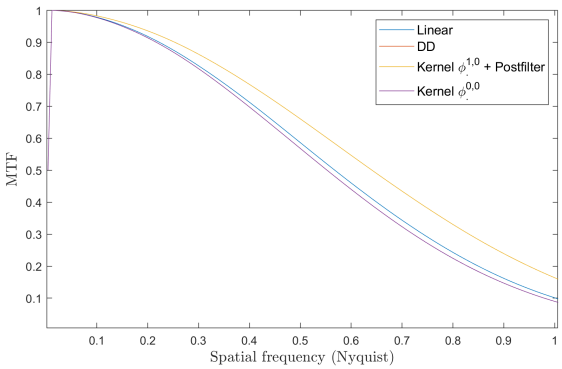


Fig. 1: MTF curve for $\mathbf{B}$

provides superimposed MTF curves just below the curve obtained with linear interpolation. Kernel $\phi_{\cdot}^{1,0}$ combined with a digital filter for the backprojection produces better spatial resolution. We further quantitatively compare their MTF curves through frequency $\nu_{0.2}$ for which the MTF reaches value 0.2. For kernel $\phi_{\cdot}^{1,0}$, $\nu_{0.2}$ is increased by 0.1. Then for FDK reconstruction with $\mathbf{R}^\top$, the same frequencies $\nu_{0.2}$ are tabulated in Table II. All MTF curves are close with a slight advantage for $\phi_{\cdot}^{0,0}$.

	Linear	DD	$\phi_{\cdot}^{0,0}$	$\phi_{\cdot}^{1,0}$
$\nu_{0.2}$	0.808	0.824	0.824	0.793

TABLE II: Spatial frequency $\nu_{0.2}$ at 20% MTF for FDK with $\mathbf{R}^\top$

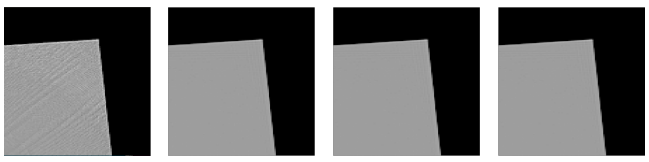


Fig. 2: Zoom on corner of reconstructed slice. From left to right: Linear interpolation, DD, $\phi_{\cdot}^{0,0}$, $\phi_{\cdot}^{1,0}$.

Finally, after 500 Landweber iterations, the reconstructed square with linear interpolation is substantially different than those obtained with the other models as shown on Figure 2. Undesirable interpolation patterns are visible with linear interpolation. This indicates that some non-null singular values of such $\mathbf{R}$ are of very low magnitude, yielding a reconstruction with details similar to discretization errors. For the other models, frequencies $\nu_{0.2}$ are reported in Table III. All MTF curves are almost superimposed: $\phi_{\cdot}^{0,0}$, $\phi_{\cdot}^{1,0}$ and DD scheme produce sharp reconstructions, comparable to that of FDK.

	DD	$\phi_{\cdot}^{0,0}$	$\phi_{\cdot}^{1,0}$
$\nu_{0.2}$	0.8945	0.8945	0.8711

TABLE III: Spatial frequency $\nu_{0.2}$ at 20% MTF for iterative reconstruction

These preliminary results suggest that the transpose of the linear interpolation-based backprojection does not qualify as a quality projection model. By using a refined discretization of the resampling transform through our convolution kernels, the performance of the transpose becomes similar to that of direct models for the basic tests given here, even with the lowest order of B-spline. Our results indicate that the kernel obtained with B-splines of order 0 perform as well as the distance-driven pair in a consistent fashion. This scheme is based on a similar rationale as ours, but is not flexible as it does not provide means for higher–order interpolation. We did not get consistent MTF increases when raising the degree of the spline modeling the volume. Further experiments are needed to tell whether it can improve reconstruction of real data at a reasonable computation cost. As advocated in [6], iterative reconstruction could have been performed in a virtual geometry, called rectified geometry. On the one hand, it incurs an additional resampling step that may degrade precision. On the other hand, homographies reduce to magnifications. The resampling task is simplified into that of the resizing problem for which the B-splines approach chosen here benefits from strong theoretical properties [7], especially in terms of optimal approximation.

IV. Conclusion

A novel interpolation scheme based on B-spline convolution has been proposed to discretize the projector and backprojector, in the context of CBCT with a flat panel detector. The expected benefits of this approach is improved reconstruction for various methods involving a matched pair of projector/backprojector requiring high-quality interpolation.

References

[1] H. Langet, C. Riddell, A. Reshef, Y. Trousset, A. Tenenhaus, E. Lahalle, G. Fleury, and N. Paragios, "Compressed-sensing-based content-driven hierarchical reconstruction: Theory and application to c-arm cone-beam tomography," *Medical Physics*, vol. 42, Aug. 2015.

[2] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame based inverse problems," *Inverse Problems*, vol. 23, no. 1, Jun. 2007.

[3] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," *Physics in medicine and biology*, vol. 49, July 2004.

[4] Y. Long, J. A. Fessler, and J. M. Balter, "3D forward and back-projection for x-ray ct using separable footprints," *IEEE Transactions on Medical Imaging*, vol. 29, no. 11, Nov. 2010.

[5] R. R. Galigekere, K. Wiesent, and D. W. Holdsworth, "Cone-beam reprojection using projection-matrices," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, Oct. 2003.

[6] C. Riddell and Y. Trousset, "Rectification for cone-beam projection and backprojection," *IEEE Transactions on Medical Imaging*, vol. 25, no. 7, July 2006.

[7] M. Unser, A. Aldroubi, and M. Eden, "Enlargement or reduction of digital images with minimum loss of information," *IEEE Transactions on Image Processing*, vol. 4, no. 3, March 1995.

Structured Regularization for Material Decomposition of Photon Counting CT Data using Collaborative Total Variation

Eckhard Wehrse, Stefania Petra, Laura Klein, Lukas Rotkopf, Christian Herbert Ziener, Marc Kachelrieß, Heinz-Peter Schlemmer, and Stefan Sawall

***Abstract*—Photon Counting CT (PC CT) is of great interest in medical imaging as it improves spatial resolution, provides higher contrast-to-noise ratios and reduced electronic noise and allows the simultaneous acquisition of multiple energy bin detector data. Although PC CT can support a higher number of energy bins compared to Dual Energy CT, multimaterial decomposition remains an ill-posed problem for many practical applications. Prior information about the involved materials, such as conservation of mass, smoothness and sparsity in the material maps can be used for regularization. As an extension of the widely-used regularization method of total variation (TV), collaborative TV (CTV) provides a regularization for vector-valued images. In this work, we apply CTV to iteratively denoise material images of an abdomen phantom by coupling material image derivatives. The results suggest that material image quality benefits from a stronger energy bin coupling.**

I. Introduction

PC CT enables tissue characterization and the visualization of multiple contrast agents by measuring the number and the energy of detected photons. For instance, virtual noncalcium and noncontrast-enhanced images have been applied for the evaluation of bone marrow edema, fatty liver disease, liver fibrosis [1] [2]. It is assumed that when using a linear measurement model, the number of basis material images to be accurately reconstructed should not exceed the number of energy bins. Moreover, the condition of the reconstruction problem suffers from spectral similarity of the involved materials and technical limitation in the correct assignment to energy bins, which lowers the quality and practical benefit of material images. Therefore, regularization aims at meaningful and reliable material maps, where textures are not oversmoothed [3] and gradual changes in material concentrations are correctly displayed. The authors of [2] proposed an image-domain multimaterial decomposition (MMD) method for dual-energy (DE) CT which adds prior information to the compromised spectral data. By implementing the conservation of mass assumption, they increase the number of basis materials by one. Additionally, it is assumed that not more than three basis materials are present in a pixel and can be described by an empirically determined sequence of material triplets. By introducing a penalized weighted least square (PWLS) method [4], the noise characteristics in the energy bins can be taken into account. The same authors denoise and promote sparsity in the material images by implementing l^0 and total nuclear variation (TNV)-based regularization terms. The TNV norm is chosen because it is the unique convex envelope of the rank of the image's gradient [5], promoting common boundaries of the material images. By introducing CTV as a framework of vectorial total variation (VTV), the authors of [6] and [7] generalize TNV and TV as proposed by [8] for single-color images.
In this paper, we investigate regularization terms of the CTV class for material images by exchanging the noise reduction term in the PWLS-TNV-l^0 method. This leads to a more general variational model for material decomposition and denoising. Additionally, we compare our results with Frequency Split (FS), a recently proposed denoising method based on a linear combination of low- and high-pass filtered material images [9]. As proposed by [4], we apply the alternating direction method of multipliers (ADMM) method to split the overall minimal cost function problem into subproblems, which can be solved effectively by the appropriate proximal operators.

Eckhard Wehrse, Laura Klein, Lukas Rotkopf, PD Dr. Dr. Christian Herbert Ziener, Prof. Dr. Marc Kachelrieß, Prof. Dr. Heinz-Peter Schlemmer and Dr. Stefan Sawall are with the German Cancer Research Center (Dkfz), Heidelberg, Germany and with the University of Heidelberg, Germany. Email: e.wehrse@dkfz.de. Prof. Dr. Stefania Petra is with the Mathematical Imaging Group, University of Heidelberg, Germany.

II. Methods and Materials

This section introduces our variational material decomposition and denoising model that builds on the PWLS-TNV-l^0 method from [4]. Let N_p, ν and $L_0 \in \mathbb{N}$ denote the number of image voxels, of energy bins and of material images, which are denoted by $\boldsymbol{X} \in \mathbb{R}^{L_0 \times N_p}$. For known linear attenuation coefficients $\mathbf{A}_0 \in \mathbb{R}^{\nu \times L_0}$ the PC CT value $\boldsymbol{y}_i \in \mathbb{R}^{\nu}$ of a voxel $i \in N_p$ is determined by

$$\boldsymbol{y}_i = \mathbf{A}_0 \boldsymbol{x}_i \ , \tag{1}$$

where $\boldsymbol{x}_i$ denotes the i-th column of $\boldsymbol{X}$.
We use $\mathbf{A} = \mathbf{A}_0 \otimes \mathbf{I}_{N_p} \in \mathbb{R}^{\nu N_p \times L_0 N_p}$, with the identity matrix $\mathbf{I}_{N_p} \in \mathbb{R}^{N_p \times N_p}$ and the vectorized representation $\boldsymbol{x} \in \mathbb{R}^{L_0 N_p}$ and $\boldsymbol{y} \in \mathbb{R}^{\nu N_p}$ of $\boldsymbol{X}$ and the measured PC CT data in image domain $\boldsymbol{Y} \in \mathbb{R}^{\nu \times N_p}$. Assuming that the reconstructed image in the spatial domain is corrupted by additive noise $\boldsymbol{\epsilon} \in \mathbb{R}^{\nu N_p}$

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{\epsilon} \ , \tag{2}$$

we employ a weighted least-squares data fidelity term as in [10] [11]

$$\bar{L}(\boldsymbol{x}) = \frac{1}{\nu}||\boldsymbol{y} - \mathbf{A}\boldsymbol{x}||^2_{2,\boldsymbol{\Sigma}^{-1}} \quad . \tag{3}$$

Here, we denote the variances in the energy bins by $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2\mathbf{I}_{N_p}, \sigma_2^2\mathbf{I}_{N_p}, \ldots, \sigma_\nu^2\mathbf{I}_{N_p})$. To make use of prior information about the material images, regularization terms are added to the energy functional, resulting in the variational model

$$\hat{\boldsymbol{x}} = \mathrm{argmin}_{\boldsymbol{x}}\{\bar{L}(\boldsymbol{x}) + R(\boldsymbol{x})\}, \tag{4}$$

where $R : \mathbb{R}^{\nu N p} \to \mathbb{R}_{\geq 0}$ consists of three different regularization terms

$$R(\boldsymbol{x}) = \beta_{\mathrm{noi}}\, R_{\mathrm{noi}}(\boldsymbol{x}) + \beta_{\mathrm{spa}}\, R_{\mathrm{spa}}(\boldsymbol{x}) + \beta_{\mathrm{mas}}\, R_{\mathrm{mas}}(\boldsymbol{x}) \quad . \tag{5}$$

$\beta_{\mathrm{noi}}, \beta_{\mathrm{spa}}, \beta_{\mathrm{mas}} \geq 0$ are user-defined weight regularization parameters that need to be carefully chosen. In the following we present these three terms in detail.

A. Noise reduction term

Since several materials are involved, we consider the material images' gradient as a tensor $\mathbf{Dx} \in \mathbb{R}^{N_p \times N_d \times L_0}$ for N_d denoting the number of spatial derivative directions [6]. We employ several mixed tensor norms that lead to different coupling of image derivatives and therefore to different penalization of noisy images. We consider two different types of norms for the noise reduction term R_{noi} in Eq. 5 above. The first is the $l^{p,q,r}$ norm

$$||\mathbf{Dx}||_{p,q,r} := \left(\sum_{i=1}^{N_p} \left(\sum_{j=1}^{N_d} \left(\sum_{k=1}^{L_0} |(\mathbf{Dx})_{i,j,k}|^p \right)^{\frac{q}{p}} \right)^{\frac{r}{q}} \right)^{\frac{1}{r}} \tag{6}$$

and the second is the $\|\cdot\|_{(S^p,l^q)}$ norm

$$\|\mathbf{Dx}\|_{(S^p,l^q)} := \left(\sum_{i=1}^{N_p} \left\| \begin{pmatrix} (\mathbf{Dx})_{i,1,1} & \ldots & (\mathbf{Dx})_{i,1,L_0} \\ \vdots & \ddots & \vdots \\ (\mathbf{Dx})_{i,N_d,1} & \ldots & (\mathbf{Dx})_{i,N_d,L_0} \end{pmatrix} \right\|^q_{S^p} \right)^{\frac{1}{q}} , \tag{7}$$

which is based on the Schatten-p-norm ($\|\cdot\|_{S^p}$), with $p, q, r \in [1, \infty]$, where ∞ denotes the maximum norm. To obtain the discontinuity-preserving properties of the TV [8], we select $r = 1$ in Eq. 6 and $q = 1$ in Eq. 7 and restrict ourselves to anisotropic CTV ($q = 1$ in Eq. 6). An alternative popular choice is the nuclear norm S^1, that is the convex envelope of a matrix rank S^0 (given as the l^0 quasinorm of the singular values) [5]. The TNV is considered optimal for enforcing aligned boundary directions, as it is generally assumed for piecewise-constant material images. Even stronger channel coupling can be promoted for $p \to \infty$ [6].

B. Sparsity promoting term

Most pixels can be described accurately by only a subset of materials. We therefore expect a meaningful solution $\boldsymbol{x}$ to be sparse. By applying the l^0 quasi-norm, which counts nonzero-entries of a vector, the second regularization term is imposed as

$$R_{\mathrm{spa}}(\boldsymbol{x}) = \sum_{j=1}^{N_p} \|\boldsymbol{X}(:, j)\|_0 =: \|\boldsymbol{x}\|_0 \quad . \tag{8}$$

Since l^0 is non-convex, the convergence of the applied ADMM is not proven [12], while l^p for $p \geq 1$ fulfills the triangular equation, favoring generally equal mass concentrations.

C. Non-Negativity and Mass Conservation

If the set of basis materials is adequate, the set of feasible solutions can be restricted to non-negative concentrations which sum up to one. While [2] increase the size of $\mathbf{A}_0$ and $\mathbf{y}_i$ by additional constant entries increasing the size of the PWLS (Eq. 3), we make use of the characteristic function

$$R_{\mathrm{mas}}(\boldsymbol{x}) = \chi_S(\boldsymbol{x}) = \begin{cases} 0, & \boldsymbol{X} \in S \\ \infty, & \boldsymbol{X} \notin S, \end{cases} \tag{9}$$

as also done by [4]. S denotes the feasible set for $\boldsymbol{X}$

$$S = \{\boldsymbol{X} \in \mathbb{R}^{L_0 \times N_p} : \sum_{l=1}^{L_0} X_{lj} = 1, 0 \leq X_{lj}, \quad j = 1, \ldots, N_p\}. \tag{10}$$

Each column of $\boldsymbol{X}$ represents the set of "barycentric," coordinates for the signal in each pixel j. Since S is a convex set, χ_S is a convex function [13].

D. Details on the ADMM algorithm

We set $\beta_{\mathrm{mas}} = 1$, $\beta_{\mathrm{spa}} = 10$ [14] and the maximal number of ADMM iterations to 2000. β_{noi} has to be optimized for each CTV norm first to allow for a fair comparison. This is achieved by minimizing the Root-Mean-Squared-Error (RMSE) as given by Eq. 11 below for a small image part (yellow square in Fig. 1). In a second step, the optimized parameters are applied to the greater image demonstrating the entire vertebral body (green square in Fig. 1). We repeated this optimization process for the same set of parameters also for the PWLS-TNV-l^0 method [4], which we denote as S^1l^1 in the context of CTV.

III. Results

A. Evaluation Metric

For the reference $\boldsymbol{x}_{\mathrm{ref}} \in \mathbb{R}^{L_0 N_p}$, the MMD algorithm is applied to the mean value of fifty adjacent and invariant slices in longitudinal direction. We compare differences of resulting bone images $\boldsymbol{x}_{\mathrm{bone}} \in \mathbb{R}^{N_p}$ to $\boldsymbol{x}_{\mathrm{ref,bone}} \in \mathbb{R}^{N_p}$ by the RMSE and all material images to the combined $\mathrm{RMSE}_{\mathrm{comb.}}$ as follows

$$\mathrm{RMSE}_{\mathrm{comb.}} = \frac{\|\boldsymbol{x}_{\mathrm{bone}} - \boldsymbol{x}_{\mathrm{bone,ref}}\|_2}{\sqrt{N_p \cdot L_0}}, \mathrm{RMSE} = \frac{\|\boldsymbol{x} - \boldsymbol{x}_{\mathrm{ref}}\|_2}{\sqrt{N_p}} \quad . \tag{11}$$

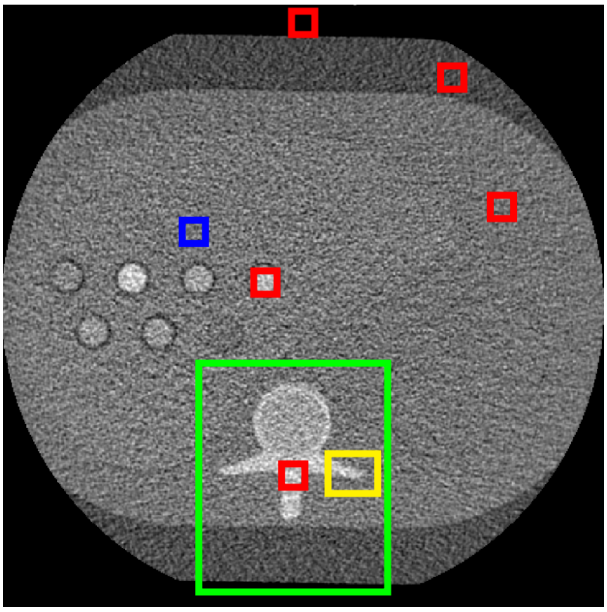


Figure 1: Energy bin image $[75, 120]\,\mathrm{keV}$, window $\mathrm{C} = 100\,\mathrm{HU}, \mathrm{W} = 900\,\mathrm{HU}$. Red squares: ROIs for averaged attenuation coefficents ($\mathbf{A}_0$). Blue square: ROI for standard deviations $\sigma_1, \ldots \sigma_4$. Green rectangle: Image part as seen in Fig. 2. Yellow rectangle: picture detail to optimize β_{noi} and β_{spa}.

B. Semi-anthropomorphic phantom study

A semi-anthropomorphic abdomen phantom (QRM GmbH, Möhrendorf, Germany) is measured by a SOMATOM CounT (Siemens Healthineers, Germany), an experimental PC CT system. Regions of interest (ROIs) are placed in the vertebral spine, in adipose and soft tissue, in an iodine vial (concentration $25\,\frac{\mathrm{mg}}{\mathrm{mL}}$) and into surrounding air (Fig. 1). The additional ROI in the liver ROI yields $\sigma_1, \ldots, \sigma_\nu$, to take photon statistics of the different energy bin data into account. For the acquisition a tube voltage of $120\,\mathrm{kV}$ with $200\,\mathrm{mAs}$ is chosen. These data are reconstructed with a D40f convolution kernel, pixel spacing set to $0.537\,\mathrm{mm} \times 0.537\,\mathrm{mm}$, and a slice thickness of $1\,\mathrm{mm}$. Fig. 2 (a)-(c) show the bone images of the MMD algorithm results, which have been obtained for different noise levels in the initial energy bin images data. The MMD result of the mean of seven slices (Fig. 2 (b)) shows a small RMSE to the reference (Fig. 2 (a)). Details like the compacta of the vertebral body and the clear boundary of the liver model are preserved. The RMSE increases monotonically for a lower number of slices (Fig. 2 (c)) while the visibility of smaller image details and the difference of compacta and spongiosa decreases. While the fast-performing filtering-based FS is able to reduce the RMSE to 0.13, the combination of CTV norms and sparsity promoting term (Eq. 8) reduce the RMSE to around 0.09, which corresponds to the MMD result for the mean of seven measurements (b). All regularization and denoising methods preserve sharp discontinuities like the boundaries of air and adipose tissue (not shown) or soft tissue and bone. Picture details like the left (anatomical) transverse and the spinous process are best visualized by the $l^{\infty,1,1}$-norm, while $l^{1,1,1}$ performs best in reducing the overall $\mathrm{RMSE}_{\mathrm{comb.}}$ (Table I). As Eq. 7 restricts the number of singular values by the number of derivative directions, the S^1l^1 and $S^\infty l^1$ norm demonstrate similar results. They do not show superior recovery characteristics at common boundaries. All applied denoising and regularization methods are not able to specify the correct composition of the spines spongiosa, a material of unknown spectral signature (no ROI as seen in Fig. 1).

Table I: $\mathrm{RMSE}_{\mathrm{comb.}}$

1 measur. (MMD [2])	0.1885
7 measur. (MMD [2])	0.1156
9 measur. (MMD [2])	0.1071
$l^{1,1,1}$	**0.1091**
$l^{\infty,1,1}$	0.1098
S^1l^1 (PWLS-TNV-l^0 [4])	0.1118
$S^\infty l^1$	0.1109

IV. Discussion of Experiments and Conclusion

Without making prior assumptions on piecewise constancy of the ground truth, we use MMD results of low-noise initial data as reference. This reference is also affected by the condition of matrix $\mathbf{A}$, leading to to a salt and pepper like noise on a small scale. All regularization and denoising methods are able to decrease noise levels and achieve material images that are equivalent to MMD results achieved with four times the radiation dose. Both in $l^{p,q,r}$ and (S^p, l^q) based CTV, the RMSE benefits slightly from a stronger coupling of material images. (S^p, l^q)-based CTV requires time-consuming singular value decompositions for each pixel and each ADMM iteration. As the $l^{\infty,1,1}$ norm also provides a strong coupling, reconstruction times for material images can be reduced. Further research could focus on whether differences between the CTV norms increase when the number of materials to be decomposed is larger. Most notably, we observe that MMD results from PC CT data benefit to a large extent from employing a variational denoising model that used a CTV based regularization term.

References

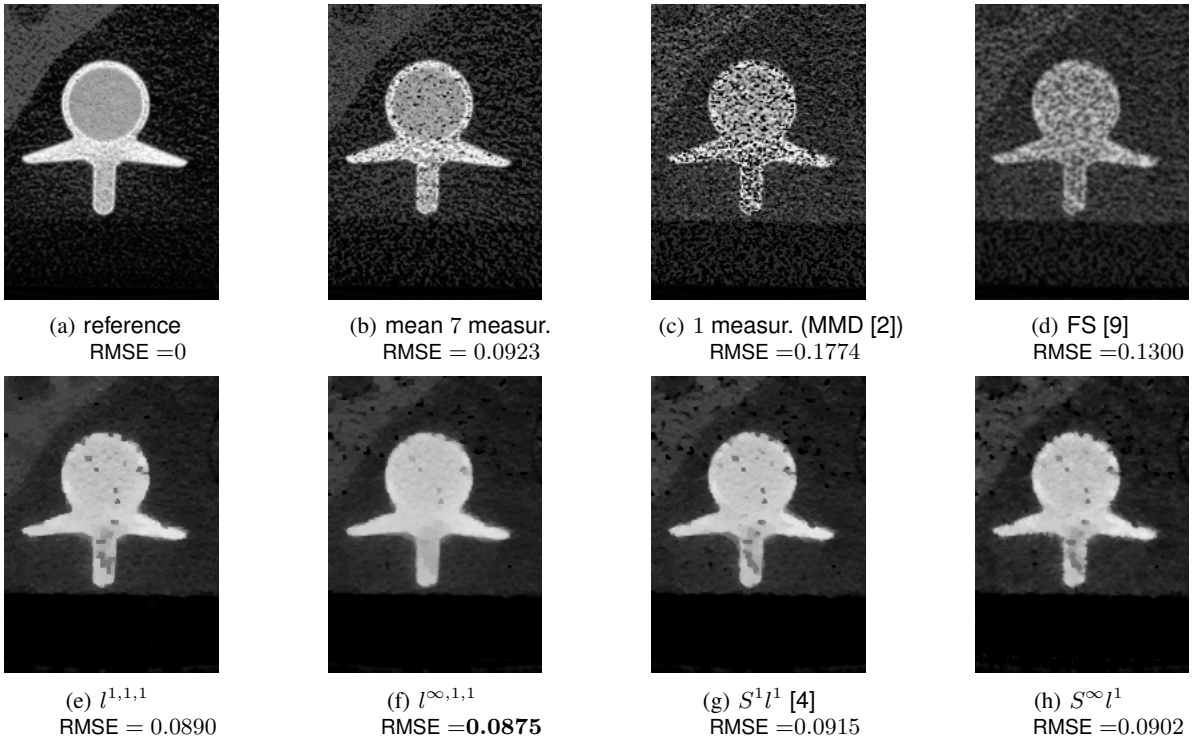
[1] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting ct: Technical principles and clinical prospects", *Radiology*, vol. 289, no. 2, pp. 293–312, 2018.

[2] P. R. Mendonca, P. Lamb, and D. V. Sahani, "A flexible method for multi-material decomposition of dual-energy CT images", *IEEE Transactions on Medical Imaging*, vol. 33, no. 1, pp. 99–116, 2014.

[3] L. Fan, F. Zhang, H. Fan, and C. Zhang, "Brief review of image denoising techniques", *Visual Computing for Industry, Biomedicine, and Art*, vol. 2, no. 1, p. 7, 2019.

[4] Q. Ding, T. Niu, X. Zhang, and Y. Long, "Image-domain multi-material decomposition for dual-energy CT based on correlation and sparsity of material images", *Medical Physics*, vol. 45, no. 8, pp. 3614–3626, 2018.

[5] K. M. Holt, "Total nuclear variation and jacobian extensions of total variation for vector fields", *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3975–3989, 2014.



(a) reference RMSE $=0$ | (b) mean 7 measur. RMSE $= 0.0923$ | (c) 1 measur. (MMD [2]) RMSE $=0.1774$ | (d) FS [9] RMSE $=0.1300$

(e) $l^{1,1,1}$ RMSE $= 0.0890$ | (f) $l^{\infty,1,1}$ RMSE $=\mathbf{0.0875}$ | (g) S^1l^1 [4] RMSE $=0.0915$ | (h) $S^\infty l^1$ RMSE $=0.0902$

Figure 2: Material image "bone" and the corresponding RMSE of all material images to the reference. First row: Results of the MMD algorithm for different noise levels in the PC CT data and result of FS [9]. The MMD solution of the average of fifty measurements (a) is taken as reference. Second row: Results of the PWLS-CTV-l^0 algorithm for $l^{p,q,r}$ and Schatten-p-norms (S^p, l^q) for different channel-couplings. For S^1l^1 the algorithm is identical to the PWLS-TNV-l^0 method [4]. C = 0.5, W = 1 (relative material concentration)

[6] J. Duran, M. Moeller, C. Sbert, and D. Cremers, "Collaborative total variation: A general framework for vectorial TV models", *SIAM Journal on Imaging Sciences*, vol. 9, no. 1, pp. 116–151, 2016.

[7] J. Duran, M. Moeller, C. Sbert, and D. Cremers, "On the implementation of collaborative tv regularization: Application to cartoon+ texture decomposition", *Image Processing On Line*, vol. 6, pp. 27–74, 2016.

[8] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms", *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[9] M. Özdemir, S. Dorn, F. Pisana, M. Uhrig, H.-P. Schlemmer, and M. Kachelrieß, "Image-based noise reduction for material decomposition in dual or multi energy computed tomography", vol. 10948, p. 109482U, 2019.

[10] D. S. Rigie and P. J. La Rivière, "Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization", *Physics in Medicine and Biology*, vol. 60, no. 5, pp. 1741–1762, 2015.

[11] D. Kazantsev, J. S. Jørgensen, M. S. Andersen, W. R. Lionheart, P. D. Lee, and P. J. Withers, "Joint image reconstruction method with correlative multi-channel prior for x-ray spectral computed tomography", *Inverse Problems*, vol. 34, no. 6, p. 064 001, 2018.

[12] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers", *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[13] R. T. Rockafellar, *Convex analysis*. Princeton university press, 1970, vol. 28.

[14] Q. Ding, T. Niu, X. Zhang, and Y. Long, "Image-domain multi-material decomposition for dual-energy CT via total nuclear norm and l0 norm", *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pp. 154–9, 2017.

Direct Iterative Reconstruction of Multiple Basis Material Images in Photon-counting Spectral CT

Obaidullah Rahman, Ken D. Sauer, Department of Electrical Engineering, University of Notre Dame
Connor J. Evans, Ryan K. Roeder, Department of Aerospace and Mechanical Engineering, University of Notre Dame

***Abstract*—In this work, we perform direct material reconstruction from spectral CT data using a model based iterative reconstruction (MBIR) approach. Material concentrations are measured in volume fractions, whose total is constrained by a maximum of unity. A phantom containing a combination of 4 basis materials (water, iodine, gadolinium, calcium) was scanned using a photon-counting detector. Iodine and gadolinium were chosen because of their common use as contrast agents in CT imaging. Scan data was binned into 5 energy (keV) levels. Each energy bin in a calibration scan was reconstructed, allowing the linear attenuation coefficient of each material for every energy to be estimated by a least-squares fit to ground truth in the image domain. The resulting 5×4 matrix, for 5 energies and 4 materials, is incorporated into the forward model in direct reconstruction of the 4 basis material images with spatial and/or inter-material regularization. In reconstruction from a subsequent low-concentration scan, volume fractions within regions of interest (ROIs) are found to be close to the ground truth. This work is meant to lay the foundation for further work with phantoms including spatially coincident mixtures of contrast materials and/or contrast agents in widely varying concentrations, molecular imaging from animal scans, and eventually clinical applications.**

***Index Terms*—Spectral CT, MBIR, mixing matrix, basis materials.**

I. Introduction

SPECTRAL CT has shown great promise for a variety of medical and scientific imaging tasks. The technology is enabled by the advent of photon-counting detectors (PCD), which can accurately count the number of received photons and bin them according to energy levels [1]. Energy bins can be adjusted to leverage the K-edge of some contrast agents for better delineation among basis materials. This also allows us to more closely model the attenuation effects of materials at differing energy levels, limiting the effects of beam hardening artifacts. This permits a more accurate linear model between image and sinogram domains.

Material decomposition is typically accomplished by reconstructing each energy bin independently as a monochromatic scan using a traditional algorithm such as FBP or an iterative technique such as model-based iterative reconstruction (MBIR). The resulting reconstructions are linear attenuation coefficient (LAC) images for respective energy levels which are then, through a "mixing matrix" describing linear attenuation by each material at each energy level, decomposed into basis material fraction images[2]. For greater accuracy than with material modeling from standard values, the mixing matrix can be calibrated using phantoms with known materials and concentrations[3]. Calibration is meant to account for system specific variations. In this work, we use a calibration dataset to compute our mixing matrix, $\mathcal{M}$ from individual LAC images of 5 bins and the ground truth. Then we use that $\mathcal{M}$ in the linear model to reconstruct material fraction images directly from subsequent scans containing similar materials.

The goal of this work is to detect, identify and quantify contrast agents within a phantom using MBIR directly from spectral CT sinogram data, and to lay a foundation for its eventual use in laboratory and clinical studies. Water, a natural choice for one of the basis materials, was chosen to simulate soft tissue. Iodine and gadolinium were chosen because of their wide usage as contrast agents in diagnostic imaging. Calcium was chosen to simulate mineralized tissue. We are attempting to estimate material fractions the order of 10^{-3} for contrast agents. The challenges in this work come from low signal contributions from the contrast which, with relatively large condition numbers of mixing matrices, poses a challenging estimation problem as the solutions become more dilute.

A great deal of research has been published in the area of spectral CT, a large fraction of it dealing with dual-energy systems [4], [5]. Direct reconstructions to date have typically used fewer energy bins [6] or fewer basis materials than our current efforts. Some have worked on more than 2 energies [7] but with 2 basis materials only. Our previous work has been focused on inversion from multi-energy attenuation coefficients to material concentrations under constrained optimization in the image domain[2]. The present work departs from previous in that we attempt direct reconstructions at low contrast among multiple contrast agents using a direct statistical iterative reconstruction algorithm in spectral CT.

II. Method

A. Background

1) MAP estimation for poly-chromatic case: Let $x = [x_l; l \in (1, L)]$ be the vector in image space. In conventional, monochromatic CT, x will represent linear attenuation coefficients for a single scan, while in our subsequent spectral case, it will represent concentrations of various materials in an expanded form. The vector $y = [y_n; n \in (1, N)]$ contains projection measurements. Let $A_{N \times L}$ be the projection matrix that maps linear attenuation to projection space according to scanner geometry. In the work presented here, A is computed according to the methods described in [8]. The MAP estimate of x for a single energy can be formulated by

$$\hat{x} = \underset{x}{\textit{argmin}}\ p(y|x)g(x), \tag{1}$$

with p and g representing conditional probability density of measurements and an *a priori* model of the image, respectively. The MAP estimate of x [9] can be approximated as

$$\hat{x} = \underset{x}{\textit{argmin}}\ \frac{1}{2}(y - Ax)^T D(y - Ax) + U(x) \tag{2}$$

where D is a weighting matrix approximately inverse to conditional variance of data and $U(x)$ is a log-prior on the image. With an iterative coordinate descent (ICD) scalar update, a $k-th$ step in optimization of (2) becomes

$$\hat{\Delta x_l^k} = \underset{\Delta x_l}{\textit{argmin}} \frac{1}{2}(e^k + A_{*l}\Delta x_l)^T D(e^k + A_{*l}\Delta x_l) + U_l(x + I_l \Delta x_l) \tag{3}$$

where Δx_l^k is the change in image value x_l at update k, $e^k = y - Ax^k$, the sinogram error vector at step k, A_{*l} represents the l^{th} column of A, I_l is zero except for unity at site l, and $x_l^{k+1} = x_l^k + \Delta x_l^k$. ICD may have advantages in convergence, and in our case, simpler computation under varied image models for $U(x)$ is possible.

2) MAP estimation for spectral case: MAP estimation for the polychromatic case can be easily extended to spectral data. In this case, $x = [x_l; l \in (1, L)]$ will again be the vector of image space with L voxels. However, each entry x_l is now an M-element vector such that $x_l = [x_{lm}; m \in (1, M)]$, where M is the number of basis materials. $x_{L\times M}$ will represent the *volume fraction* for each material, rather than linear attenuation. $y = [y_e; e \in (1, E)]$, where E is the number of energy bins. y_e represents e^{th} energy sinogram. Each entry in y_e is a N-element vector. $y_e = [y_{en}; n \in (1, N]$, where N is the number of sinogram datapoints in each energy bin. $y_{E\times N}$ represents total sinogram measurement. Let $A_{N\times L}$ be the standard projection matrix as before. x can be estimated by the following equation:

$$\hat{x} = \underset{x}{\textit{argmin}} \left[\frac{1}{2} \sum_{e=1}^{E} (y_e - A(\mathcal{M}x^T)^T)^T D(y_e - A(\mathcal{M}x^T)^T) + U(x) \right] \tag{4}$$

The dimension of y_e is $N \times 1$, $\mathcal{M}$ is $E \times M$, X^T is $M \times L$, $\mathcal{M}X^T$ is $E \times L$, D is $N \times N$, and $A(\mathcal{M}x^T)^T$ is $N \times 1$ which is same as y_e.

The vector ICD update for each voxel is identical to (3), except that x_l is a vector which is converted from material fractions to linear attenuation by $\mathcal{M}$ for forward projection by A before insertion into the log-likelihood term.

B. Optimization

1) Convexity: The Hessian of first cost term w.r.t. x_l is positive definite, since it is quadratic in x_l with a positive, diagonal weighting. We apply a non-negativity constraint on all material fraction values, and force the material fractions for a voxel to sum to no more than 1. The optimization problem now becomes the modified equation (3) with addition of the following constraints

$$\text{s.t., } x_l \geq \mathbf{0} \qquad x_l^T \mathbf{1} \leq 1 \tag{5}$$

where $\mathbf{1}$ and $\mathbf{0}$ are M-element vectors with all ones and all zeros respectively. Since the cost is convex and the constraints are linear, it is a convex problem with a unique minimum. During every ICD voxel update, the M-dimensional optimization problem is solved. We employ a modified simplex method to solve the problem.

The regularizer term $U(x_l)$ was chosen to be a simple quadratic function that penalizes the difference in the current voxel and its neighbors.

$$U(x_l) = \frac{1}{2\sigma^2} \sum_{m=1}^{M} \sum_{k \in \mathcal{N}_l} \alpha_k (x_{lm} - x_{km})^2 \tag{6}$$

where $\mathcal{N}_l$ denotes the neighborhood of x_l, and α_k depends on the position of the neighbor such that $\sum_{k\in\mathcal{N}_l} \alpha_k = 1$. Its Hessian w.r.t. x_l is a diagonal matrix with all entries $\frac{1}{\sigma^2}$, since we have not yet applied inter-material regularization. The Hessian of total cost stays positive definite and it still has a unique minimum. If we are concerned with the average value of material fraction within ROIs, we can afford to have somewhat noisy reconstructions, as this helps avoid bias. Therefore the value of σ was chosen to yield very light regularization for current experiments.

2) Simplex Implementation: Since both terms in cost are quadratic, it can be re-written in a standard quadratic expression. Let $\mathcal{H}$ be the hessian and $\mathcal{D}$ be the derivative of the total cost w.r.t x_l. The optimization problem can be written in vector notation as

$$\frac{1}{2} x_l^T \mathcal{H} x_l + \mathcal{D}^T x_l \tag{7}$$

The single-site optimization problem, from the KKT condition, now becomes

$$\hat{x_l} = \underset{x_l}{\textit{argmin}}\ \frac{1}{2} x_l^T \mathcal{H} x_l + \mathcal{D}^T x_l - \mu^T x_l + \lambda(\mathbf{1}^T x_l - 1)$$
$$\text{s.t., } \mathbf{1}^T x_l \leq 1$$
$$x_l \geq \mathbf{0},\ \mu \geq \mathbf{0},\ \lambda \geq 0,\ \mu_m x_{lm} = 0 \text{ for } m = 1 \text{ to } M \tag{8}$$

where μ is a vector Lagrangian multiplier for positivity constraint, λ is a scalar Lagrangian for the sum-less-than-or-equal-to-1 constraint. $\mu_m x_{lm} = 0$ comes from complementary slackness. We solve this problem during every ICD voxel update using a modified simplex method.

C. Spectral CT Data Acquisition

Phantom solutions were created with varying concentrations of iodine, gadolinium, and calcium in water in 7 tubes. Please refer to Table I for concentration and spatial information of different solutions. Note that iodine, gadolinium, and calcium are not mixed with each other and are kept spatially orthogonal in this setup. We intend to mix contrasts in our future work. K-edges of iodine and gadolinium are around 33 KeV and 53 KeV respectively. The bins are chosen to take advantage of the difference in the photon absorption properties of different basis materials. Some defective detector pixels did not register any signal, or were unreliable, and data correction was therefore

Material	Calibration	Test
	$Conc.$	$Conc.$
I_{high}	15.86	9.52
I_{low}	7.93	4.76
Gd_{high}	19.66	11.79
Gd_{low}	9.83	5.90
Ca_{high}	146.29	87.77
Ca_{low}	73.14	43.89

TABLE I: Concentration of materials for Calibration data and Test data in mg/mL

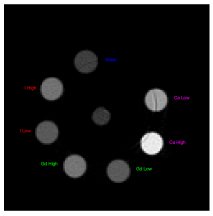


Fig. 1: Material Concentration and spatial location

performed to eliminate outliers and make sinograms consistent. The first set of phantom solutions was used to calibrate the mixing matrix (please refer to the 2^{nd} column of Table I). Then that mixing matrix was used in the forward model of reconstruction for a 2^{nd} set of phantom data (please refer to the 3^{rd} column of Table I).

The scanner employed is a photon-counting spectral CT scanner manufactured by MARS Bioimaging (Christchurch, NZ). The photon-counting detector comprises 5 CdZnTe sensors separated by small gaps and bonded to a Medipix 3RX chip. Each detector panel has 128 rows and 128 channels, and $0.11\ mm \times 0.11\ mm$ "pixel pitch". Scan voltage was $80\ kVp$ with an X-ray tube current of $35\ \mu A$. The total number of views was 720 in an axial scan mode. The resulting dimension of the sinogram was 128 rows, 663 channels, 740 views.

Source to iso-center distance was $200\ mm$, and source to detector distance was $250\ mm$. The energy bins are (1) $7.0-19.0$, (2) $19.0-29.0$, (3) $29.0-38.8$, (4) $38.8-51.1$, and (5) $51.1-82.6\ keV$. The photon count rate per pixel was kept to 10 photons/ms to mitigate pulse pileup effects. The exposure time was 125 ms. No external beam filtration was used. Sensor pixels collect with "charge-summing mode." Average counts per detector pixel was 1500. Voxel size of reconstructions was $0.4\ mm \times 0.4\ mm \times 0.5\ mm$. Reconstruction field of view was $61\ mm$.

D. Mixing matrix computation

Individual reconstructions from every bin were first performed with scalar MBIR. This furnished 5 LAC images, each from a different energy bin. 10 slices of each LAC image volume were used to compute $\mathcal{M}$. The mixing matrix was computed from ROIs shown in red in Fig 2. The following equation must be satisfied ideally

$$\mathcal{M}x^T = \mu \tag{9}$$

where x is a $Q \times M$ matrix that represents material fraction of Q voxels in ROIs that is known from ground truth, μ is a $E \times Q$ matrix that denotes LAC of Q voxel in ROIs from E reconstructed bins. $\mathcal{M}$ is computed with least-squares linear regression using the following

$$\mathcal{M} = \mu x(x^T x)^{-1} \tag{10}$$

The resulting attenuation coefficients may be found in Table II.

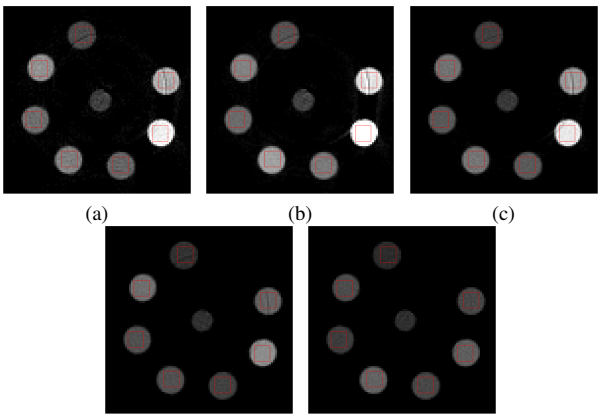


Fig. 2: Middle slice of LAC image reconstruction of each energy bin sinogram for mixing matrix estimation. Display window [0, 0.1]
(a) Bin1:$7.0-19.0\ keV$, (b) Bin2: $19.0-29.0\ keV$, (c) Bin3: $29.0-38.8\ keV$, (d) Bin4: $38.8-51.1\ keV$, (e) Bin5: $51.1-82.6\ keV$

Energy	Linear Attenuation Coefficient			
	$Water$	$Iodine$	$Gadolinium$	$Calcium$
$Bin\ 1$	0.0301	8.0544	7.5169	0.7722
$Bin\ 2$	0.0345	4.9786	11.1868	1.0905
$Bin\ 3$	0.0253	5.9366	8.1421	0.7815
$Bin\ 4$	0.0196	7.2125	4.8177	0.4259
$Bin\ 5$	0.0177	3.7628	8.4091	0.2395

TABLE II: Mixing matrix computed from 5 sinograms and ground truth. All values of linear attenuation coefficient are in mm^{-1}

E. Spectral Reconstruction

MBIR of basis material fractional concentration images is now performed directly from the collection of 5 new sinograms. The output reconstruction images are material maps for the basis materials' volume fractions. The result presented here comes from iteration 150.

III. Results

The computed mixing matrix $\mathcal{M}$ is used in the model based iterative algorithm to reconstruct the material fraction images. The convergence is slow because of the poor condition number of the Hessian in (4). Since the tube solutions are nearly constant in z direction, we show only one reconstructed slice of the total image volume. Each ROI is the center square patch of the 2-D image within each tube.

Ground Truth (volume fraction)				
Spatial Loc.	Basis Material			
	$Water$	$Iodine$	$Gadolinium$	$Calcium$
ROI_1	**1.000000**	0	0	0
ROI_2	**0.998073**	**0.001926**	0	0
ROI_3	**0.999036**	**0.000964**	0	0
ROI_4	**0.998509**	0	**0.001490**	0
ROI_5	**0.999253**	0	**0.000746**	0
ROI_6	**0.946408**	0	0	**0.053591**
ROI_7	**0.972463**	0	0	**0.027536**

TABLE III: Ground Truth from the phantom solutions. Contrast agents are spatially orthogonal. Note that the values in volume fraction have been converted from concentration (mg/mL) in Table I 3^{rd} column

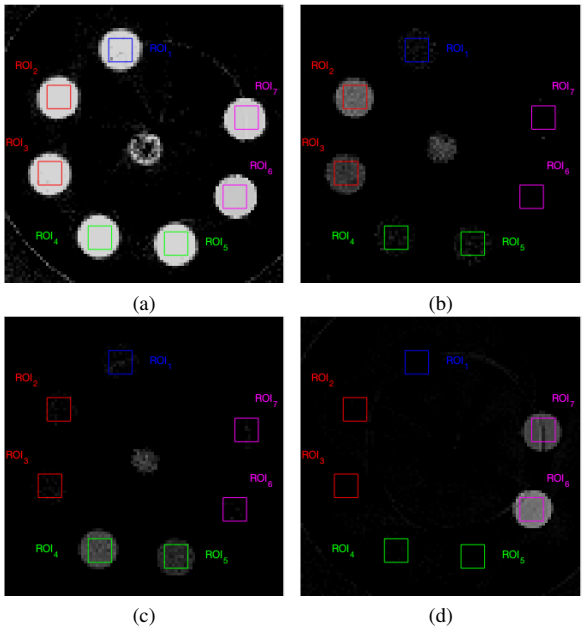


Fig. 3: Reconstructed image of material fraction of (a) Water: Display window [0, 1.2]. (b) Iodine: Display window [0, 0.006]. (c) Gadolinium: Display window [0, 0.006]. (d) Calcium: Display window [0, 0.15]

Estimated values (volume fraction)				
Spatial Loc.	Basis Material			
	Water	*Iodine*	*Gadolinium*	*Calcium*
ROI_1	**0.987255**	0.000248	0.000196	0.000035
ROI_2	**0.997248**	**0.002288**	0.000137	0.000102
ROI_3	**0.997725**	**0.001237**	0.000149	0.000000
ROI_4	**0.997719**	0.000190	**0.001621**	0.000468
ROI_5	**0.998700**	0.000238	**0.000951**	0.000110
ROI_6	**0.935048**	0.000016	0.000119	**0.064815**
ROI_7	**0.961056**	0.000049	0.000096	**0.038797**

TABLE IV: Estimated average values of material fraction within ROIs.

Error (volume fraction)				
Spatial Loc.	Basis Material			
	Water	*Iodine*	*Gadolinium*	*Calcium*
ROI_1	**1.27 %**	—	—	—
ROI_2	**0.08 %**	**18.81%**	—	—
ROI_3	**0.13 %**	**28.35%**	—	—
ROI_4	**0.08 %**	—	**8.80%**	—
ROI_5	**0.05 %**	—	**27.48%**	—
ROI_6	**1.20 %**	—	—	**20.94%**
ROI_7	**1.17 %**	—	—	**40.89 %**

TABLE V: Error in estimation of volume fraction within ROIs (in %).

In Table IV, We see that some amount of iodine, gadolinium, and calcium erroneously appear in some tubes, but overall the values are close to the ground truth values of Table III. In the center of the image there is a Teflon screw that does not match any of the basis materials. It can be seen in the material fraction images that the outer periphery of the screw is approximated by water and the inside is approximated by iodine and gadolinium (and air). As concentration decreases, in Table V the percentage error is observed to increase for each material, which makes sense since SNR value drops and the condition number of the mixing matrix is large.

Some of the individual reconstructed LAC images have ring artifacts and some have circular bands that have voxels with clearly different LAC value from their neighbors. This is caused by inconsistent behavior of some detectors and contributes to error in the estimate of material fraction's average values within the ROI. Another potential source of error could be the noise in mixing matrix from the calibration dataset as pointed out in [3]. Inter-material difference in material fraction values was not penalized in the regularizer term $U(x_l)$ but will be explored in future work.

IV. Conclusion

Accurate estimation of the spectral mixing matrix is important for a good linear model. After computation of the matrix from a set of calibration measurements, direct reconstruction of material fraction images was performed successfully. Average values within ROIs are consistent with the ground truth. Future work will include mixed or spatially coincident contrast agent and tissue compositions of varying concentration, as well as small-animal contrast studies.

V. Acknowldegement

This work was supported by the Martell Family PhD Fellowship and the Equipment Renewal and Restoration Program, both at the University of Notre Dame, the Department of Homeland Security ALERT Center for Excellence under Grant Award 2013-ST-061-ED0001, the Kelly Cares Foundation and St. Joseph Health System. The authors acknowledge also the assistance of Notre Dame's Integrated Imaging Facility and Center for Environmental Science and Technology.

References

[1] Roessl, E., and R. Proksa *K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors*, Physics in Medicine & Biology 52, no. 15 (2007): 4679.

[2] Curtis, Tyler E., and Ryan K. Roeder. *Quantification of multiple mixed contrast and tissue compositions using photon-counting spectral computed tomography*, Journal of Medical Imaging 6.1 (2019): 013501.

[3] Curtis, Tyler E., and Ryan K. Roeder. *Effects of calibration methods on quantitative material decomposition in photoncounting spectral computed tomography using a maximum a posteriori estimator*, Medical physics 44.10 (2017): 5187-5197.

[4] Zhang, Ruoqiao, et al *Model-based iterative reconstruction for dual-energy X-ray CT using a joint quadratic likelihood model*, IEEE transactions on medical imaging 33.1 (2013): 117-134.

[5] Mendona, Paulo RS, Peter Lamb, and Dushyant V. Sahani. *A flexible method for multi-material decomposition of dual-energy CT images*, IEEE transactions on medical imaging 33.1 (2013): 99-116.

[6] Long, Yong, and Jeffrey A. Fessler *Multi-material decomposition using statistical image reconstruction for spectral CT*, IEEE transactions on medical imaging 33.8 (2014): 1614-1626.

[7] Schmidt,Taly Gilat, Rina Foygel Barber, and Emil Y. Sidky, *A Spectral CT method to directly estimate basis material maps from experimental photon-counting data*, IEEE transactions on medical imaging 36.9 (2017): 1808-1819.

[8] Thibault, JeanBaptiste, et al. *A three-dimensional statistical approach to improved image quality for multislice helical CT*, Medical physics 34.11 (2007): 4526-4544.

[9] Bouman, Charles A., and Ken Sauer *A unified approach to statistical tomography using coordinate descent optimization*, IEEE Transactions on image processing 5.3 (1996): 480-492.

Perturbation Response of Model-based Material Decomposition with Edge-Preserving Penalties

Wenying Wang, Grace J. Gang, Matthew Tivnan, and J. Webster Stayman
Department of Biomedical Engineering, Johns Hopkins University, Baltimore MD, 21205

***Abstract*—Spectral CT permits material discrimination beyond the structural information in conventional single-energy CT. Model-based material decomposition facilitates direct estimation of material density from spectral measurements, incorporating a general forward model for arbitrary spectral CT system, a statistical model of spectral CT measurements, and flexible regularization schemes. Such one-step approaches are promising for superior image quality, but the relationship between regularization parameters, imaging conditions, and reconstructed image properties is complicated. More specifically, the estimator is inherently nonlinear and may include additional nonlinearities like edge-preserving regularization, making image quality metrics intended for linear system evaluation difficult to apply. In this work, we seek approaches to quantify the image properties of this inherently nonlinear process through an investigation of perturbation response - the generalized system response to a local perturbation of arbitrary shape, location, and contrast. Such responses include cross-talk between material density channels, and we investigate the application of this metric in a sample spectral CT system. Inspired by the prior work under assumptions of local linearity and shift-invariant we also propose a prediction framework for perturbation response using a perceptron neural network. The proposed prediction framework offers an alternative to exhaustive evaluation and is a potential tool that can be used to prospectively choose optimal regularization parameters based on imaging conditions and diagnostic task.**

***Index Terms*—Spectral CT, CT image quality, Nonlinear system analysis**

I. Introduction

SPECTRAL CT is finding an increasing number of clinical applications with its ability to perform material discrimination, provide quantitative density estimates, reduce artifacts, and provide specialized scans (e.g. virtual non-contrast, monoenergetic reconstructions, etc.). There are a diversity of approaches that enable spectral CT. Fundamentally, such scanners require multiple spectral channels achieved, e.g., through kV-switching, source-side filtration, multiple sources, or detectors that have channel-dependent energy sensitivity such as dual-layer or photon-counting detectors.

Conventionally, material decomposition is accomplished in two steps in one of two ways: 1) Image-domain decomposition methods[1] estimate material density maps from attenuation values individually reconstructed from each spectral channels. This method estimates the densities as a linear or polynomial combination of attenuation coefficients in different spectral channels, and often needs further correction for inherent beam-hardening artifacts. 2) Projection domain decomposition takes spectral measurements and first estimates line integrals for each basis material.[2] These line integrals are then reconstructed into material density maps. This method inherently accounts for the beam-hardening effect but generally requires spectral data where the geometry is consistent between channels (e.g., common line integrals are required across channels).

Recently, a number of one-step model-based material decomposition (MBMD) approaches have been developed to estimate material densities directly from the spectral CT measurements.[3;4;5] Such MBMD methods are flexible and allow for different system geometries between spectral channels[5], improved system modeling[6], and inherent modeling of spectral effects for beam hardening, etc. Moreover, integration of measurement statistics and prior knowledge through advanced regularization has the potential to achieve better image quality or higher sensitivity for lower-dose protocols.

Despite the benefits and flexibility of MBMD, the image properties of the reconstructed material density maps can be complex. Similar to traditional model-based iterative reconstruction approaches for single-energy imaging, MBMD has dependencies on a number of imaging conditions including reconstruction parameter values, patient size and anatomy, location in the patient, x-ray technique, etc. Moreover, image properties are more complex than standard single-energy imaging where the focus is on a single volume. In additional to single image considerations like noise and resolution there are typically biases in one material basis that can result in correlated *"cross-talk"* in the other material bases. The inherently nonlinear forward model makes it difficult to apply traditional image properties metrics like point spread function and noise power spectrum. However, under a locally linear and shift-invariant assumption, one can consider local forms of point spread function and noise power spectrum. The authors have previously explored such approximations to derive closed-form predictors of noise and resolution (and cross-response) in the case of quadratic regularization.[7]

Simple quadratic regularization is limited in the degree of image quality improvements available as compared to more sophisticated regularization approaches. In this work, we seek to analyze and predict more general image properties under more nonlinear processing. Specifically, we apply the concept of perturbation response (previously used in a single-energy reconstruction[8]) to spectral CT and MBMD. We consider the particular case of MBMD with an edge-preserving penalty and investigate the response to arbitrary perturbations over a range of imaging conditions and parameters. The perturbation response is general and includes the same kind of cross-response correlations between material bases described above.

In addition to the investigation of perturbation response as a tool to quantify image properties in MBMD, we also seek an approach to enable a general prediction framework

for perturbation response. Specifically, we investigate the use of a perceptron neural network - trained using a sampling of perturbation responses - to prospectively predict more arbitrary responses. The structure of this network is motivated by prior analysis under quadratic regularization conditions. Network outputs are validated for a class of spiculated iodine perturbations, with special attention to the cross-talk between material bases.

II. Methods

A. Background

1) Model-based material decomposition: The polychromatic forward model from Tilley et al[5] relates the mean measurements vector $\bar{y}$ (over angles, pixels, and spectral channels) to the material density vector ρ (over materials and voxels):

$$\bar{y} = \mathbf{GS}\exp\{-\mathbf{QA}\rho\} \tag{1}$$

where $\mathbf{A}, \mathbf{Q}, \mathbf{S}, \mathbf{G}$ denote the system projector, material mass attenuation, spectral sensitivity, and non-spectral gain, respectively. Note that the spectral sensitivity matrix $\mathbf{S}$ carries both source-side spectral modulation and spectral sensitivity of detectors. With a flexible system matrix $\mathbf{A}$, the forward model is suitable for various spectral CT systems including kV-switching, photon-counting detectors, dual-layer detector, etc.

Presuming the spectral measurements follow multivariate Gaussian with covariance $\mathbf{K}_y$, the material decomposition $\hat{\rho}$ is estimated by minimizing the following objective function including a data fidelity term and a regularization term $R(\rho)$:

$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}} \frac{1}{2}(y-\bar{y})^T\mathbf{K}_y^{-1}(y-\bar{y}) + R(\rho). \tag{2}$$

Here we adopt an in-basis Huber penalty as the regularization term. The total penalty is computed as the sum of a Huber penalty with independent regularization parameters (β_k, δ_k) in k^{th} material basis:

$$R(\rho) = \sum_k \beta_k \sum_j \sum_{n\in\mathcal{N}_j} \psi_k(\rho_k^j - \rho_k^n) \tag{3}$$

$$\psi_k(x) = \begin{cases} \frac{1}{2}x^2, & |x| \le \delta_k \\ \delta_k|x| - \frac{1}{2}\delta_k^2, & |x| > \delta_k \end{cases}. \tag{4}$$

2) Local impulse response prediction in quadratic MBMD: In previous work[7], we derived a mathematical expression for the local impulse response in MBMD with a quadratic penalty which is locally linearizable. The response to a local impulse at the j^{th} voxel in the k^{th} material basis is

$$l_k^j = [\mathbf{F}+\mathbf{R}]^{-1}\mathbf{F}e_k^j, \tag{5}$$

$$\mathbf{F} = \mathbf{A}^T\mathbf{Q}^T\mathbf{W}^T\mathbf{K}_y^{-1}\mathbf{WQA}, \tag{6}$$

$$\mathbf{W} = \mathbf{GS}D\{\exp\{-\mathbf{QA}\rho\}\}, \tag{7}$$

where e_k^j is a Kronecker delta vector that has a single unity element at the j^{th} voxel in the k^{th} material basis, $\mathbf{F}$ is the system Fisher information term, $\mathbf{R}$ is the Hessian of the penalty term $R(\rho)$, and $\mathbf{W}$ is the spectral channel-wise weights. The above closed-form expression carries dependencies on imaging system characteristics $(\mathbf{A}, \mathbf{Q}, \mathbf{S}, \mathbf{G})$, patient anatomy (through $\mathbf{W}$, $\mathbf{K}_y$), and the regularization ($\mathbf{R}$).

B. Perturbation response in MBMD with Huber penalty

Extending previous work[8] to material decomposition, we define the system response, $\mathcal{H}(\rho_s)$, to an arbitrary perturbation ρ_s as the difference between mean MBMD estimates with and without the perturbation:

$$\mathcal{H}(\rho_s) = \overline{\hat{\rho}(y(\rho+\rho_s))} - \overline{\hat{\rho}(y(\rho))} \tag{8}$$

where $\rho_s, \rho, \hat{\rho}, y$ denote the multi-material perturbation, unperturbed anatomy, decomposition method, and spectral measurements, respectively. The perturbation response characterizes the dependencies on imaging conditions through forward model $y(\cdot)$, decomposition method through decomposition operator $\hat{\rho}(\cdot;\beta_k,\delta_k)$ and patient anatomy ρ. Note that this expression reduces to the local impulse response used to form (5) when ρ_s is an impulse in material k at location j.

To investigate perturbation response dependencies under varying imaging conditions, we emulated a dual-material phantom consisting of soft tissue and iodine (Figure 1) for spectral CT measurements in an 80/140 kV-switching dual-energy CT system, where the source-to-detector distance was 1200 mm and the source-to-axis distance was 600 mm. We chose perturbations to be spiculated lesions of 1.5, 2.0, or 2.5 mm radius, which were inserted at 17 different locations. The contrast of each perturbation was 10 mg/mL or 5 mg/mL. The perturbation responses at the central location ① and the location ② near the edge were investigated in detail. We emulated 360 angles × 540 bins noiseless projections and reconstructed 2 material bases, each composed of 150 × 270 0.5 mm cubic voxels. For each perturbation evaluation, we performed a 4D sweep over regularization strengths and thresholds in soft tissue and iodine bases $(\beta_s, \beta_I, \delta_s, \delta_I)$ with values summarized in the second column of Table I. For MBMD optimization we used the preconditioned approach in Tivnan et al[9] with 10 iterations of preconditioned Newtons method, each with 20 subiterations of conjugate gradient descent to apply inverse Hessian, followed by 200 iterations of separable paraboloidal surrogates (Tilley *et al* 2019[5]). Perturbation responses were computed as defined in (8) and analyzed with varying decomposition parameters and perturbation size.

C. Perturbation response prediction framework

While the derivation in II-A2 presumes local linearity and does not strictly apply to Huber-penalized MBMD, this general form can approximate the Huber approach if the quadratic penalty is appropriately tuned. Specifically, choosing a quadratic regularization strength based on selection of an operating point for the Huber penalty sufficiently close to the penalty as evaluated at the solution. (Figure 2a). This surrogate indicates that for MBMD with Huber, the effective Hessian of the Huber penalty term, $\tilde{\mathbf{R}}$, is dependent on location and data in addition to the regularization parameters. Furthermore, the regularization is locally shift-variant, thus a direct inverse is improper. More generally, the perturbation response can be written as a nonlinear function of the Fisher information term operating on the perturbation and the effective Hessian of Huber penalty term:

$$\mathcal{H}(\rho_s;\rho) = f(\mathbf{F}\rho_s, \tilde{\mathbf{R}}(\rho, \rho_s, \{\beta_k\}, \{\delta_k\})) \tag{9}$$

TABLE I: Parameters for training and testing

Dataset		Training	Testing
Perturbation radius (mm)		1.5, 2.0, 2.5	1.8, 2.3
Regularization strength	β_w β_I	1.8E8, 1.8E9, 1.8E10 2.7E10, 2.7E11, 8.5E11	5.7E8, 5.7E9 4.3E10, 4.3E11
Regularization threshold(mg/mL)	δ_w δ_I	150, 15, 1.5 3, 0.3, 0.03	50, 5 1, 0.1

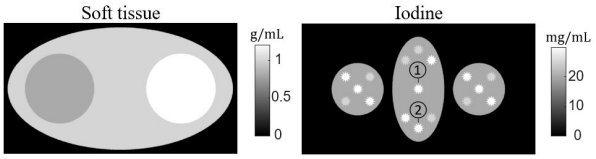


Fig. 1: Dual-material (soft tissue and iodine) digital phantom with spiculated lesion perturbations in iodine. Perturbation responses are evaluated at the central location ① and location ② near the edge.

Note that the above function still carries dependencies on imaging conditions and penalty design but represents a potential minimal input response estimation. Gang *et al.*[10] used a multi-layer perceptron network (Figure 2b) to predict the perturbation response in penalized-likelihood reconstruction with Huber penalty in conventional single-energy CT. In this work, we adopt a similar network structure to predict the nonlinear relationship between the input, including a local region-of-interest circumscribing the perturbation in $\mathbf{F}\rho_s$ and regularization parameters (β_k,δ_k), and the perturbation response in all material bases.

A perceptron neural network (2b) was trained using retrospective perturbation evaluation with a sweep of regularization parameters and perturbation size as the training set. Prediction feasibility was validated with a testing set with differing perturbation size and regularization parameters. The perturbation radius and MBMD parameters for training and testing are summarized in I.

III. Results

A. Perturbation response dependencies

The dependencies on pairs of regularization parameters are illustrated with perturbation responses computed retrospectively with varying reconstruction parameters and perturbation sizes. We denote the in-basis iodine perturbation response as $\mathcal{H}_I(\rho_I)$, outlined in red, while $\mathcal{H}_s(\rho_I)$ marks the response in the soft-tissue material basis to the inserted iodine perturbation, outlined in blue.

The 2.5 mm perturbation response at the central location ① with varying regularization strengths or thresholds are summarized in Figures 3a and 3b. Figure 3a shows perturbation responses with varying regularization strengths (β_s,β_I). When we increased β_I, we observed shrinkage of response in iodine, but also increased cross-talk in soft tissue. With increased β_s, the cross-talk was reduced, while the contrast in iodine was less biased. Figure 3b illustrates the dependency on varying regularization strength and threshold in iodine (β_I, δ_I). Again, the size of perturbation response shrank with increased β_I, and was more distinct with increased regularization threshold δ_I.

Figure 3c shows the responses to perturbations of three different sizes with the same regularization parameters. The spiculated feature of 2.5 mm and 2.0 mm perturbation was maintained, while in the 1.5 mm perturbation response was eliminated. The cross-talk in water also varied with perturbation size. Compared with the perturbation responses at location ①, the perturbation response at location ② was more isotropic, showing shift-variant image properties.

B. Perturbation response prediction

Figures 4a and 4b summarize the comparison between measurement and prediction of perturbation responses at the central location ① with varying β_I or δ_I, respectively. The prediction exhibits good agreement with the measured perturbation response in general shape and intensity with varying decomposition parameters. Figure 4c displays the measured and predicted perturbation responses of two different sizes at location ② . Again, we observe good agreement between the measurement and the prediction, where the spiculated feature diminished with decreased perturbation size. The proposed prediction framework successfully captures the nonlinear properties in Huber MBMD. However, small mismatches are notable between measurement and prediction, especially in cross-talk in the water basis. Note that current prediction network was trained with limited samples for each variable. We conjecture the residual mismatch can be reduced with more samples included in the training set.

IV. Discussion and Conclusion

In this work, we adopted the perturbation response as a generalized image quality metric for edge-preserving MBMD. The dependencies on decomposition parameters and locations are observed and analyzed with retrospective evaluations. We adopt a generalized prediction framework using neural networks and extend it to the spectral scenario. With the proposed prediction framework, perturbation response can be quantified with Fisher-information-like perturbation inputs, providing a potential avenue for prospective image quality analysis and control.

Ongoing work includes application of the perturbation response to additional stimuli including varied (1) perturbation shape, (2) surrounding anatomy, e.g, lung nodules in thorax imaging or liver lesions in abdomen imaging, and (3) varying material combinations, for example, a hyper-dense perturbation in iodine and a hypo-dense perturbation in soft tissue. Future work will further investigate the ability of the neural network approach to generalize the prediction response to varied patient anatomy, location, stimuli, etc.

Acknowledgments

This work is supported, in part, by NIH grants R01EB025470, R21CA219608, and R21EB026849; and, in part, by computational resources at the Maryland Advanced Research Computing Center (MARCC).

References

[1] X. Liu, L. Yu, A. N. Primak, and C. H. McCollough, "Quantitative imaging of element composition and mass fraction using dual-energy CT: Three-material decomposition," *Medical Physics*, vol. 36, no. 5, pp. 1602–1609, 2009.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerized tomography," *Phys Med Biol*, vol. 21, no. 5, pp. 733–744, 1976.

[3] R. F. Barber, E. Y. Sidky, T. G. Schmidt, and X. Pan, "An algorithm for constrained one-step inversion of spectral CT data," *Physics in Medicine and Biology*, vol. 61, no. 10, pp. 3784–3818, may 2016.

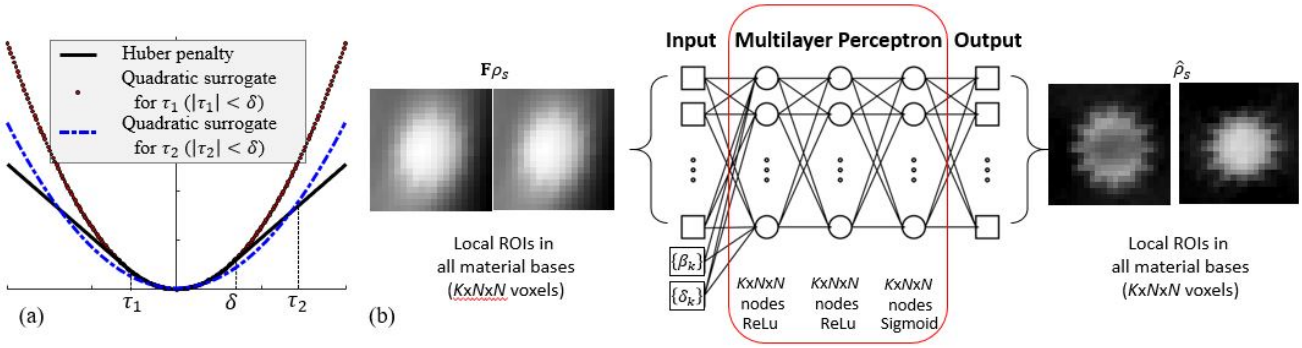


Fig. 2: (a) Approximation of the Huber penalty with a quadratic surrogate. The red dotted line shows the quadratic surrogate for the operating point τ_1 within the quadratic region, and the blue dashed line shows the quadratic surrogate for the operating point τ_2 in the linear region. (b) The perceptron neural network structure for perturbation response prediction in MBMD.[10] The input includes a local $20x20$ ROI in $\mathbf{F}\rho_s$ and the regularization parameters, while the output is the same $20x20$ ROI in perturbation response, including in-basis and cross-basis response.

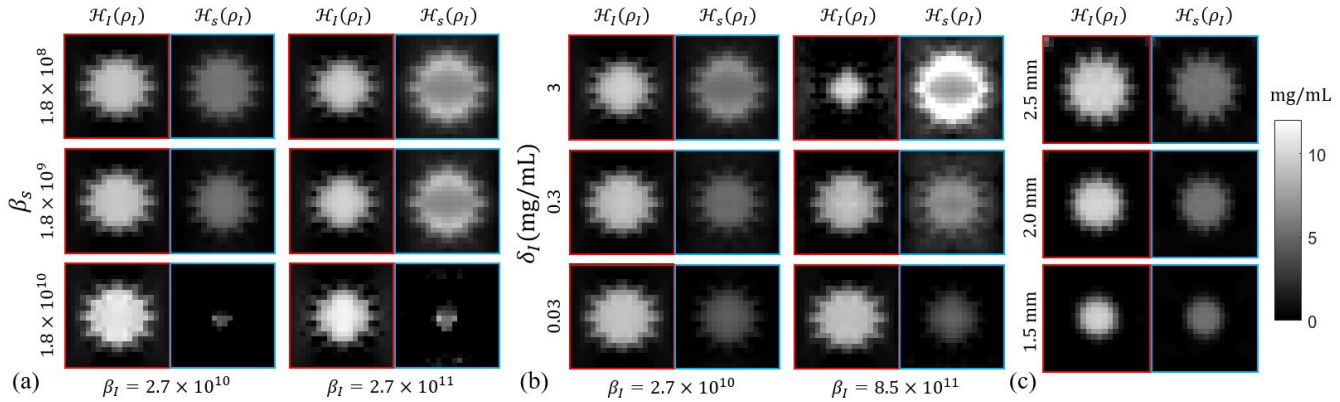


Fig. 3: Perturbation responses with varying (a) regularization strengths, (b) regularization strength and threshold in iodine, and (c) perturbation sizes

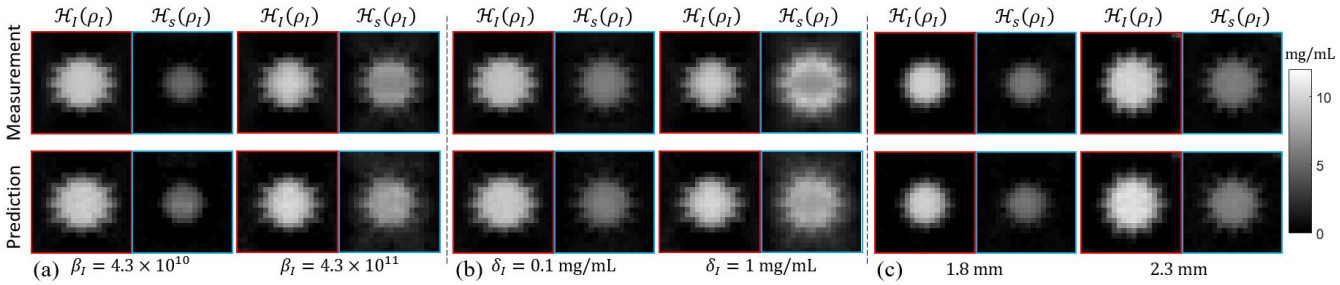


Fig. 4: Measured and predicted perturbation responses with varying (a) regularization strength in iodine, (b) regularization threshold in iodine, and (c) perturbation sizes

[4] K. Mechlem, T. Sellerer, S. Ehn, D. Münzel, E. Braig, J. Herzen, P. B. Noël, and F. Pfeiffer, "Spectral Angiography Material Decomposition Using an Empirical Forward Model and a Dictionary-Based Regularization," *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2298–2309, 2018.

[5] S. Tilley II, W. Zbijewski, and J. W. Stayman, "Model-based material decomposition with a penalized nonlinear least- squares CT reconstruction algorithm," *Phys Med Biol*, vol. 64, no. 3, 2019.

[6] W. Wang, M. Tivnan, G. J. Gang, Y. Ma, Q. Cao, M. Lu, R. E. Colbeth, W. Zbijewski, and J. W. Stayman, "Model-based Material Decomposition with System Blur Modeling," in *SPIE Medical Imaging*, 2020.

[7] W. Wang, S. Tilley II, M. Tivnan, and J. W. Stayman, "Local response prediction in model-based CT material decomposition," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. SPIE, may 2019, p. 30.

[8] S. Ahn and R. M. Leahy, "Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET," *IEEE Transactions on Medical Imaging*, vol. 27, no. 3, pp. 413–424, 2008.

[9] M. Tivnan, W. Wang, and J. W. Stayman, "Preconditioned algorithms for model-based iterative reconstruction and material decomposition from spectral ct data," *arXiv:3010843*, 2020.

[10] G. J. Gang, K. Cheng, X. Guo, and J. W. Stayman, "Generalized prediction framework for reconstructed image properties using neural networks," in *Medical Imaging 2019: Physics of Medical Imaging*, no. March. SPIE, mar 2019, p. 20.

Convolutional Neural Network based Iterative Metal Artifact Reduction

David Weller, Maik Stille, Philipp Koch, Mark Dreier, Alfred Mertins and Thorsten M. Buzug

***Abstract*—The presence of metal objects within the human body is an ongoing problem for medical computed tomography despite the availability of various artifact reduction strategies. Especially, iterative metal artifact reduction has shown to perform well. Over the last decade, convolutional neural networks have achieved remarkable results in various fields of image processing. Recently, the application of this concept has produced the first promising results on the reduction of metal artifacts in computed tomography. We propose a Convolutional Neural Network based Iterative Metal Artifact Reduction (CI-MAR). The iterative scheme is defined by the following steps: An artifact reduction by filtering a corrupted image, followed by a sinogram update with projection data of the filtered image. An image, reconstructed from this updated sinogram is the output of one iteration. Results show an improvement in artifact reduction compared to a linear interpolation approach.**

I. Introduction

The presence of metal implants or objects with high attenuation coefficients can cause severe image artifacts in computed tomography (CT). X-rays that pass through these structures experience a high attenuation that leads to inconsistencies in the sinogram. A common assumption in the field of computed tomography is that the attenuation behavior of the employed radiation can be approximated by the laws of monochromatic attenuation. When X-rays pass through structures with very high attenuation, their lower energy spectral components experience a considerably stronger attenuation than their higher ones. As a result of this so-called *beam-hardening* it is no longer possible to sufficiently model the attenuation behaviour as monochromatic. Furthermore, the strong attenuation of metal objects can lead to an increased ratio of scattered photons, as well as an enhanced *partial-volume-effect* due to sharp edges. Applying the *filtered-backprojection* (FBP) on this corrupted data generates images that show strong artifacts in the form of streaks that spread from the high-attenuation structures. In general, strategies for metal artifact reduction (MAR) can be separated into two categories: correcting corrupted data with a *Beam Hardening Correction* (BHC) and interpolation of the corrupted data with methods such as the *Linear Interpolation* (LI), *Normalised Metal Artifact Reduction* (NMAR) [1] or iterative schemes such as the *Augmented Likelihood Image Reconstruction* (ALIR) [2]. The BHC tends to fail in the presence of large or very high density objects while interpolation in the projection domain can introduce new artifacts in the reconstructed images [3] [4]. Iterative reconstruction techniques based on these strategies are able to reduce artifacts very well but can be highly sensitive to hyperparameters that have to be adapted according to the size, position or even material of the implants. Therefore, none of these strategies has been able to establish an universally and satisfyingly applicable MAR-approach. Over the last decade, deep learning has proven to be highly effective for classification as well as regression problems. In the area of image processing, the concept of the *Convolutional Neural Network* (CNN) has repeatedly outperformed conventional methods. Therefore, CNNs have already been applied to various problems of medical imaging, including metal artifact reduction. By now, concepts have been introduced that apply CNN based correction in the image or projection domain as well as dual domain approaches. Previous works in this area have shown that CNNs are able to significantly reduce metal artifacts but tend to leave some remaining artifacts or introduce unintended image smoothing. An area that has not been examined yet is the usage of neural networks in the context of an iterative CT reconstruction or correction scheme. Therefore, the goal of this work is to implement a *CNN Based Iterative Metal Artifact Reduction* (CI-MAR) that is capable of a stepwise reduction of metal artifacts while avoiding over-smoothing by ensuring a maximum consistency to uncorrupted projection data.

D. Weller, M. Stille and T.M. Buzug are with the Institute of Medical Engineering, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, e-mail: (see {weller, stille, buzug}@imt.uni-luebeck.de).

P. Koch, M. Dreier and A. Mertins are with the Institute for Signal Processing, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, e-mail: (see {ph.koch, mark.dreier@student, alfred.mertins}@uni-luebeck.de)

II. Methods

A. Basic concept

To provide a context for the following detailed explanations of experiments and training data, the goal of this subsection is to outline the basic approach of our algorithm:

1) **Identification of corrupted data:** The proposed CI-MAR can be interpreted as an enhancement of the LI-MAR because its aim is to reduce the characteristic streak artifacts of this method. Therefore, as a preparation for an iterative artifact correction, the LI-MAR has to be applied. From an FBP of an uncorrected sinogram S_0 an initial image f_0 is reconstructed. From this image, a threshold-based segmentation obtains a binary segmentation image f_{seg} of the metal objects. A forward projection (FP) of f_{seg} generates a sinogram S_{impl}. All entries of S_{impl} that differ from zero can easily be identified as projection data of the implant and be used for segmentation of the corrupted data.
2) **LI-MAR:** By using the segmentation of the implant, the corrupted projection data is replaced with a linear

interpolation. An FBP of the resulting sinogram $S_{\text{LI-MAR}}$ reconstructs an image that contains the typical artifacts of LI-MAR. This image f_{LI} is now progressively refined regarding the streak artifacts.

3) **CNN-based filtering:** This step marks the begin of the iterative scheme. A CNN ϕ_i, specifically trained for this task, is used to reduce artifacts in an image $f_{\text{i-1}}$ that is the output of the previous iteration $i-1$. This artifact reduction is performed by filtering the image. At the start of the first iteration, a CNN ϕ_1 applies such a filter on the LI-MAR image f_{LI}. We denote the CNN-based filtering as *CNN-step*. The outcome of the CNN-step is an intermediate image $f_i^{'}$.
4) **Sinogram update:** Applying a FP, a sinogram $S_i^{'}$ is generated from the filtered image $f_i^{'}$. By using the segmentation of the metal trace from Step 1 the trace of the implant in $S_i^{'}$ can be identified. The values at the corresponding entries in $S_i^{'}$ are used to replace the corrupted entries in the original sinogram to acquire an updated sinogram S_i.
5) **Image update:** By using the FBP, an updated image f_i is obtained from the updated sinogram S_i. This image is used as input for the next iteration $i+1$.

The iterative character of the scheme is given by repeating Steps 3 to 5. The basic assumption of this approach is that the CNN in Step 3 can reduce artifacts by replacing the streaks with values that are closer to the correct physical attenuation. This filtering can even cause an unintended over-smoothing. But we expect this effect to be negated by Steps 4 and 5, which we summarize as the *reco-step*. The FP of the filtered image $f_{\text{i}}^{'}$ yields a replacement for the corrupted projection data that is more consistent with the uncorrupted projection data. Based on this, an FBP of the sinogram S_i, which is optimized with respect to its overall consistency, should again generate an image with reduced artifacts. By involving all the uncorrupted projection data in the reco-step, the potentially unintended over-smoothing in the CNN-filtering should be reduced. The workflow of the CI-MAR is visualized in figure 1.

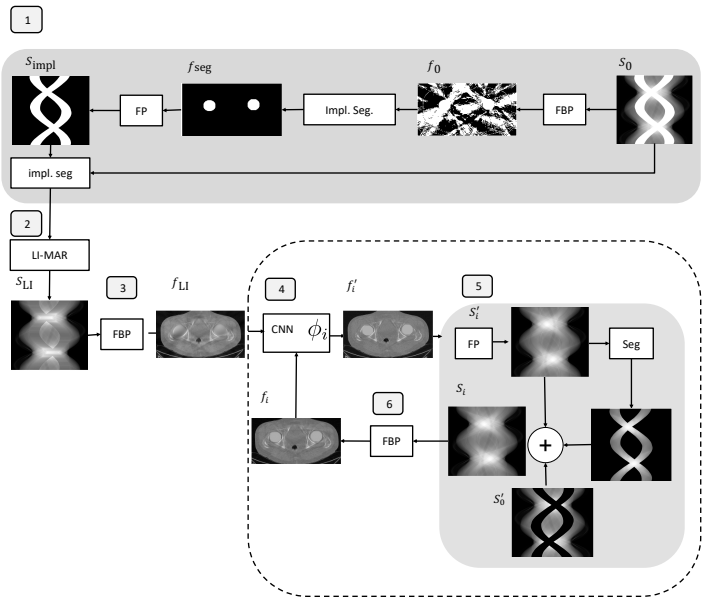


Fig. 1. Basic approach of the CI-MAR: The artifact reduction starts with a conventional LI-MAR (1-3). Its outcome is iteratively refined by a CNN-based filtering and a subsequent sinogram update (4-6).

B. Network Architecture

The applied CNN is shown in figure 2. Its main features are based on the work of Zhang et al. [5]. The input layer of our architecture handles a crop I_{cor} of the corrupted image with a size of 64×64 pixels that is obtained by a prior LI-MAR. This input is propagated through 5 consecutive convolutional layers. A final residual layer connects the input of the network to the output of the last layer. The convolutional layers 1 to 4 are designed identically, each one contains 32 3×3 convolutions. The *rectified-linear-unit* (ReLU) is applied as a non-linear activation function for these layers. Layer 5 consists of a single 1×1 convolution without activation function and therefore produces a single channel output tensor as output. The final residual layer predicts an image by subtracting the output of convolutional layer 5 from the input of the network. A zero padding is applied before each convolution so that all layers can operate at the same resolution. This architecture matches the concept of convolutional neural networks. The basic assumption behind the architecture is that within the first four convolutional layers, image features in increasing levels of abstraction are detected. Based on these features a residual image I_{res} can be synthesized by the last convolutional layer. In general, a residual image I_{res} can be regarded as a prediction of an image error. In this case, it matches a prediction of streak artifacts that overlay an uncorrupted image I_{unc}. Therefore, an artifact-free or artifact reduced image I_{unc} can be obtained by the linear operation:

$$I_{\text{unc}} = I_{\text{cor}} - I_{\text{res}}. \tag{1}$$

This operation is executed within the residual layer whose inputs are the residual image I_{res} and the corrupted input image I_{cor}. The residual layer was added to simplify the training of the networks. It has previously been suggested that it is easier for a neural network to learn the prediction of an error than to learn the synthesis of a complete error-free outcome [6].

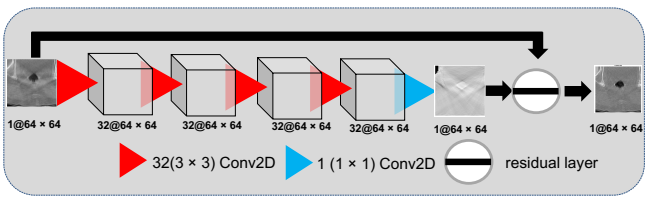


Fig. 2. The architecture of the CNN: From 4 subsequently generated feature maps, a residual image is synthesized. Within the residual layer, the prediction of the error is subtracted from the corrupted image.

C. Training

1) Training data: In order to simulate training data for the CNN, transversal slices of the hip region of clinical CT-images

were collected. Spherical objects were manually inserted into the hip joints of these images to generate an artifact-free and implant-inserted image f_{impl}. By performing an FP of the images f_{impl} and f_{seg}, two sinograms were generated. Analogous to the workflow of an LI-MAR, the projection data of the implants, stored in the sinogram S_{impl}, could be identified from the sinogram of f_{seg} and subsequently be replaced by linear interpolated values. By performing an FBP of the interpolated sinogram, an image f_{LI} is reconstructed. As a consequence of introducing interpolated values, this image contains streak artifacts and therefore resembles the outcome of a LI-MAR.

To simulate the hip endoprosthesis problem relevant for medical CT, this procedure was applied to CT images of the hip. The data consists of 20 CT volumes of the hip area of 20 different subjects. The transversal resolution of 0.977 mm and an axial resolution of 1.405 mm. From 18 of those subjects, 495 slices are collected as training data, with 25 to 31 slices per sample and a resolution of 150×220 pixels. We reserved 56 slices from the two remaining subjects as validation data. This validation data was employed to calculate a validation loss during the training of the networks and to evaluate the performance of the established CI-MAR. For the training of a CNN 495 images as training samples are a very low number. Therefore, as a tool of data augmentation, 40 randomised crops with a resolution of 64×64 pixels were taken from every LI-MAR image, as well as its artifact free version. Overall, 19800 training and 2240 validation crops x_i, y_i were generated with this scheme.

2) Training of the networks: The execution of the proposed artifact correction for a fixed number of iterations n can be described as:

$$\text{CI-MAR}_n(f_{\text{LI}}) = \psi_{\text{reco}}(\phi_n(\psi_{\text{reco}}(\dots \phi_2(\psi_{\text{reco}}(\phi_1(f_{\text{LI}})\dots))). \tag{2}$$

In this formula f_{LI} denotes the LI-MAR input image and ϕ_i a CNN that is specifically trained for the reduction of artifacts at iteration i. The function ψ_{reco} sums up the Steps 4 and 5 of CI-MAR, as shown in figure 1. To ensure that at each iteration a properly adapted filtering is performed, we train a separate network ϕ_i for each iteration i. These networks ϕ_i all match the architecture that is shown in figure 2. The first network $\phi_0^{'}$ is trained on the set of random sampled LI-MAR image crops. Following this, the LI-MAR get sampled in a sorted fashion and subsequently filtered by the now trained network ϕ_0. The complete training schedule, including the generation of the training data, is outlined in figure 3.

3) loss function: The *mean-squared-error* (MSE) is a common loss function for training a neural network on regression problems and has already been applied to CNN-MAR approaches [5] [7]. The basic MSE of a networks prediction image $y^{'}$ and its ground truth y, both with a size of $n \times m$ pixels, is given by

$$\text{MSE}(y, y^{'}) = \frac{\|(y^{'} - y)^2\|}{n \times m}. \tag{3}$$

A property of MSE is that all elements of its input arguments affect its outcome, independently of the content they describe.

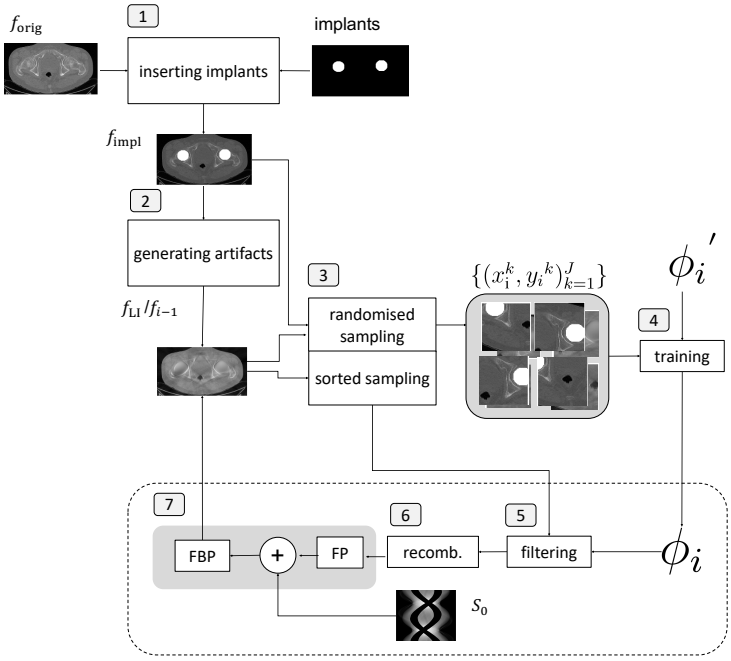


Fig. 3. Training concept:(1-2) collection with images, resembling the outcome of an LI-MAR. (3) randomised training samples are collected from the images, to train a CNN (4). By executing the CI-MAR with the trained network on the training data, the next generation of training data is produced (5-7). This data can then be used for training another CNN for the next iteration.

Specifically referring to training a CNN on the reduction of metal artifacts, the network would also have to learn to assign a metal implant the *correct* intensity values. Therefore, we want a neural network to solely *focus* on the reduction of streak artifacts that overlay diagnostically relevant anatomical structures. To enable this behavior, a mask $y_{\text{mask}} \in [0,1]^{64\times64}$ is used to determine the MSE-based loss function MSE_{mask}, in which the pixel values of the metal implants have no influence on the result and thus on the training process.

With the Hadamard product being $\odot$ this loss function can be denoted as

$$\text{MSE}(y, y^{'})_{\text{mask}} = \frac{\|y_{\text{mask}} \odot (y - y^{'})^2\|}{64 \times 64}. \tag{4}$$

D. Experiment setup

We implemented the algorithm for a total number of three iterations and therefore subsequently trained three networks. For the training of all three networks, the same parameters were chosen: The AdAm Optimizer, a batch-size of 240 and an initial learning rate of 0.001 that was reduced by a factor of 10 when the training loss stagnated for 10 consecutive epochs. We included an early stopping regularisation to prevent overfitting. This regularisation finishes the training, when the validation loss does not improve for 15 epochs. The network ϕ_1 of iteration 1 was trained until epoch 53, network ϕ_2 until epoch 43 and training of network ϕ_3 was stopped after epoch 42. The training was implemented in Keras using a Tensorflow backend and executed on a Geforce GTX Titan Black GPU with storage of 6144 MB.

E. Evaluation

We use the MSE and *Structural Similarity Index Metric* (SSIM) for the evaluation of the proposed algorithm. The SSIM is a method for measuring the similarity between two images. It is motivated by predicting the human perception of such a similarity [8]. For two images a and b SSIM can be calculated as

$$\mathrm{SSIM}(\mathrm{a},\mathrm{b}) = \frac{(2\mu_\mathrm{a})\mu_\mathrm{b} + c_1)(2\sigma_\mathrm{ab} + c_2)}{(\mu_\mathrm{a}^2 + \mu_\mathrm{l}^2 + c_1)(\sigma_\mathrm{a}^2 + \sigma_\mathrm{l}^2 + c_2)}, \qquad (5)$$

where μ is the mean of an image, σ^2 the variance and σ its covariance. To avoid singularities, the constants c_1 and c_2 are introduced. A constant c_i is given by $c_i = (k_i L)^2$ with the default factors $k_1 = 0.01$ and $k_2 = 0.03$, L specifies the dynamic range of the pixel values. The SSIM produces outcomes in the range of $[-1, 1]$, where 1 indicates a complete match between two images.

III. Results

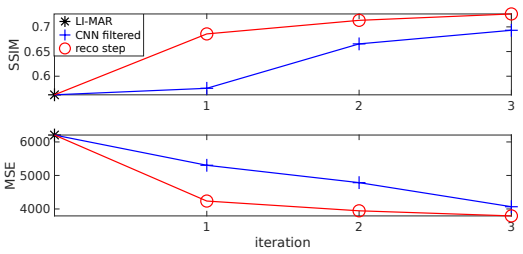


Fig. 4. Image scores for three iterations of the CI-MAR: Except for the first iteration, CNN filtering worsens both evaluation metrics as expected, sicnce over-smoothing as been introduced. However, the subsequent reco-step always improves the image quality beyond the level of the last iteration.

Figure 4 shows image scores for SSIM and MSE based on a subject of our evaluation data. For each iteration i the image scores are separately plotted: The intermediate result in form of the CNN-filtered image $f_i^{'}$ and the outcome f_i of the subsequent reco-step. To compare the performance of the algorithm with a conventional MAR method, the same image scores are plotted for LI-MAR. Both SSIM and MSE show an improvement over all three iterations. For the first epoch, the CNN-filtered image $f_i^{'}$ alone is an improvement in comparison to LI-MAR that is noticeably increased by the reconstruction step. As expected, the following two iterations show a slightly different behavior: For both consecutive iterations, the CNN-filtering reduces the image quality, compared to the result of the previous iteration. But the following reconstruction step raises both image scores above the previous iterations level. Figure 5 shows the outcome of the third iteration of our method for two slices with artifacts in different intensities. Two slices are shown with different levels of corruption, due to the varying spread of the implants at these positions. The first row shows the original CT image with an artificially inserted implant, the results of a LI-MAR are depicted in the second one. Row three shows the outcome of our iterative correction. For both levels of corruption, the streak artifacts are significantly reduced. Especially the image quality of the area between the implants benefits from the reduced artifacts.

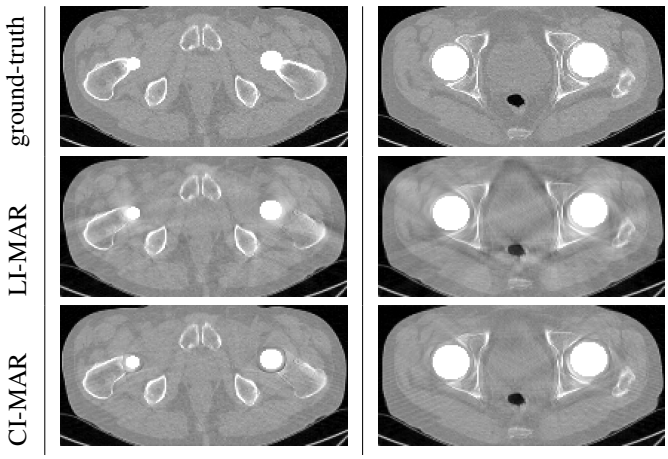


Fig. 5. Row wise comparison of metal inserted CT data as ground-truth, a conventional LI-MAR and our CI-MAR after three iterations. The images are plotted in a window of [-1000 HU, 700 HU]. Our correction significantly reduces the streak artefact, especially the image quality of the area between the hip joints profits from our correction.

IV. Discussion & Conclusion

We conclude that the proposed iterative correction scheme can significantly improve various image scores on our test data, compared to a conventional LI-MAR approach. The quality of the images profits from the iterative use of a CNN-based filtering, sinogram and subsequent image update. Both evaluation metrics indicate an over-smoothing due to the CNN-filtering. The following sinogram and image update can still lift the image quality beyond the level of the former iterations, as was assumed. In a next step, the effect of different loss functions should be evaluated, as well as more sophisticated data augmentation to enable the training of larger networks.

References

[1] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Normalized metal artifact reduction (NMAR) in computed tomography," *Medical Physics*, vol. 37, no. 10, pp. 5482–5493, Sep. 2010.

[2] M. Stille, M. Kleine, J. Haegele, J. Barkhausen, and T. M. Buzug, "Augmented Likelihood Image Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, p. 158–173, 2015.

[3] A. Mehranian, M. R. Ay, A. Rahmim, and H. Zaidi, "X-ray ct metal artifact reduction using wavelet domain l_0 sparse regularization," *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1707–1722, Sep. 2013.

[4] S. Seltzer, "Xcom-photon cross sections database, nist standard reference database 8," 1987.

[5] Y. Zhang and H. Yu, "Convolutional neural network based metal artifact reduction in x-ray computed tomography," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1370–1381, June 2018.

[6] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *CoRR*, vol. abs/1608.03981, 2016.

[7] W. Lin, H. Liao, C. Peng, X. Sun, J. Zhang, J. Luo, R. Chellappa, and S. K. Zhou, "Dudonet: Dual domain network for ct metal artifact reduction," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019, pp. 10 504–10 513.

[8] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, pp. 600 – 612, 05 2004.

Prior information for CT reconstruction from thermal data

Alessandro Piol, Patricia M. Berdon, Manuel Desco, Monica Abella

***Abstract*— In conventional cone beam computed tomography systems, the source-detector pair rotates around the patient through 360 degrees (full angular span) to acquire generally more than 360 projections. Sometimes, only a low number of projections can be acquired and/or a smaller span angle can be covered by the machine, due to movement limitations. In these scenarios the 3D image reconstruction may be affected by severe artifacts. The lack of information can be compensated by using iterative reconstruction algorithms introducing prior assumptions to help restricting the solution search space. The sample surface, which could be obtained with a 3D scanner just before the CT acquisition, would provide good prior information. The problem with a 3D scanner is the difficulty in acquiring some kind of surfaces, as shiny surfaces, in which the light scatters and bounces in uncontrollable directions. Moreover, many times patients are covered by cloaks or clothes, which prevent the scanner light to reach directly the interested surface. This work presents an alternative approach by using an infrared camera. Evaluation on small-animal data shows that two-dimensional thermal images can be used to obtain the surface mask, with the purpose to use it as prior information in iterative reconstruction algorithms.**

I. Introduction

In cone beam computed tomography (CBCT) systems, source and detector rotate simultaneously around the patient, generally covering a complete span angle of 360 degrees, to acquire a set of planar radiographies from different view angles. In this case, the three-dimensional image reconstruction can be done through simple filtered-backprojection algorithms [1], reaching a good image quality.

Manuscript submitted January 19, 2020. This work has been supported by Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación, projects "DPI2016 79075 R AEI/FEDER, UE ", Instituto de Salud Carlos III, project "DTS17/00122 ", co funded by European Regional Development Fund (ERDF), "A way of making Europe" Europe". The CNIC is supported by the Ministerio de Ciencia Innovación y Universidades and the Pro CNIC Foundation, and is a Severo Ochoa Center of Excellence (SEV 2015 0505).

A. Piol, M. Desco and M. Abella are with Dpto. Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid.

A. Piol, P.M. Berdón, M. Desco and M. Abella are with the Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España (e-mail: apiol@hggm.es, pmoreno@hggm.es, mabella@ing.uc3m.es).

M. Abella and M. Desco are with the Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, España (e-mail: manuel.desco@cnic.es).

M. Desco is with the Centro de investigación en red salud mental (CIBERSAM), Madrid, España. (e-mail: desco@hggm.es).

There are situations in which the minimum span angle for an exact reconstruction may not be possible, due to movement limitations, such as in the operating room, where the patient is connected to monitoring equipment. A high step angle or an insufficient span angle results in streak artifacts and edge distortions that can hinder the correct diagnosis. A possible solution to compensate for this lack of data is the use of new advanced iterative reconstruction methods that include prior assumptions, in order to restrict the search space and find an optimal solution. In some applications, these prior assumptions can be based on a preliminary reconstruction of a complete dataset such as in respiratory or cardiac gating [2]. However, in static CT with limited data this type of information is not available, thus it is advisable to find a different prior. Another option for the prior data is to assume the non-negativity of the solution or to restrict the search field to the field of view (FOV). In [3], the sample contour was proposed as prior information in a surface-constraint reconstruction algorithm (SCoLD), in order to further restrict the search space. Fig. 1 shows an example of the result of applying that methodology in order to reconstruct limited data (only 70 projections and 140 span). As can be seen in the figure, the use of the advanced reconstruction method including the surface mask compensates for the lack of data, achieving an image quality close to the gold standard (360 projections and 360 span).

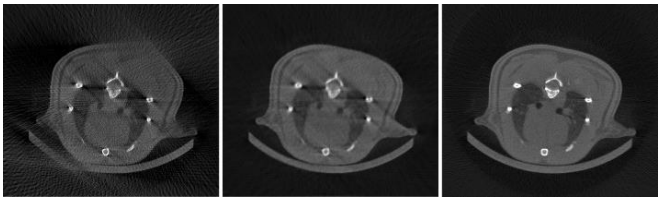


Fig. 1. Axial slice of a small-animal thorax study. Left: Reconstruction of limited data using FDK-based reconstruction. Center: Reconstruction of limited data using advanced iterative reconstruction including the surface information. Right: FDK-based reconstruction of full data (gold standard).

The work in [3] uses a 3D structured light scanner, which is a fast and easy way to measure the three-dimensional shape of an object, but provides the real surface of the sample only if the interested surface can be directly reached by the light emitted from the device. In surgery, for example, patients may be covered by surgical drapes, and the 3D light scanner would not provide the real surface of the patient. If the drape material satisfies specific spectral conditions, the patient surface could be analyzed without removing the drape, using thermal imaging, defined as the conversion of infrared radiation (heat) into visible images.

Since temperature is considered a useful marker of various diseases, thermal cameras are currently used in the medical field. Despite this, the use of thermal images in the medical field is generally limited to the 2D space; they can be used for detecting inflammation and changes in microcirculation, indices of diseases such as cancers [4] or diabetes [5]. In the case of 3D clinical thermography, reconstructions which include thermal information are only obtained by adding the thermal data to a previously obtained 3D model, as in [6].

In this work, we present the evaluation of the use of a thermal camera to reconstruct the sample surface, with the possibility of using it as prior data in advanced iterative reconstruction methods.

II. MATERIALS AND METHODS

The evaluation was done on a rodent chest study, comparing an ideal mask with the masks obtained from surface analysis, which was done in two different ways: (1) by acquiring thermal images, which were successively processed, and (2) by acquiring directly the contour information with a 3D light scanner. All animal procedures conformed to EU Directive 2010/63EU and national regulations (RD 53/2013) and were approved by the local ethics committees and the Animal Protection Board of the Comunidad Autónoma de Madrid (PROEX 241/19).

The rodent was shaved in the region of interest since the animal hair, which is not visible in the CT image, would appear in the thermal images due to its spectral features. We need detectable landmarks for helping in the thermal surface reconstruction and for the image registration process between the surface masks and the ideal mask.

Ceramic markers were chosen for the different thermal emissivity they have with respect to the animal skin, which results in a difference in color in the thermal images. Six markers were located on the back of the rat, three near the neck and three in row in the lower part of the back, avoiding the lung region, in which the CT field of view was centered. Fig. 2 shows the animal ready for the acquisitions.

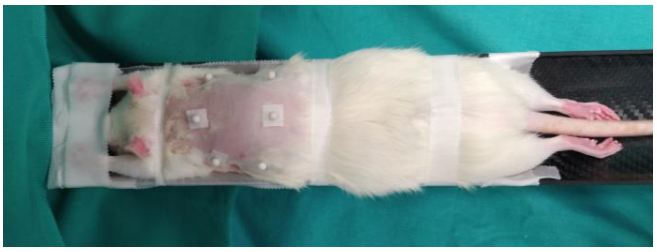


Fig. 2. Rodent set up for the acquisitions. Ceramic markers are placed over a medical tape for a better stability.

Thermal images were acquired using a *Seek thermal compact PRO* thermal camera, which has a detector of 320×240 thermal sensors array. The acquisitions were done trying to cover the rodent from different positions and distances, in order to reduce errors and deformations during surface generation. We also made a thermal acquisition of the animal covered by a simple white plastic bag. Fig. 3 shows how the plastic bag does not affect the rodent visualization, while it would prevent the surface extraction with the structured light scanner.

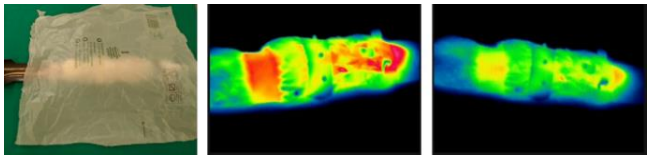


Fig. 3. Animal covered with a plastic bag (left), thermal image of the rodent under analysis not covered (center) and covered by a plastic bag (right).

The 3D reconstruction software for thermal images we used, the *3DF Zephyr* (3D Flow, Italy), is based on finding common critical points within the images, in order to obtain the different angle views from where the various photos have been taken and, hence, to combine them in a three-dimensional model. The surface was extracted using 90 thermal images. The look up table (LUT) of the images was found to be critical for the software performance, and best results were found by converting the standard thermal LUT into greyscale images.

We compared the 3D surface obtained with the thermal images with that obtained with the *Artec Eva* 3D structured light scanner, which comes with a dedicated software to process the acquired data.

Fig. 4 shows the surfaces obtained with the two methods. Meshes were cleaned and cropped to be similar in dimensions to the reference CT volume.

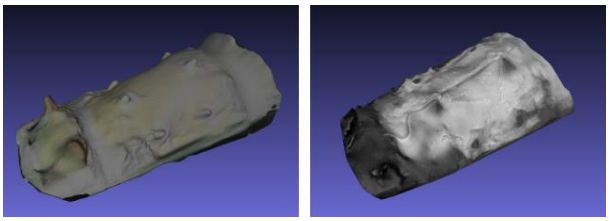


Fig. 4. Cleaned and cropped surfaces obtained with the Artec 3D structured light scanner (left) and with the Seek Thermal Compact PRO camera (right).

The animal was acquired with the *Argus* micro-CT (SEDECAL), obtaining 360 projections, covering the entire span angle of 360 degrees. The CT image was reconstructed with Mangoose, an FDK-based algorithm [7, 8], using all the acquired projections and resulting in a 514×514×574 volume without artifacts (reference volume).

The ideal mask used as reference was obtained by thresholding from the reference image (360 projections taken within a 360-degree span), as shown in Fig. 2.

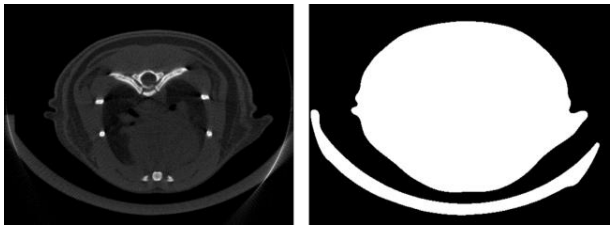


Fig. 5. Axial view of the reference FDK reconstruction (left) and the correspondent mask of the same slice (right).

We used *3D Slicer* [9] to perform a fiducial registration between the reference CT volume and, individually, the two surface models. The registration was performed in the same way for both thermal and light-scanner surfaces.

The ceramic markers were used as landmarks for the generation of a rigid registration transform including scaling,

which position was marked manually in the 3D space. Fig. 6 shows the landmarks in the 3D space, before generating the transform based on the spatial position of the blue markers from the surface model, with respect to the red ones from the reference volume.

The cost function for the transform is based on the distances between the landmarks that represent the same critical points; the final quality of the registration is quantified as the root mean square error (RMSE) in mm.

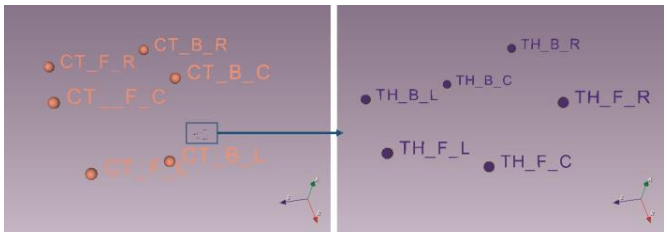


Fig. 6. Landmarks visualized in the 3D space. Red markers refer to the FDK reference volume (right panel). Blue markers refer to the 3D scanner surface model (left panel).

The registered surface model was converted into a voxelized volume with the same voxel size of that of the reference volume. Finally, this 3D surface model is cropped to be adapted to the reference FOV in the longitudinal axis and extended in order to recover missing information, not visible for the thermal camera (below the animal). For this last step, we acquired a CT of the bed alone and obtained a mask by thresholding. We added this mask to the light and thermal mask volumes and filled the gap between the bed and the surface masks.

To evaluate the accuracy of the mask obtained with both methods (thermal and light based) we compared them with the ideal mask using the Dice-Sørensen coefficient (DSC) [10]:

$$\mathrm{DSC} = \frac{2\cdot|X\cap Y|}{|X|+|Y|} \qquad (1)$$

where X and Y are the two logical volumes to be compared.

To evaluate the effect of the mask on the final reconstruction, when used as a prior, we used SCoLD to reconstruct a limited-view acquisition using 160 projections over a 160-degrees span from the complete dataset. The effect of the errors in the masks on the quality of the reconstruction was evaluated based on the RMSE between the reconstructed images and the reference FDK reconstruction (full-dose). Since the objective was only the comparison of the effects on the reconstructed image using the different masks, no effort was done to optimize the reconstruction parameters.

III. RESULTS

Fig. 7 shows the raw thermal and light masks (transparent green) overlapped to the reference CT volume, in the three projection planes. Regarding the quality of the registration step, we obtained an RMSE of 0.6581 mm and RMSE of 0.6712 mm for the thermal and light masks respectively. As we can see, the mask contours well follow those of the reference CT, in the upper part of the volume.

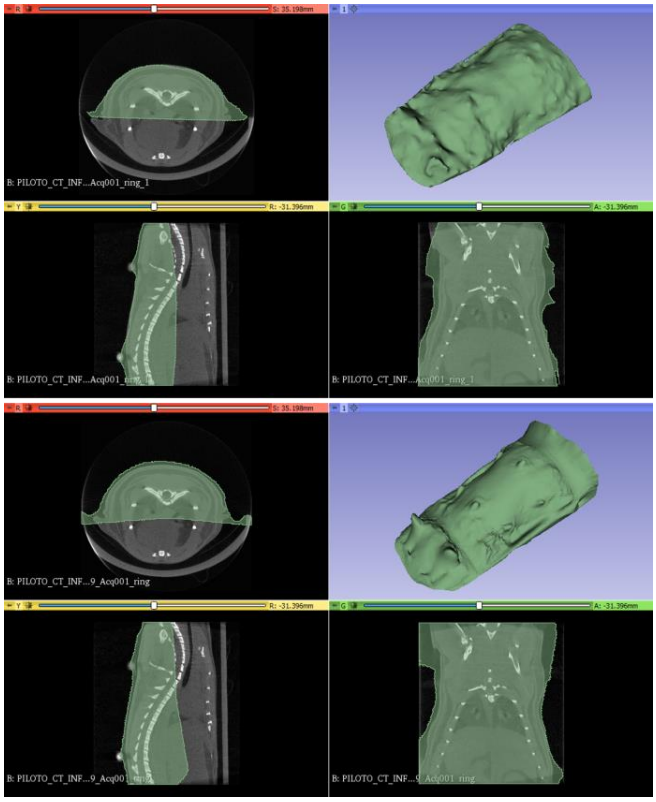


Fig. 7. Screenshot of Slicer3D interface, showing the 3D surface model obtained from the thermal data (top) and the light scanner (bottom). The raw masks are shown overlapped to the reference CT volume for the three views.

Fig. 8 shows the final masks (514×514×574 with a voxel size of 0.123 mm) after filling the missing data in the lower part of the CT.

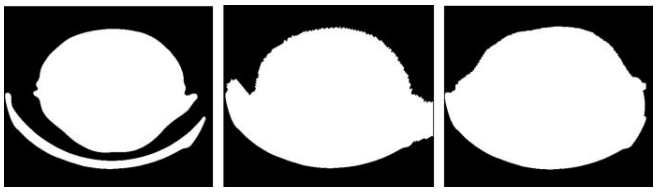


Fig. 8. Ideal reference mask (left), 3D scanner mask (center) and thermal surface mask (right).

Table I shows the Dice-Sørensen coefficients calculated between the final masks and the RMSE between the volumes reconstructed using the two surface masks and the ideal one.

TABLE I

	Light scanner	Thermal
DICE-SØRENSEN	0.8724	0.8456
RMSE	1.1743	1.4300

Fig. 9 shows the central axial slice of the different reconstructed volumes. Results with SCoLD using the two masks are similar to that using the ideal one, providing a removal of the limited-view artifact present in the FDK reconstruction (see arrow).

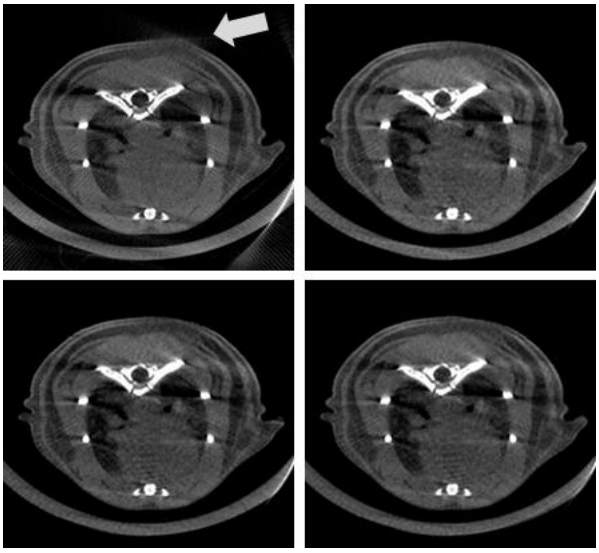


Fig. 9. Limited-view FDK reconstruction (top-left). Reconstructed volumes obtained with SCoLD using as prior information the ideal (top-right), 3D scanner (bottom-left) and thermal (bottom-right) masks. Arrow marks the contour distortion in FDK that is compensated with the information of the sample contour.

IV. DISCUSSION AND CONCLUSIONS

In this work, we have evaluated the use a thermal camera to obtain the sample surface, with the purpose of using it as prior data in advanced iterative reconstruction algorithms, and compared the results with those obtained using a structured light scanner.

Both the thermal and light scanner surfaces have missing data on the bottom part of the FOV since it is not accessible, but this is easily overcome by filling the gap using a mask of the bed, which can be acquired previously.

Errors shown in the Dice coefficients can be due to three factors. First, the lack of complete hair removal in the FOV, which results in differences in the obtained masks with respect to the ideal mask. This issue is mostly related to the veterinary case, since the body hair of a clinical patient will not affect the thermal analysis. Second, the errors in the registration were around 0.6 mm, which would also be responsible for the increase of the Dice coefficients. Third, the limited spatial resolution of the thermal camera hindered the correct identification of the markers. Further work will include the optimization of the markers (shape, size, material) in order to improve their identification and enhance the registration process.

The final reconstructions show that these errors have little effect in generating the masks, resulting in similar outcomes for ideal, thermal and light masks. Since the objective was only to compare the effects on the reconstructed image using the different masks, no effort was made to optimize the reconstruction parameters.

The preliminary results shown here suggest the possibility of using thermal images as prior information for image reconstruction of limited-view acquisitions, even when the patient is covered with certain materials, given that these materials have a specific spectral response within the camera's thermal sensors.

REFERENCES

[1] L. A. Feldkamp, L. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Josa a*, vol. 1, pp. 612-619, 1984.

[2] J. F. Abascal, M. Abella, E. Marinetto, J. Pascau, and M. Desco, "A novel prior-and motion-based compressed sensing method for small-animal respiratory gated CT," *PloS one*, vol. 11, p. e0149841, 2016.

[3] M. Abella, C. de Molina, N. Ballesteros, A. García-Santos, Á. Martínez, I. Garcia, *et al.*, "Enabling tomography with low-cost C-arm systems," *PloS one*, vol. 13, p. e0203817, 2018.

[4] H. G. Zadeh, J. Haddadnia, and A. Montazeri, "A model for diagnosing breast cancerous tissue from thermal images using active contour and lyapunov exponent," *Iranian journal of public health*, vol. 45, p. 657, 2016.

[5] A. Selvarani and G. R. Suresh, "Infrared Thermal Imaging for Diabetes Detection and Measurement," *J Med Syst*, vol. 43, p. 23, Jan 2 2019.

[6] R. F. van Doremalen, J. J. van Netten, J. G. van Baal, M. M. Vollenbroek-Hutten, and F. van der Heijden, "Infrared 3D Thermography for Inflammation Detection in Diabetic Foot Disease: A Proof of Concept," *Journal of diabetes science and technology*, p. 1932296819854062, 2019.

[7] J. G. Blas, M. Abella, F. Isaila, J. Carretero, and M. Desco, "Surfing the optimization space of a multiple-GPU parallel implementation of a X-ray tomography reconstruction algorithm," *Journal of Systems and Software*, vol. 95, pp. 166-175, 2014.

[8] M. Abella, J. J. Vaquero, A. Sisniega, J. Pascau, A. Udias, V. Garcia, *et al.*, "Software architecture for multi-bed FDK-based reconstruction in X-ray CT scanners," *Comput Methods Programs Biomed*, vol. 107, pp. 218-32, Aug 2012.

[9] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J. Fillion-Robin, S. Pujol, *et al.*, "3D Slicer as an image computing platform for the quantitative imaging network," *Magn Reson Imaging*, vol. 30, pp. 1323–41, 2012.

[10] T. J. Sørensen, *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons*, 1948.

elsa - an elegant framework for precision learning in tomographic reconstruction

David Tellenbach*, Tobias Lasser*

***Abstract*—Deep learning techniques are currently dominating medical image processing methods. However, tomographic reconstruction, as employed in X-ray computed tomography, is not very well suited to end-to-end learning, due to the very high model complexity of its inverse problem. Precision learning seeks to drastically lower model complexity by incorporating prior knowledge into deep neural networks, such as known physical models like forward and backward projection operators discretizing the X-ray transform. In this work we present an approach of integrating highly efficient low-level deep learning libraries, such as oneDNN or cuDNN, into our existing open source framework for tomographic reconstruction *elsa*. We present our elegant architecture of the framework and the efficient integration of deep learning techniques, together with a simple show case of precision learning in X-ray computed tomography.**

I. Introduction

In recent years, deep learning has taken the image processing world by storm. Almost all of the state-of-the-art methods in medical image processing tasks, such as detection, recognition, segmentation, registration, and physical simulation, are now based on deep learning approaches [6].

However, for solving inverse problems in tomographic reconstruction, progress has been a lot slower [2] due to the complexity of the reconstruction problem. Many approaches focus only on pre- or post-processing, for example to suppress noise or remove artifacts. End-to-end learning the tomographic reconstruction process has met with less success due to the high model complexity this requires. For example, for the novel imaging modality Anisotropic X-ray Dark-field Tomography where spherical scattering functions are reconstructed from X-ray dark-field measurements [11, 12], modeling the inverse problem as a deep network would require on the order of 6 quintillion ($6e18$) parameters. This is not feasible both in terms of memory required during training and inference, and in terms of training data required.

One solution approach to this is to incorporate prior knowledge into neural networks, such as known physical models, like the forward and backward projection operators modeling the X-ray transform in X-ray Computed Tomography (CT). This reduces model complexity drastically and allows tighter maximum error bounds [7]. This approach is also known as *precision learning*. Typically, this is implemented by integrating that prior knowledge into deep learning frameworks such as TensorFlow or PyTorch.

Since known operators such as the forward and backward projection operators in X-ray CT are computationally quite intensive, they are typically implemented using one or multiple GPUs to accelerate the computations in their own software frameworks [4, 1, 3, 5]. Hence, efficient integration into the deep learning frameworks, which use GPUs themselves to accelerate their computations, is quite difficult [10]. Additionally, the frequent releases of the deep learning frameworks often involve API changes breaking that integration.

In this work, we go the opposite way of integrating deep learning algorithms into our existing framework for tomographic reconstruction *elsa* [5]. Building on highly optimized low-level libraries such as the "oneAPI Deep Neural Network Library" (oneDNN) by Intel [9] and the "CUDA Deep Neural Network library" (cuDNN) by NVIDIA [8], which also drive popular deep learning frameworks such as TensorFlow or PyTorch, we aim at an elegant and very efficient framework that can easily handle precision learning tasks in tomographic reconstruction, while taking full advantage of the computational power of GPUs and specialized hardware such as tensor cores.

Our framework *elsa* is implemented in modern C++17, using the CMake build system for platform independence, featuring high unit-test coverage and continuous integration for reliability and correctness. It is published as open source under the Apache 2.0 license and can be found at https://ciip.in.tum.de/software.html. The corresponding code to this abstract currently resides in the branch "precision-learning" in the linked git repository.

In the following we briefly present the general architecture of our framework *elsa*, followed by our architecture for integrating deep learning into the framework to enable precision learning. As an example we present two simple code examples to show how tomographic reconstruction and precision learning in X-ray CT can be expressed in *elsa*. Finally we conclude with a short summary.

II. elsa Architecture

A. General architecture

elsa expresses the inverse problem of tomographic reconstruction as a generic optimization problem using series expansion,

$$\arg\min_x E(x), \quad E(x) = D(Ax, m) + \sum_{k=1}^{K} \lambda_k R_k(x)$$

where x is the quantity to be reconstructed, $D(Ax, m)$ denotes a data term and the $R_k(x)$ are regularization terms with corresponding regularization parameters λ_k. The data term $D(Ax, m)$ is expressed as a linear operator A being applied to x, the measured data m, as well as a functional

*D. Tellenbach and T. Lasser are with the Technical University of Munich, Department of Informatics, Munich, Germany.
Contact: lasser@in.tum.de

D evaluating Ax and m into a real number. As an example, A could represent the discretized X-ray transform, with D a weighted least squares functional $\frac{1}{2}\|Ax - m\|^2_{2,W}$, while R_k could be a total variation regularization term $\|Fx\|_1$, with F denoting a finite differences operator. Together with suitable approximation algorithms solving the optimization problem, this design allows a flexible, operator- and optimization-based framework for tomographic reconstruction.

These mathematical concepts are represented in code by the following classes:

- *DataContainer*. This class stores a vector (such as x or m) of arbitrary dimension, for example a linearized representation of a 3D volume using N^3 voxels on the CPU or the GPU. A *DataDescriptor* is associated with each DataContainer, providing information on the dimensionality and size of the data, and various convenience routines and operator overloads for arithmetic operations.
- *LinearOperator*. This class represents a linear operator A, providing *apply* and *applyAdjoint* methods to compute Ax and $A^T y$, respectively, and operator overloads for concatenating operators. Derived classes provide actual operators such as *Scaling*, *FiniteDifferences*, or *JosephsMethod*, which implements a discretized X-ray transform. Vertically or horizontally stacked versions of other operators are supported by *BlockLinearOperator*.
- *Functional*. This class represents a functional $f : x \mapsto f(x) \in \mathbb{R}$ mapping a vector x (a *DataContainer*) to a real value. Derived classes implement, for example, various norms such as *L2NormPow2* or *L1Norm*. Together with *Residual* representing the term $Ax - m$ for a given *LinearOperator* A (which may be a composite of other linear operators via operator overloading) and *DataContainers* x and m, these functionals enable the formulation of concrete data terms such as $D(Ax, m)$ and regularization terms $R_k(x)$.
- *Problem*. This class and its derivatives tie together a functional $D(Ax, m)$ for the data term together with functionals for regularization terms $R_k(x)$ into an optimization problem. Examples are classes like *WLSProblem* and *TikhonovProblem*. Routines are provided to compute the value, gradient, or Hessian of the data and regularization terms for use in first- or second-order methods.
- *Solver*. This base class allows the computation of an approximate solution to a given *Problem*. Derived classes include, for example, *CG* implementing the conjugate gradient method, or *StochasticGradientDescent*. Each solver class will try to convert a provided *Problem* into a form that it can handle (such as *QuadricProblem* for *CG*), otherwise it will fail.

It should be noted that each class modeling a mathematical expression knows how to evaluate its value as well as its first and second derivatives, with automatic appropriate applications of the chain rule.

Various other classes provide support, for example projective *Geometry* to model acquisition geometries, input/output helpers, as well as convenience classes such as *TrajectorGenerator* and *PhantomGenerator* that enable easy implementation of toy problems without requiring availability of real data.

B. Deep learning architecture

In order to integrate deep learning techniques into the *elsa* framework, we chose a layer-based approach to represent neural networks, which mirrors the low-level libraries such as oneDNN [9] or cuDNN [8]. The backend implementing the low-level layer operations efficiently is abstracted away from the frontend interface to the *elsa* library, in order to support choosing different low-level libraries at compile-time while providing an elegant user interface.

The base class *Layer* provides a generic interface, providing information on the *DataDescriptors* describing the input and output dimensionality and sizes, as well as the currently activated backend.

Derived classes currently include *DenseLayer* and *ConvLayer* for dense and convolutional layers, *PoolingLayer* for pooling operations, *ActivationLayer* to support various activation functions, as well as *LRNLayer* for local response normalization and *SoftmaxLayer* to implement the softmax function.

Each layer has a corresponding backend layer that is chosen at compile-time and which contains all essential logic to perform inference, training or learning. Figure 1 shows the hierarchy of all implemented network layers and their corresponding backend layer classes.

Various initializers that are frequently used in Deep Learning are supported by the *RandomInitializer* class. Tying the layers together is the *SequentialNetwork* class, which allows building up networks layer by layer. Due to the modular backend structure, network inputs can be propagated through the backend layers, reducing the need for expensive data copies.

Inference is directly supported by the network via the functions *forwardPropagate* and *getOutput*, while training (using solvers, see below) is support by *train*.

Of particular importance for precision learning is the *FixedLayer* class, that allows wrapping a *LinearOperator* for inclusion as a layer without trainable parameters into the neural network, with appropriate gradients for back-propagation. This design allows to integrate already implemented operators into a common Deep Learning setup without any need for reimplementations.

For network training, several loss functions are implemented such as *MeanSquareError* or *CrossEntropy*. Training itself is done via *StochasticGradientDescent*, with optional support for batch-training. The popular optimizer *ADAM* is available as well to improve training performance.

In the following section we will demonstrate how these class can be used to implement deep neural network inference and training, as well as how to formulate precision learning network architectures.

III. Show case

First we will show a brief example program illustrating the use of *elsa* for tomographic reconstruction, followed by a simple show case illustrating how to perform precision learning with our deep learning integration in *elsa*.

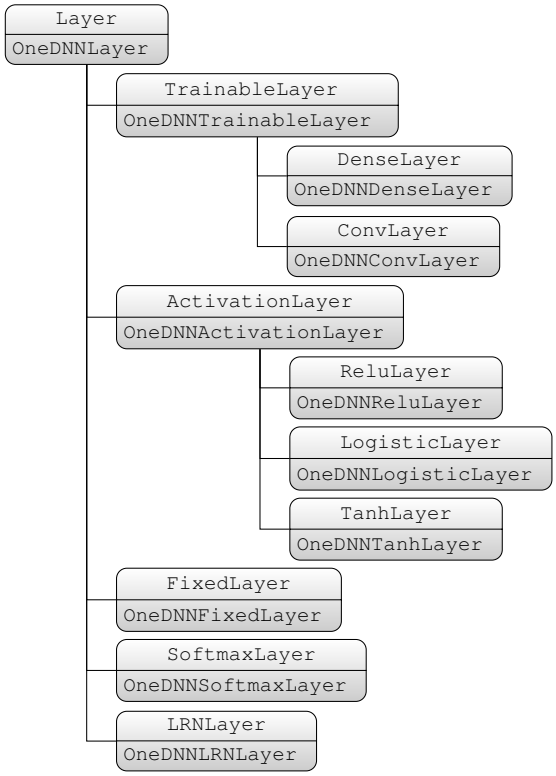


Fig. 1: Class hierarchy for interface network layers and their corresponding backend classes for the oneDNN backend.

A. Tomographic reconstruction

The following example code shows how to setup a 2D Shepp-Logan phantom, how to simulate a sinogram, and how to perform a reconstruction.

```
// generate 2d phantom of size 256x256
auto size = IndexVector_t::Constant(2, 1, 256);
auto phantom = PhantomGenerator<real_t>::
    createModifiedSheppLogan(size);

// setup circular trajectory with 180 angles
index_t numAngles{180}, arc{360};
auto [geometry, sinoDescriptor] =
    CircleTrajectoryGenerator::createTrajectory(
        numAngles, phantom.getDataDescriptor(),
        arc, size(0)*100, size(0));

// setup operator for 2d X-ray transform
JosephsMethod projector(phantom.getDataDescriptor(),
    *sinoDescriptor, geometry);

// simulate the sinogram
auto sinogram = projector.apply(phantom);

// setup reconstruction problem
WLSProblem problem(projector, sinogram);

// run conjugate gradient solver for 20 iterations
CG solver(problem);
auto reconstruction = solver.solve(20);
```

B. Precision learning

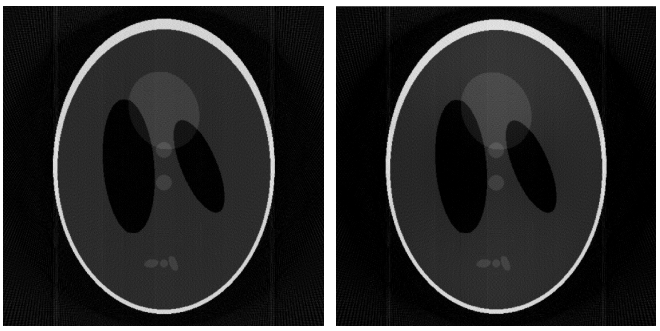
As a simple show case of our deep learning integration into *elsa*, we show how filtered backprojection reconstruction (FBP) with a filter learned by precision learning can be expressed in *elsa*, similar to what was done already in [10].

The network architecture is very simple: the sinogram is input to a convolution layer (*ConvLayer*) modeling the filter in FBP, the backprojection is implemented as a *FixedLayer* wrapping the *SiddonsMethod* operator discretizing the X-ray transform with no trainable parameters, followed by an *ActivationLayer* implementing ReLU as a non-negativity constraint. The final network output is the reconstructed image. The convolution layer is initialized with a Ramp-Filter, which is improved during the training process.

The network architecture can be produced by this code:

```
// set phantom size to 512 x 512
auto phantomDims = IndexVector_t(2);
phantomDims << 512, 512;
DataDescriptor phantomDesc(phantomDims);

// setup geometry with 180 poses applied on a
// 360 angle
index_t numAngles{180}, arc{360};
auto [geometry, sinoDesc] =
    CircleTrajectoryGenerator::createTrajectory(
        numAngles, phantomDesc,
        arc, 51200, 512);

// setup operator for 2d X-ray transform
SiddonsMethod projector(phantomDesc,
    *sinoDesc, geometry);

// construct sequential network with oneDNN backend
// that accepts sinograms as input
auto model = SequentialNetwork<float,
              MlBackend::OneDNN>(*sinoDesc);

// set options for convolution layer
auto weightsDims = IndexVector_t(2);
filterDims << 1, 512;
auto weightsDesc = DataDescriptor(weightsDims);

auto strides = IndexVector_t(2);
strides << 1, 1;

// add Conv layer with 1 filter as described above
model.addConvLayer(weightsDesc, strides,
                   Padding::Same);

// add operator for 2d X-ray transform as FixedLayer
model.addFixedLayer(&projector);

// add ReLU as non-negativity constraint
model.addActivationLayer(Activation::Relu. 1e-10f);

// set MSE as loss function
model.setLoss(Loss::MeanSquareError);

// Use a learning rate of 0.2e-9
model.setLearningRate(0.2e-9f);

// compile the model, i.e., allocate memory,
// create backend topology etc. ...
model.compile();

// Initialize Conv layer with Ramp-Filter
model.layers(0).setWeights(rampFilter);
```

We trained our model on Shepp-Logan-like Phantoms with randomized modifications and corresponding sinograms. For learning we chose a simple mini-batch gradient decent scheme with a learning rate of $0.2 \cdot 10^{-9}$ applied in 100 iterations. *MeanSquareError* served as our model's loss function.



(a) (b)

Fig. 2: (a) shows a reconstruction of a randomized Shepp-Logan phantom using FBP with a Ramp filter. This leads to a PSNR of 23.53dB. (b) shows the reconstruction using a learned filter which yields a PSNR of 24.64dB.

The following code shows how training can be done in *elsa*:

```
// simple mini-batch gradient descent training
// with 100 epochs and a batch size of 10
model.train(100, 10, sinograms, phantoms);

// forward propagate a sinogram through the
// network to perform inference
model.forwardPropagate(sinogram);

// get reconstruction
auto inferredRecon = model.getOutput();
```

To demonstrate the learning capabilities of our model we compared the *peak signal-to-noise ratio* (PSNR) reconstructions using a Ramp-Filter and our learned filter (Figure 2). Even using our simple approach we were able to increase the PSNR from 23.53dB to 24.64dB using a trained filter.

IV. Outlook

As always, software is never a finished product, in particular when trying to achieve high standards in code quality and correctness. While our architecture is designed to use several different low-level backend libraries, as of submission of this manuscript, so far we have only implemented the oneDNN backend, with support for cuDNN still a work in progress. For more complex network architectures, we will also have to implement more layer types.

In terms of practical usability, an easily usable Python interface for the deep learning algorithms is critical, as C++ does not lend itself well to rapid prototyping. Such an interface based on the pybind11 library is scheduled as future work, once the work on the backend is completed. An important factor for practical success will be a design of the Python interface that closely mirrors popular deep learning frameworks such as TensorFlow/Keras or PyTorch, ideally allowing free interchange of, for example, network architectures.

V. Conclusion

In summary, we have presented our elegant and flexible reconstruction framework for tomographic reconstruction *elsa*, and the integration of deep learning techniques based on highly efficient low-level libraries such as oneDNN or cuDNN, in order to enable precision learning for tomographic reconstruction tasks. Our code is available as open source under the Apache 2.0 license at https://ciip.in.tum.de/software.html.

References

[1] W van Aarle et al. “Fast and Flexible X-ray Tomography Using the ASTRA Toolbox”. In: *Optics Express* 24.22 (2016), pp. 25129–25147. DOI: 10.1364/OE.24.025129.

[2] S. Arridge et al. “Solving inverse problems using data-driven models”. In: *Acta Numerica* (2019), pp. 1–174.

[3] A Biguri et al. “TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction”. In: *Biomedical Physics & Engineering Express* 2.5 (2016).

[4] A. Fehringer et al. “A versatile tomographic forward- and back-projection approach on multi-GPUs”. In: *Proceedings of SPIE*. Vol. 9034. SPIE, 2014, 90344F.

[5] T. Lasser, M. Hornung, and D. Frank. “elsa - an elegant framework for tomographic reconstruction”. In: *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. Vol. 11072. SPIE, 2019, pp. 570–573. DOI: 10.1117/12.2534833. URL: https://elsa.campar.in.tum.de.

[6] A. Maier et al. “A gentle introduction to deep learning in medical image processing”. In: *Zeitschrift für medizinische Physik* 29.2 (2019), pp. 86–101. DOI: 10.1016/j.zemedi.2018.12.003.

[7] A. Maier et al. “Learning with known operators reduces maximum error bounds”. In: *Nature Machine Intelligence* 1 (2019), pp. 373–380. DOI: 10.1038/s42256-019-0077-5.

[8] *NVIDIA CUDA Deep Neural Network library (cuDNN)*. https://developer.nvidia.com/cudnn. Accessed: 2020-01-18.

[9] *oneAPI Deep Neural Network Library*. https://01.org/dnnl. Accessed: 2020-04-29.

[10] C. Syben et al. *PYRO-NN: Python Reconstruction Operators in Neural Networks*. 2019. arXiv: 1904.13342.

[11] M. Wieczorek et al. “Anisotropic X-Ray Dark-Field Tomography: A Continuous Model and its Discretization”. In: *Phys. Rev. Lett.* 117.15 (2016), p. 158101. DOI: 10.1103/PhysRevLett.117.158101.

[12] M. Wieczorek et al. “Brain Connectivity Exposed by Anisotropic X-ray Dark-field Tomography”. In: *Scientific Reports* 8.1 (2018), p. 14345. DOI: 10.1038/s41598-018-32023-y.

Reconstruction of Blood Vessel Paths from Sparse Cone Beam Projection Images using Neural Networks and Ray-Tracing

Maximilian Wattenberg, Lasse Hansen, Philipp Klein, Mattias P. Heinrich, Maik Stille, and Thorsten M. Buzug

***Abstract*—X-ray C-arms are routine instruments in interventional procedures, providing instant information on anatomical structures. Thereby, physicians often rely on 2-dimensional projection images taken from different view angles estimating the spatial arrangement. Assumptions based on these partial information are prone to errors and may risk a successful outcome of the intervention. Modern X-ray C-arms track their spatial positions or can even move along specified trajectories. This enables the reconstruction of a patient's anatomy and provides volumetric image data with reasonable speed. However, due to safety reasons the C-arms' motion speed is limited. Furthermore, a certain minimum number of images have to be acquired in order to achieve satisfying reconstruction results. This paper aims to introduce a tomogram-less reconstruction algorithm, which calculates vessel paths from very few projection images for guidance during interventions.**

***Index Terms*—Blood vessel, C-arm, reconstruction, ray-tracing, X-ray, structured prediction, convolutional neural networks**

I. Introduction

VASCULAR illnesses like coronary artery disease, thrombosis, or aneurysms are potentially life-threatening and indicate an angiographic examination. Different methods for angiography on various imaging modalities like magnetic resonance imaging (MRI) or computed tomography (CT) are available in the clinical routine. Whereas MRI- and CT-angiography provide detailed 3-dimensional representations of a vascular tree, they come with the drawback of low temporal resolution or higher exposure to ionizing radiation. Especially for interventional angiography fast and low-dose methodologies providing a 3-dimensional vessel tree representation are desired. At present, in order to navigate through arteries physicians often rely upon multiple separately acquired 2-dimensional projections and therefore have to estimate the true 3-dimensional vessel path. The works of Navab, Wiesent, et al. introduce an approach, for which multiple projection images are acquired using an X-ray C-arm. Subsequently, a tomogram is reconstructed with a cone-beam reconstruction approach [1], [2]. Even though their approach requires far fewer projection images than complying the Nyquist-Shannon sampling theorem would require, still 50-100 projections and a few minutes for the reconstruction are required [2], [3]. Tomogram-less alternatives are topic of research [4]. In the following, we will introduce a proof of concept for a tomogram-less reconstruction approach, which aims to calculate centerlines of vessels from 5 projections. A clinical application with such features would increase the quality of interventional surgeries as the 3-dimensional navigation paths could be consulted in short periods.

This work was funded by the YXLON International GmbH.

M. Wattenberg, M. Stille, and T. M. Buzug are with the Institute of Medical Engineering, Universität zu Lübeck, Lübeck, Germany (e-mail: {wattenberg, stille, buzug}@imt.uni-luebeck.de).

L. Hansen and M.P. Heinrich are with the Institute of Medical Informatics, Universität zu Lübeck, Lübeck, Germany (e-mail: {hansen, heinrich}@imi.uni-luebeck.de).

P. Klein is with the YXLON International GmbH, Hamburg, Germany (e-mail: klein@hbg.yxlon.com).

Manuscript received January 19, 2020.

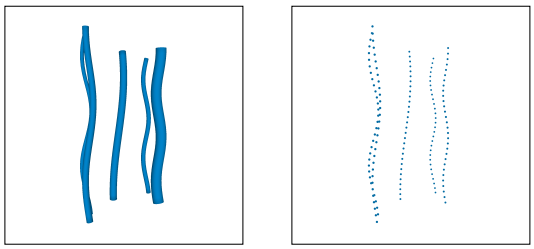


Fig. 1. Polygon mesh vessel model (a) and the according spline vertices (b).

II. Material and Methods

The following sections describe: the simulation process of the digital blood vessel phantoms used throughout the experiments, the artificial intelligence-driven detection of the vessel centers in the projection images, the reconstruction approach for calculating the vessel paths in the 3-dimensional space, and the methods used for evaluating the accuracy.

A. Data generation

In order to prove the concept of a tomogram free vessel path reconstruction, simplified vessel models were generated synthetically. The vessel models consist of multiple isolated pipe-like structures represented as polygon-surface meshes. These models were scanned using a circular trajectory generating the desired projection images. As the vessel reconstruction relies on the detection of the vessel center in the projections a neural network was trained in order to undertake this task. The polygon-mesh generation, the virtual scanning, and the training data generation are described in the following paragraphs:

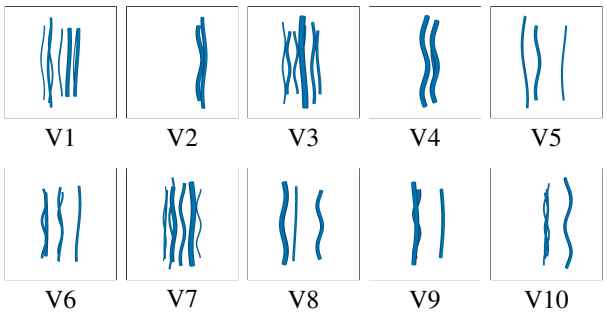


Fig. 2. Side view of a subset of the generated phantoms. The models may serve as a visual reference of a specific model in the results section.

1) Digital phantom: The vessel models are generated in such a way that the modeling process provides the centerline of the vessel as ground truth and the volumetric polygon mesh as a scannable model.

The modeling process starts with the generation of the vessel centerline, which is modeled as a helix-like 3-dimensional spline. Whereas helices often consist only of a single circular path, the orbit of the used helix is the sum of three helices with different amplitudes, phases, and frequencies. This path calculation can be seen as the sum of Fourier coefficients. The sample coordinates of the spline are defined by:

$$\begin{aligned}
\varphi_m(n) &= \frac{2(n-1)c_m\pi}{N} + \theta_m, \\
x(n) &= \sum_{m=1}^{3} A_m \cos(\varphi_m(n)), \\
y(n) &= -\frac{L}{2} + (n-1)\frac{L}{N-1}, \\
z(n) &= \sum_{m=1}^{3} A_m \sin(\varphi_m(n)),
\end{aligned}$$

where $n \in N$ denotes the sample index, N represents the number of samples, L is the length of the vessels. The variables for the number of cycles per vessel length: $c_m \in [0.2, 3.0]$, the amplitude of the helix: $A_m \in [0.8, 1.8]$, and the phase of the helix: $\theta_m \in [0, \pi)$ are three randomly chosen parameters out of the specified intervals. The specified spline is widened into a pipe-like polygon surface mesh by creating a circle[1] of the size $r \in [0.8, 1.2]$ and assign it to the spline using Blender 2.81 (Blender Foundation, Amsterdam, Netherlands). This procedure is performed for up to 8 single vessels, which subsequently are set apart and joined into a single phantom containing multiple vessels (see Figure 1).

2) Projection images and training data: The phantoms are scanned under a sparsely sampled circular trajectory of 18 steps and an angle increment of $10\,^\circ$. The scans are performed using aRTist (Bundesanstalt für Materialforschung und -prüfung, Berlin, Germany), an X-ray simulation software often used for research in the field of industrial CT. One example of the simulated projection images is shown in Figure 3.a. Furthermore, the splines, which were used for the generation of the blood vessels, are projected under the same geometries onto the detector using a projection matrix derived from the scan geometry. The projected spline is converted into a binarized projection image, providing the ground truth for the desired vessel center detection (see Figure 3.b).

As the vessel center detection relies on a data-driven approach, a base data set and an additional training data set were generated. For the base data set 20 phantoms were used and scanned in the sparse trajectory resulting in 360 image pairs. The additional data set used 64 vessel phantoms and the same trajectory leading to a total number of 1872 training data pairs. An image size of 512×512 pixels was chosen.

[1]In Blender the circle is represented in the Bezier format.

B. Blood vessel detection

The task of detecting the centers of a broad image structure suggests the use of skeletonizing algorithms as described in [5] or [6]. Standard skeletonizers rely on a binarized image, whereas the potentially overlaying blood vessels would require a sophisticated image processing, which separates the blood vessels into several image layers. However, the latest developments in the use of neural networks for image segmentation, e.g. the U-Net architecture, established a quick alternative, given the fact suitable training data are at hand [7]. The architecture of the neural network used for the centerline detection is based on a U-Net like architecture. The encoder part consists of eight convolutional layers, where every second layer uses stridden convolutions to downsample the feature maps by a factor of two. With each downsampling step, the filters are doubled starting with 16 filters in the first layer. To upsample the feature maps in the decoder part we employ four layers of bilinear interpolation and succeeding convolutions, which yields a final feature map at the resolution of the original input image. Skip connections at corresponding resolution levels enrich semantic information with low-level features. The blood vessel centerlines are predicted from the last feature maps using a final 1×1 convolutional layer. All other convolutional layers are applied with a kernel size of three and are followed by a LeakyReLU activation function with a slope of 0.1. To supervise the training of the convolutional neural network (CNN) we use a loss function L that combines the binary cross-entropy (BCE) and Dice loss ($DICE$) to

$$L = 0.001 \cdot BCE + DICE,$$

which was shown to produce sharp edges in previous works on deep learning based structured prediction [8]. The network was trained for 120 epochs. ADAM (adaptive moment estimation) optimization with an initial learning rate of 0.001 is used to stabilize training, the learning rate is decayed by a multiplicative factor of 0.1 after 60 and 90 epochs [9]. Additionally, batch normalization is applied throughout all layers with a batch size of 16.

C. Blood vessel reconstruction

The vessel center reconstruction uses the tomogram-less object edge detection algorithm described in detail in [10], with minor modifications to suit the different input data of detected vessel center projections. In general, the algorithm

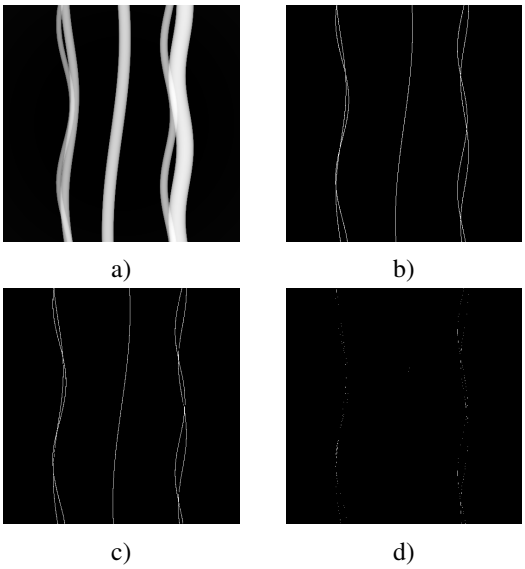


Fig. 3. The projection image of one of the phantoms (a) is paired with the corresponding projection of the ideal vessel centerline (b) and serve as training data for the data driven detection algorithm. The detection result (c) of the trained neural network is subtracted from the ground truth. The difference image (d) indicates the quality of the detection exemplary.

locates the position of a specific object feature in the 3-dimensional space with a ray-tracing approach, thus inverting the projection matrix. The object feature can be of different types, like object edges or vessel centers. The ray-tracing spawns potential locations of the object features as a point cloud. As the inversion of a projection matrix is an ill-posed problem the point clouds of potential feature locations may have false-positive elements. Therefore, multiple point clouds of potential feature locations need to be generated from different pairings of projection geometries. These point clouds are merged by clustering. Finally, the clusters are filtered by a threshold over the number of contained potential feature locations.

For the reconstruction of the vessel centerlines a subset of 5 of the 18 projection geometries, used for training, is taken into account. The ray-tracing is performed for all pairing combinations of the 5 geometries. A threshold was manually chosen on the final clustering result.

D. Evaluation

The following evaluation methods were either performed for the accuracy of the CNN-based skeletonizing approach or for the reconstruction algorithm.

1) Vessel center detection: For the evaluation of the CNN the output images were compared with the ground-truth images driven from the projection of the initial vessel centerline. The detection result for an example image and the difference image to the corresponding ground truth are calculated and presented in Figure 3. As a measure for the accuracy of the vessel center detection the ratios of the true positive positioned pixels were calculated. In order to investigate potential influences of either the different projection geometries or the different models, the percentage of the true positives were evaluated for the first 10 phantoms and the 18 geometries separately. Thus, the mean percentage over the 10 models calculated for each of the read-out angles and the mean percentage over the angles were calculated for each phantom (see Figures 4 and 5). To produce the centerline predictions for the complete training data set (N=1152) we performed a 5-fold cross-validation (fold/phantom ids: 1/1-13, 2/14-26, 3/27-39, 4/40-52, 5/53-64), where the CNN was trained on the base data set (N=360) and four of the five folds of the training data set and outputs were predicted for all images in the remaining fold (see Table I).

2) Vessel tree reconstruction: The vessel center reconstruction results in a point cloud, where each coordinate lies along the centerline of the vessel. The reconstruction approach is evaluated using a subset of the Hausdorff distance, which is defined as

$$h(P,Q) = \max\left\{\sup_{p\in P}\inf_{q\in Q}(||\mathbf{p}-\mathbf{q}||), \sup_{q\in Q}\inf_{p\in P}(||\mathbf{p}-\mathbf{q}||)\right\},$$

where P and Q represent the reconstructed point cloud and the samples of the splines and $||\mathbf{p}-\mathbf{q}||$ represents the Euclidean distance between any two points of the point clouds. Figure 6 shows the infima of both point clouds as box-plots for the first ten vessel phantoms.

III. Results and discussion

The results for the evaluation of vessel center detection in the projection images and the accuracy of the reconstruction will be discussed in the following two subsections.

A. Vessel center detection

A first impression of the quality of the vessel center detection can be gained considering Figure 3. The four images show: a) the simulated projection image of one of the vessel phantoms, b) the ideal projection of the spline, which was used during modeling, c) result of the vessel center detection, performed by the CNN, and d) the difference image between b) and c).

Figure 4 shows the mean percentage of true positive detected vessel center pixels for multiple vessel phantoms. The ratio of the true positives is plotted for the 18 different projection geometries, which are indicated by the corresponding read-out angle, separately. The average percentage of the true positively detected pixels is above 93.09%. The data show no influence of the projection geometry on the quality of the detection procedure, indicating independence of the CNN regarding the scan geometry.

The generalization capability of the CNN can be derived from the Figure 5. For each of 10 different phantoms the mean percentage of true positive pixels averaged over all of its projections, is shown. The accuracy of the vessel center detection deviates slightly more than the angle related counterpart. This is most likely due to the different complexity levels in between the phantoms, wherein the number of simulated blood vessels may vary between 2 and 8. It is evident, that a higher number of blood vessels intersect each other in projection images more

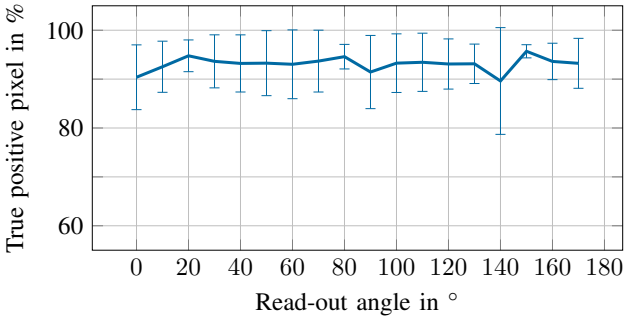


Fig. 4. Mean percentage of true positive detected vessel center pixels over 10 vessel phantoms plotted against the circular trajectory positions.

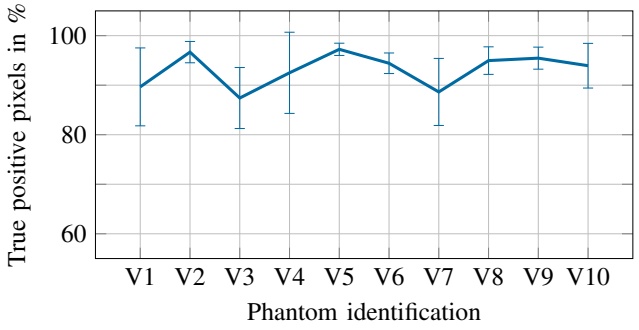


Fig. 5. Mean percentage of true positively detected vessel center pixels over the 18 projection geometries shown for 10 phantoms individually.

TABLE I
RESULTS OF THE 5-FOLD CROSS-VALIDATION.

Phantom	True positives in %
V_1 to V_{13}	93.6
V_{14} to V_{26}	92.7
V_{27} to V_{39}	94.9
V_{40} to V_{52}	94.8
V_{53} to V_{64}	93.6

often, impeding the detection. Also the different diameters of the single vessels may have an influence, as small blood vessels have higher transparency as larger ones and therefore might be partially hidden by those. Table I summarizes the result of the cross-validation and show very similar results overall combinations.

B. Vessel tree reconstruction

Figure 6 shows the similarity of the reconstruction result and the simulated vessel center spline for the first 10 phantoms indicated by V_1 to V_{10}. For each phantom the infinitum from the reconstructed vessel center to the ground truth spline is represented by the dark box-plot, whereas the infinitum from the ground truth to the reconstruction result is represented by the bright box-plots. Even though the reconstruction results have multiple outliers, the majority of the reconstructed pixels are below 5 pixels to the ideal vessel centers. Taking the high resolution of the simulated detector of 512×512 detector elements into account, this shall provide the physician with an approximation of the true vessel positions.

IV. CONCLUSION

In the present work, the concept of a tomogram free vessel reconstruction approach for navigation during minimal invasive intravascular procedures has been introduced. For simulated phantoms the CNN showed high detection accuracy of the vessel centers in the projection data. The ray-tracing algorithm was able to calculate the vessel path in 3-dimensional space from only 5 projection images, with the most data points having a small distance to the ideal position. More comprehensive adjustments to the reconstruction algorithms will most likely reduce the number of outliers, which will increase the quality of the reconstruction result.

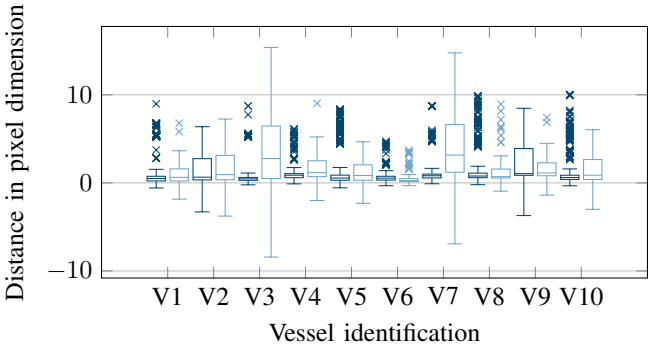


Fig. 6. The box-plots show the infinitum between the true vessel center location and the reconstruction result for different vessel phantoms (V_1 to V_{10}). Dark box-plot: infinitum from reconstructed vessel centers to the ground truth. Bright box-plot: infinitum from ground truth to reconstructed vessels.

REFERENCES

[1] N. Navab, A. Bani-Hashemi, M. S. Nadar, K. Wiesent, P. Durlak, T. Brunner, K. Barth, and R. Graumann, "3D reconstruction from projection matrices in a C-arm based 3D-angiography system," in *International Conference on Medical Image Computing*. Springer, 1998, pp. 119–129.

[2] K. Wiesent, K. Barth, N. Navab, P. Durlak, T. Brunner, O. Schuetz, and W. Seissler, "Enhanced 3-D-reconstruction algorithm for C-arm systems suitable for interventional procedures," *IEEE transactions on medical imaging*, vol. 19, no. 5, pp. 391–403, 2000.

[3] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[4] S. Çimen, A. Gooya, M. Grass, and A. F. Frangi, "Reconstruction of coronary arteries from X-ray angiography: A review," *Medical image analysis*, vol. 32, pp. 46–68, 2016.

[5] T. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, 1984.

[6] T.-C. Lee, R. L. Kashyap, and C.-N. Chu, "Building skeleton models via 3-D medial surface axis thinning algorithms," *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 6, pp. 462–478, 1994.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing*. Springer, 2015, pp. 234–241.

[8] R. Deng, C. Shen, S. Liu, H. Wang, and X. Liu, "Learning to predict crisp boundaries," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 562–578.

[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[10] M. Wattenberg, P. Klein, M. Stille, and T. M. Buzug, "3D tomogram-less reconstruction for industrial computed tomography using a clustering approach," in *Book of Abstracts Digital Industrial Radiology and Computed Tomography*, 2019.

Ring Artifact Reduction in Sinogram Space Using Deep Learning

M. Nauwynck[*], *S. Bazrafkan*[*], *A. H. van Heteren*[†‡§], *J. De Beenhouwer*[*], *J. Sijbers*[*]

[*] imec-Vision Lab, Department of Physics, University of Antwerp, Belgium
[†] Sektion Mammalogie, Zoologische Staatssammlung München, Staatliche Naturwissenschaftliche Sammlungen Bayerns, Munich, Germany [‡] GeoBio-Center, Ludwig-Maximilians-Universität München, Munich, Germany
[§] Department Biologie II, Ludwig-Maximilians-Universität München, Munich, Germany

***Abstract*—Ring artifacts are a type of reconstruction artifact that is common in X-Ray CT. Recently, methods based on deep learning have been proposed to reduce ring artifacts in reconstructed images. These methods are dependent on the choice of reconstruction algorithm and often rely on a polar coordinate transformation. Methods that directly operate in sinogram space do not feature this dependency, do not require a coordinate transformation while also operating in the space where ring artifacts originate.**

In this paper, we propose a deep neural network with a custom loss function that operates exclusively in sinogram space for ring artifact reduction. Results on real and simulated data show that our method has similar or better performance compared to other ring artifact reduction techniques that also operate exclusively in sinogram space.

***Index Terms*—Deep learning, Tomography, Ring Artifact**

I. Introduction

RING artifacts are a type of reconstruction artifact that is common in X-ray CT. They appear as a series of lines that overlay the sinogram. Ring artifacts typically originate from the X-ray hardware. In a synchrotron, for instance, the X-ray detector, source and/or the monochromator may cause ring artifacts [1, 2] due to variations in detector element sensitivity [3, 4], defective elements [5], changes in sensitivity between flatfield corrections [4], insufficient flatfield correction [5], variable (non-linear) response to beam hardening [4–6], dusty or damaged scintillator screens [5, 7], varying scintillator thickness [2, 4, 8], imperfections in the optical coupling system [2] and drift or mechanical vibrations of a monochromator [2]. After tomographic reconstruction, these deviations appear as (partial) concentric circles [4] (see Fig. 1). As these rings have an adverse effect on further image analysis, it is desirable to reduce them.

Common methods that attempt to reduce ring artifacts can be categorized in three groups: methods that operate in sinogram space (e.g. [1–3, 5, 7, 8]), image space (e.g. [3, 4, 9]) or both spaces (e.g. [10–12]).

Recently, deep learning methods have been proposed to reduce ring artifacts in CT images [11–13]. Wang, Li, and Enoh [11] use a generative adversarial network with adversarial loss, perceptual loss and a unidirectional relative total variation loss. Reconstructed images are first transformed from cartesian to polar coordinates as a preprocessing step. The

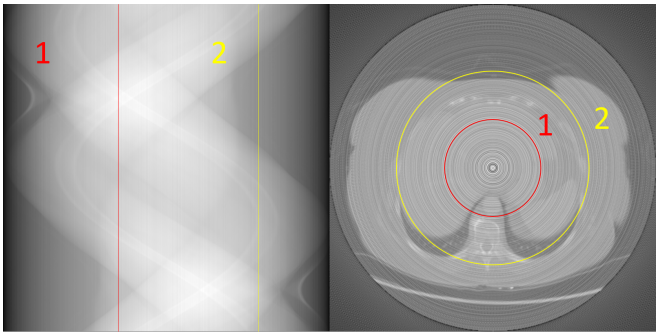


Figure 1: Artifacts on sinogram and reconstructed image.

method by Chao and Kim [9] also operates in image space with use of polar coordinates. Ring artifacts are first detected by using a smoothing operation, after which a trained radial basis function neural network is applied to reduce the artifacts on a pixel-basis. Finally, Fang, Li, and Chen [12] combined both spaces in their work to estimate an artifact map, which is subsequently subtracted from the initial reconstruction.

Multiple disadvantages are encountered when using methods that operate exclusively in image space. One is the need for correctly selecting the center for the coordinate transformation. A second limitation arises when working with reconstructed images of low quality. Few-view or limited angle CT are examples where the reconstructed image quality is significantly degraded and the artifacts have a less distinct pattern. A final limitation when working exclusively in image space relates to the fact that ring artifacts originate in sinogram space. They can be viewed as signals that are local, i.e. at some detector channel. After tomographic reconstruction, this locality is lost. Finally, the appearance of ring artifacts in image space depends on the tomographic reconstruction method that was applied to obtain the image. This dependency implies that separate deep neural networks need to be trained for different reconstruction algorithms as the reconstruction of a sinogram with the same line artifact may vary significantly depending on the choice of algorithm.

To avoid the above mentioned, a deep learning approach is proposed that works exclusively in sinogram space. To the best of our knowledge, deep learning has not yet been applied to sinogram preprocessing for ring artifact reduction. The

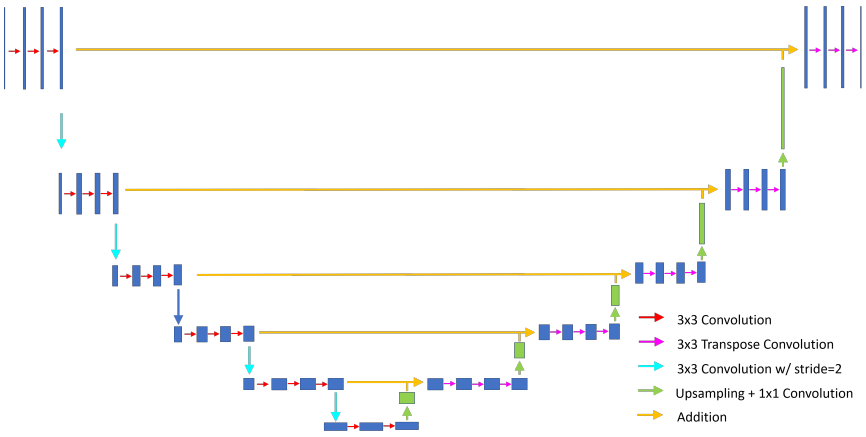


Figure 2: Network architecture

proposed method is independent of the choice of reconstruction algorithm, does not rely on a coordinate transformation and operates in the space where the artifact originates. To achieve this, a deep neural network, similar to U-Net [14], is implemented and trained on simulated data. Results on both simulated and real data are compared to other methods that operate in sinogram space. Our findings support the relevance of exclusively using sinograms in ring artifact reduction.

II. Methodology

Similar to other papers that operate in sinogram space to reduce metal artifacts [15, 16], a deep neural network was implemented that is based on a U-Net architecture [14]. A deep neural network (DNN) is a model that consists of several processing units also known as layers. These layers could be convolution/deconvolution, pooling/unpooling, and dense operations. In the current study, a DNN similar to U-Net was implemented (Fig. 2) to process sinogram data as to reduce ring artifacts. It is a fully convolutional DNN with a downsampling operation after every three convolutional layers. Downsampling is accomplished by using a 3×3 convolutional operation with a stride equal to 2. Note that the number of channels is doubled after each downsampling operation in order to preserve the information flow throughout the network. The second half of the network consists of deconvolution and upsampling layers to recover the signal back to the input size. Note that every upsampling operation is followed by a 1×1 convolutional layer. There are several skip connections inside the network architecture indicated by yellow arrows in Fig. 2. The skip connections are not used in the original implementation and make the model more parameter-efficient. The skip connections transfer the high-frequency bands from the early stages into the deeper layers which plays an important role in preserving the sharpness of the output. The ReLU activation function [17] is applied after each convolution, except for the last layer which has no activation function.

The parameters were optimized using a combined loss function. The first component, l_1, is a normalized L1 loss between the network output and the target tensor, defined by Eq. (1), with N the number of samples and y and $\hat{y}$ the expected and predicted result, respectively.

$$l_1 = \frac{\sum_{n=1}^{N} |y_n - \hat{y}_n|}{\sum_{n=1}^{N} |y_n|} \tag{1}$$

The second loss, l_2, is given by Eq. (2) wherein a normalized L1 loss is computed after convolving the difference between the predicted and expected result with the vertical kernel, w, of the Sobel operator.

$$l_2 = \frac{\sum_{n=1}^{N} |w * (y_n - \hat{y}_n)|}{\sum_{n=1}^{N} |w * y_n|} \tag{2}$$

The final loss function, l, is the weighted summation of the two separate components as shown in Eq. (3).

$$l = \alpha l_1 + \beta l_2 \tag{3}$$

Normalizing the loss functions allows to easily set and understand their relative contribution. Our results were acquired with weights α and β set to 0.8 and 0.2, respectively. These values were selected through experimentation.

The Adam optimizer [18] was utilized in the training procedure with an initial learning rate equal to 3×10^{-3} and other hyperparameters left to their default values. A dynamic learning rate was used during training, reducing the learning rate by 10% if the loss on training data had not decreased in the 5 previous epochs. The model was built in Python version 3.6 using Tensorflow version 2.0 [19] as the framework of choice and was trained for 180 epochs on a NVIDIA Tesla K80 GPU using a batch size of 128.

III. Experiments

A. Dataset

In order to provide the training database, ring artifacts were simulated on clean images. To this extent, images from the The Cancer Imaging Archive [20] were used (45640 medical images, resized to 256 by 256 pixels). Every image, no matter the modality, was included in the dataset. This ensured that the network observed a large variation of sinograms during training. Fig. 1 shows an example of a sinogram with simulated ring artifacts and the reconstructed image.

Sinograms were simulated in a parallel beam set-up through use of the ASTRA Toolbox [21] and line artifacts were added onto them. To accurately model statistical parameters of the artifact magnitude, the available datasets from Tomobank [22] with ring artifacts were analyzed. A smoothing method, part of the Tomopy toolbox [23], was applied to an image as to strongly reduce ring artifacts. By comparing the smoothed and original image, a normal distribution for the mean artifact magnitudes was estimated. Corresponding variances were estimated as well. To generate artifacts, a mean artifact magnitude value was randomly sampled from the estimated distribution for every detector channel. Variances for each mean artifact magnitude were computed, resulting in normal distributions for every affected detector channel. The resulting normal distributions were randomly sampled to generate ring artifacts for all projection angles. Finally, the artifacts were added onto the artifact-free sinogram. The dataset was split in a training, validation and test set with 42240, 1000 and 2400 samples, respectively.

B. Evaluation

The performance of our method in terms of peak signal-to-noise ratio (PSNR) and structured similarity (SSIM) [24] was compared to that of four other methods that operate in sinogram space: wavelet-Fourier filtering (FW) [25], the generalized Titarenko's algorithm (TI) [26], a based filtering approach (BF) [1] and a smoothing filtering. Out of all the available methods in the Tomopy library, these four were selected based on their better performance on real samples with ring artifacts.

Ring artifacts in 2400 simulated sinograms were suppressed using our approach and the four reference methods. These samples were not seen by the neural network during training. The simulation principle, however, was the same as for the training samples. In all evaluations, filtered backprojection was used as the reconstruction technique. PSNR and SSIM were subsequently computed for the reconstructed images.

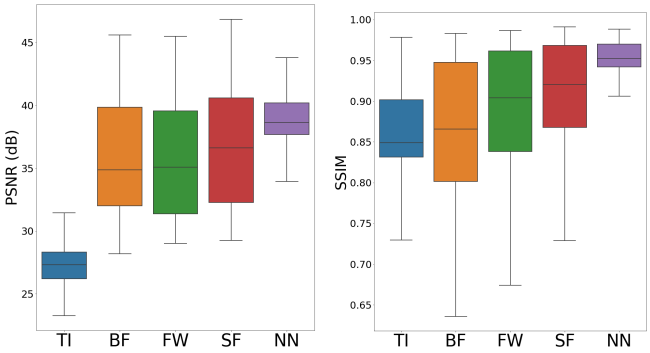
To assess the effect of the proposed method on real data, it was tested on real sinograms of a bone sample, alongside with the reference methods. The sample was acquired with TESCAN XRE's UniTOM XL scanner using a cone beam set-up.

IV. Results

A. Simulated data

Fig. 3a shows PSNR values for the different ring artifact reduction algorithms. Our method has a median PSNR that is at least 2dB higher than the reference methods. These results show that our method is able to reduce ring artifacts to a larger extent relative to the other methods. The variation seen in the results is small compared to other methods, with the exception of the generalized Titarenko's algorithm. These results indicate that our method performs well on sinograms that belong to a variety of images (differing in contrast and brightness values for instance).

Similar observations can be made when considering the SSIM values (Fig. 3b). In general, the best results are still achieved with a deep learning approach. The SF, FW and BF approach show a larger variance, while also having a lower median. The generalized Titarenko's approach shows a similar range of scores like the other methods when using SSIM as metric.



(a) (b)

Figure 3: PSNR (a) and SSIM (b) values for different ring artifact reduction techniques.

B. Real data

Fig. 4 illustrates the results of all methods on real data. In this case, only a visual comparison can be made, as no ground truth is available. All methods are able to reduce ring artifacts to a comparable degree. The thin ring artifacts are largely reduced in all cases. However, there are still some (newly introduced) artifacts present. Most notable is the remaining presence of the thicker ring artifacts for all methods. These types of artifacts are not included in the simulation data and are consequently not reduced by the neural network.

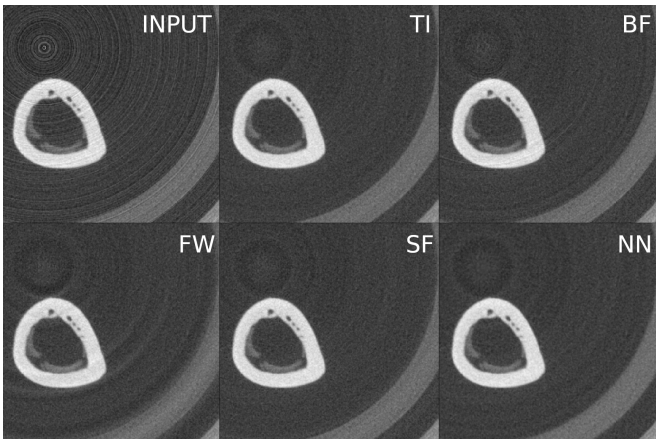


Figure 4: The unprocessed reconstruction (INPUT) as well as five different ring artifact reduction techniques applied to a reconstruction of a real sample with ring artifacts.

.

V. Discussion

In this paper, a U-Net-like deep neural network for sinogram preprocessing was proposed to prevent ring artifacts in the

reconstructed image. On a simulated dataset of 2400 samples, unseen by the neural network, significantly higher PSNR and SSIM values were reached compared to four reference methods. These results indicate that our method was able to reduce ring artifacts to a larger degree.

Due to the lack of a dataset consisting of the same sinograms with and without ring artifacts, data was simulated in this work. Nevertheless, visual assessment of ring artifact reduction on a real sample showed results comparable to other methods that operate in sinogram space. This is an indication that our method is able to generalize beyond simulated data and can also be applied to real data. However, despite our best attempts to accurately model ring artifacts, they inevitably differ from real world samples. The presence of thicker rings in Fig. 4 illustrates this problem. Consequently, caution has to be taken when making interpretations on performance for real data, based on the results for simulated data.

In future work, ring artifact reduction methods operating in image space will be included in the study. Ring artifacts will also be simulated more realistically, for instance by including thicker bands. This will allow us to see whether or not ring artifacts can be reduced to a larger degree.

Acknowledgment

This research received funding from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme as well as from the Research Foundation Flanders (FWO) through the MetroFlex project S004217N. Real data was acquired with TESCAN XRE's UniTOM XL scanner.

References

[1] N. T. Vo, R. C. Atwood, and M. Drakopoulos, "Superior techniques for eliminating ring artifacts in X-ray micro-tomography", *Optics Express*, vol. 26, no. 22, p. 28 396, Oct. 2018.

[2] L. C. P. Croton, G. Ruben, K. S. Morgan, *et al.*, "Ring artifact suppression in X-ray computed tomography using a simple, pixel-wise response correction", *Optics Express*, vol. 27, no. 10, p. 14 231, May 2019.

[3] R. A. Ketcham, "New algorithms for ring artifact removal", in *Developments in X-Ray Tomography V*, U. Bonse, Ed., SPIE, Aug. 2006.

[4] J. Sijbers and A. Postnov, "Reduction of ring artefacts in high resolution micro-CT reconstructions", *Physics in Medicine and Biology*, vol. 49, no. 14, N247–N253, Jul. 2004.

[5] E. M. A. Anas, S. Y. Lee, and M. K. Hasan, "Removal of ring artifacts in CT imaging through detection and correction of stripes in the sinogram", *Physics in Medicine and Biology*, vol. 55, no. 22, pp. 6911–6930, Nov. 2010.

[6] F. Vidal, J. Létang, G. Peix, *et al.*, "Investigation of artefact sources in synchrotron microtomography via virtual X-ray imaging", *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 234, no. 3, pp. 333–348, Jun. 2005.

[7] S. Rashid, S. Y. Lee, and M. K. Hasan, "An improved method for the removal of ring artifacts in high resolution CT imaging", *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, Apr. 2012.

[8] W. Vågberg, J. C. Larsson, and H. M. Hertz, "Removal of ring artifacts in microtomography by characterization of scintillator variations", *Optics Express*, vol. 25, no. 19, p. 23 191, Sep. 2017.

[9] Z. Chao and H.-J. Kim, "Removal of computed tomography ring artifacts via radial basis function artificial neural networks", *Physics in Medicine & Biology*, vol. 64, no. 23, p. 235 015, 2019.

[10] M. Salehjahromi, Q. Wang, Y. Zhang, *et al.*, "A new iterative algorithm for ring artifact reduction in CT using ring total variation", *Medical Physics*, vol. 46, no. 11, pp. 4803–4815, Sep. 2019.

[11] Z. Wang, J. Li, and M. Enoh, "Removing ring artifacts in CBCT images via generative adversarial networks with unidirectional relative total variation loss", *Neural Computing and Applications*, vol. 31, no. 9, pp. 5147–5158, Jan. 2019.

[12] W. Fang, L. Li, and Z. Chen, "Comparison of ring artifacts removal by using neural networks in different domains", Poster presented at the IEEE Nuclear Science Symposium and Medical Imaging Conference, Manchester, UK, 2019.

[13] Z. Chao and H.-J. Kim, "Removal of computed tomography ring artifacts via radial basis function artificial neural networks", *Physics in Medicine and Biology*, vol. 64, no. 23, p. 235 015, Dec. 2019.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 234–241.

[15] H. S. Park, S. M. Lee, H. P. Kim, *et al.*, "CT sinogram-consistency learning for metal-induced beam hardening correction", *Medical Physics*, vol. 45, no. 12, pp. 5376–5384, 2018.

[16] W.-A. Lin, H. Liao, C. Peng, *et al.*, "DuDoNet: Dual Domain Network for CT metal artifact reduction", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 512–10 521.

[17] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines", in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807–814.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014.

[19] Martin Abadi, Ashish Agarwal, Paul Barham, *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015.

[20] K. Clark, B. Vendt, K. Smith, *et al.*, "The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository", *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[21] W. van Aarle, W. J. Palenstijn, J. Cant, *et al.*, "Fast and flexible X-ray tomography using the ASTRA toolbox", *Optics Express*, vol. 24, no. 22, pp. 25 129–25 147, 2016.

[22] F. De Carlo, D. Gürsoy, D. J. Ching, *et al.*, "Tomobank: A tomographic data repository for computational X-ray science", *Measurement Science and Technology*, vol. 29, no. 3, p. 034 004, 2018.

[23] D. Gürsoy, F. De Carlo, X. Xiao, *et al.*, "TomoPy: A framework for the analysis of synchrotron tomographic data", *Journal of Synchrotron Radiation*, vol. 21, no. 5, pp. 1188–1193, 2014.

[24] Z. Wang, A. C. Bovik, H. R. Sheikh, *et al.*, "Image quality assessment: From error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[25] B. Münch, P. Trtik, F. Marone, *et al.*, "Stripe and ring artifact removal with combined wavelet—Fourier filtering", *Optics Express*, vol. 17, no. 10, pp. 8567–8591, 2009.

[26] E. X. Miqueles, J. Rinkel, F. O'Dowd, *et al.*, "Generalized Titarenko's algorithm for ring artefacts reduction", *Journal of Synchrotron Radiation*, vol. 21, no. 6, pp. 1333–1346, 2014.

An OpenCL pipeline implementation on Intel FPGA for 3D backprojection

Daouda Diakite[1], Maxime Martelli[1,2], and Nicolas Gac[1]
[1] Université Paris-Saclay, CNRS, CentraleSupélec, L2S, 91190, Gif-sur-Yvette, France.
[2] Thales DMS, Elancourt, France.

***Abstract*—To accelerate the time-consuming 3D back-projection used by CT reconstruction algorithms, hardware accelerators are needed. Since 2007, GPU is by far the main one. In this article, we propose a pipeline implementation of the 3D back-projection algorithm on a high-end FPGA using Intel FPGA SDK for OpenCL while presenting some optimization metrics for task parallelism. Compared to a non-optimized version on Arria 10, we achieved a speedup of 23 regarding execution time, by applying these techniques correctly. We then compared these results with a low-end FPGA, CPU, and GPU in terms of execution time and energy efficiency.**

***Index Terms*—Algorithm architecture co-design, Intel FPGA SDK for OpenCL, hardware acceleration, FPGA, Computed Tomography.**

I. Introduction

Programmable Gate Arrays (FPGAs) is becoming an attractive technology for High-Performance Computing thanks to increasing floating-point computation power and High-Level Synthesis (HLS) tools becoming more and more mature. These tools allow developers to generate a hardware implementation on FPGAs using software programming languages like C, C++, or OpenCL. One of the promises of those tools is to make FPGA development accessible by software engineers. Like GPUs or multi-core CPUs, FPGAs are often considered and included in heterogeneous systems, to speed up time-consuming applications. One main advantage is also to be more energy-efficient compared to GPUs and CPUs.

Some recent work focused on explorations of HLS tools efficiency includes [1], [2], [3] for Intel OpenCL SDK, or [4], [5] for Xilinx Vivado HLS. In particular, [6] presented an OpenCL implementation that was better than a Verilog/VHDL design in terms of frequency, latency, and resource utilization. The authors of [7] proposed a novel method that separates memory access from computation for OpenCL kernels based on a Low-Level Virtual Machine (LLVM) compiler. The authors of [8] use OpenMP to generate OpenCL code from a higher abstraction level, which is then synthesized by OpenCL SDK for bitstream generation. These works underline the interest for FPGAs in the High-Performance Computing (HPC) field.

In the past years, FPGAs for Computed Tomography (CT) were configured with Hardware Description Language (HDL) [9], [10], [11], [12], where the goal was to model a pipeline (from the algorithm) and perform a voxel update at every clock cycle. The main concern of the HDL implementation is the very long development time, even though the results were significant in terms of execution time. More recently, new approaches based on HLS tools for FPGAs were developed, to obtain better reconstruction times and energy efficiency while accelerating development time. One such work proposed a hardware acceleration based-FPGA of the Maximum Likelihood Expectation Maximization (MLEM) algorithm [4] and achieved a significant acceleration in execution time compared to an optimized CPU multi-core version.

FPGAs are reconfigurable, and this allows programmers to have a better algorithm architecture adequacy, by leveraging data or task parallelism. Furthermore, tools included in the Intel FPGA SDK for OpenCL enable a developer to implement a wide range of optimizations quickly but give less control compared to the traditional HDL approach.

In this paper, we propose a pipeline implementation of the 3D back-projection algorithm based on the Reflex Attila Arria 10 device, using Intel FPGA SDK for OpenCL 18.1. From a first implementation of this CT algorithm on a low-end DE1-SoC board [1], we show in this paper, the benefits of using a higher-end FPGA and explore new optimization possibilities based on task parallelism.

The remainder of this paper is organized as follows: in Section II, we present the back-projection algorithm. Section III introduces some OpenCL metrics relevant to pipeline parallelism, and Section IV details different optimization techniques. The results are discussed in section V, and we conclude this work.

II. 3D computed tomography reconstruction

3D tomography used in medical imaging or non-destructive testing (NDT) aims to acquire the internal structure of 3D objects from external measurements. As illustrated in Fig. 1, an X-ray source and an array of detectors are rotating around the object (3D volume). The radiation emitted from the source is attenuated depending on the object's local density. A sinogram is created by stacking all of the measured attenuations from the detectors acquired at different angles.

The 3D back-projection described in detail in [9] algorithm is given by

$$d(c) = \int S_{CT}(u(\varphi, c), v(\varphi, c), \varphi).w(\varphi, c)^2 d\varphi$$

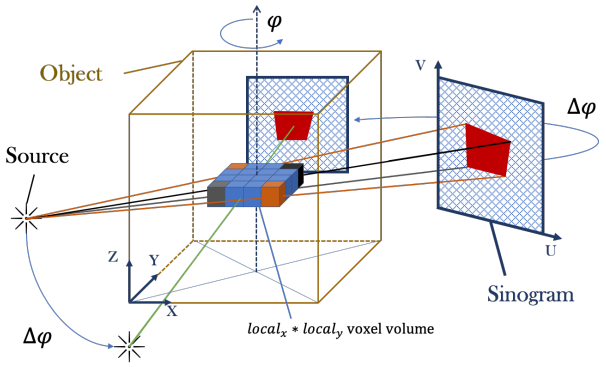


Fig. 1: 3D Computed Tomography

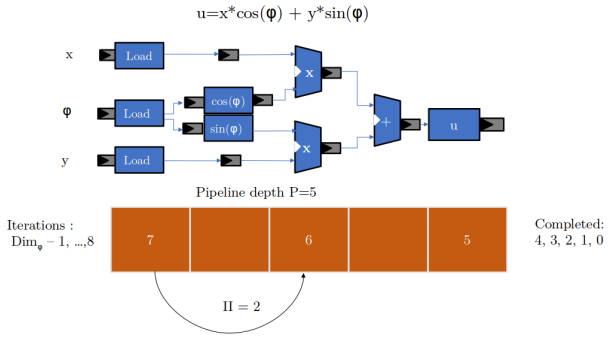


Fig. 2: Pipeline fragmentation of a function on FPGA

Where (u, v) are the cone bean coordinates, φ is the angular trajectory of the detector and w is the distance weight.

$$u(\varphi, c) = x * cos(\varphi) + y * sin(\varphi)$$

$$v(\varphi, c) = x * sin(sin\varphi) - y * cos(\varphi) + z$$

As the distribution of the sensor is discrete, the integral transforms in a sum for all φ values. This algorithm is particularly suited for SIMD cores, because this sum has to be computed for every voxel of the object, and is best executed on massively parallel architectures. However, one way to accelerate our algorithm is to make use of coalesced memory access patterns, also referred to as memory coalescing. Many times, memory objects are retrieved in large blocks and cached in smaller but faster caches.

III. OpenCL metrics for pipeline characterization

For a given program, instead of having to transform the source code into instructions that depend on the target processor, High-level tools will create an architecture made of elementary blocks for each algorithm. Figure 2 represents a computation architecture that corresponds to the calculation of u and v coordinates for the 3D back-projector. All data goes through the pipeline, and we can see that a new iteration starts every two clock cycles with a pipeline depth of 5 due to data dependency. The overall efficiency of this generated hardware is only 50% in theory.

Three main metrics characterize a pipeline. First, the **Initiation Interval (II)** represents the number of clock cycles between the launch of consecutive loop iterations. It means that one loop iteration is launched every II hardware clock cycles. The ideal II value is 1, and the offline compiler attempts to achieve that value for a given loop whenever possible, and it is achieved when dependency is absent or handled in one cycle. In most cases, internal iteration dependencies causes poor II value and it can be fixed with some optimization techniques, like shift register pattern. Secondly, the **Pipeline depth (P)** is the number of cycles required by one data to pass through the entire calculation pipeline, i.e. the number of stages of the pipeline. For a given amount of iterations n, the global cycle number of a pipeline can be computed with those two metrics: $cycles = P + II(n-1)$. By adding a third metric, the **operating frequency (F)**, it is possible to compute the execution time of a pipeline, as given here: $T = \frac{P+II(n-1)}{F}$.

Optimizing an OpenCL implementation is finding a balance between those three metrics.

IV. Pipeline parallelism

OpenCL on FPGA supports data-parallel (NDRange) as well as task parallel (Single Work Item) programming models, and this makes FPGAs useful for a wide range of applications.

NDRange kernel is the GPU-like data-parallel computing model on FPGAs using OpenCL, which aims to perform the same set of instructions on multiple elements of a memory object. Each instance of execution is called work-item, which are grouped in work-group. In [1] we performed an NDRange implementation of 3D back-projection and presented the OpenCL memory architecture as well as the impact of their utilization.

Fig. 3: Back-projection algorithm pipeline architecture

As Intel FPGA SDK for OpenCL allows both data and task (pipeline) parallelism, the choice of a type of parallelism, based on the application, requires adequate knowledge of both the application specificity and the tools. In our case, since FPGA resources do not allow massive data parallelism, we chose the pipeline parallelism approach to implement the 3D back-projection algorithm. Even before HLS tools, FPGAs were well known for their deep and efficient pipeline execution pattern, resulting in good performances and low latency. For instance, in computed tomography, a highly efficient hardware architecture for forward projection in Computed Tomography based on Xilinx Virtex-5 FPGA was proposed in [12] using a floating-point to fixed-point conversion and a two-level

memory, for separable-footprint (SF) forward-projector. Some OpenCL optimization techniques were presented in previous work [1], such as memory architecture and memory pre-fetching, as well as the difference between NDRange kernel and Single Work Item (SWI) kernel. Here, we will present other SWI optimizations based. The 3D back-projector is divided into several pipeline stages as shown in Fig. 3, which correspond to the crucial steps of the algorithm, this function level pipelining and each block can be split at loop level.

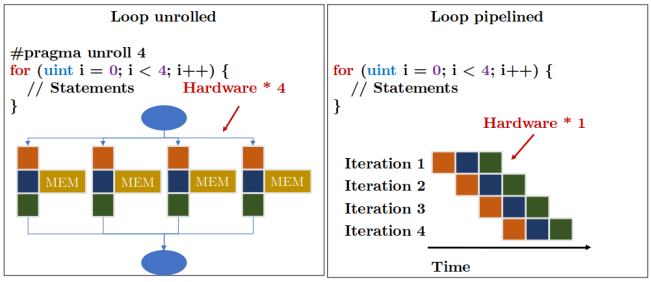


Fig. 4: Loop unroll versus loop pipeline

A. Loop pipelining

In SWI design, the offline compiler automatically tries to pipeline all the loops in the kernel. The idea is to split the body of the loop in a succession of micro-operations that can execute concurrently. The advantage is that the pipelined loop can launch a new iteration at every clock cycle by using the same hardware resources as the traditional sequential one but with better performances.

On FPGAs, the loop pipeline is the best solution in terms of resources and performance ratio. However, an efficient pipeline must not stall which means it has to perform an ideal initiation interval.

In certain rare cases, achieving an II value of 1 might lower the kernel maximum operating frequency, therefore lead to poor performances; for such a loop, it is more convenient to use the ii pragma to fix the II value that performs the best frequency. It is also possible to fix the maximum operating frequency by using the F_{max} attribute or by including it in the aoc command at compile time. However, if the offline compiler cannot meet all the constraint bitstream generation will be impossible.

B. Loop unrolling

Another loop optimization is the unrolling, which consists of replicating, with a configurable factor, the body of the loop. Unlike pipelining, unrolling also replicates hardware resources, which by the way, maximize memory bandwidth utilization i.e maximally unrolling the most nested loop increases the amount of memory access per iteration; this is because the unrolled loop is closest to the computation. A loop can be fully or partially unrolled depending on the available logic element or the size of the cache memory (if applicable), the cache misses will increase if it cannot deliver all the data needed for computation.

V. Results and discussion

We used for this work the FPGA Attila Arria 10 device, which offers 8 GB of DDR4 memory and a maximum FPGA frequency of 405 MHz. The FPGA was in PCIe connection with the host system. The considered volume is a 256^3 voxel, with 256 angles variations. Each kernel execution is monitored through the Intel FPGA dynamic Profiler for OpenCL. For each kernel, this tool provides, amongst other things, the operating frequency, the execution time, the logic utilization, and the latency, bandwidth, and stall of most memory access. The coefficient α and β are stored in the local memory due to their size and their frequent solicitation for each detector coordinates calculation.

In SWI kernel, the loops are pipelined automatically; the challenge is to feed the loop in the most efficient way, which brings us back to the loop unrolling. For a nested loop algorithm, which loop to unroll and which unrolling factor have to be chosen. One way to get to achieve the finest granularity possible is first to unroll the most nested loop and gradually consider the rest of the loops as long as the hardware resources allow. Based on this, in our case, the loop to be unrolled is the φ. This unrolling does not mean that several voxels are updated by clock cycle but allows the accumulation of the sensor contributions to be accelerated according to the projection angles.

Board	HLS	Logic (%)	Freq.(Mhz)	time(s)
Arria 10	SWI φ unroll 1	24	196.97	124.82
Arria 10	SWI φ unroll 16	50	170.45	19.09
Arria 10	SWI φ unroll 32	62	150	5.34
Arria 10	SWI φ unroll 40	80	134.94	68.63
Arria 10	NDRange	86	168.23	28
DE1-Soc	SWI	36	63.6	67.5
DE1-Soc	NDRange	96	140	16.9

TABLE I: Performances comparison between FPGAs Arria 10 (764k LUTs) and DE1-SoC (118k LUTs)

However, our first implementation on Arria, in table I, is the CPU-like version FPGA-friendly implemented to get an efficient pipeline with ideal II value, and the *unroll*1 forces the compiler not to unroll the loop. The offline compiler achieves for this version an operating frequency of 196,97 Mhz with an execution time of 124,82 s. This is the result of under-utilization (only 24% of hardware resources In table I) of available resources on Arria 10 device. Therefore it does not take full advantage of the parallelism potential. Exceeding this frequency by applying the optimizations based on SWI, will be almost impossible as the depth of the pipeline will be variable according to the number of replication.

By replicating 16 times (unroll 16), we double the number of resources and get a 6.5 fold acceleration compared to the first version. We notice that the offline compiler sacrificed F_{max} to achieve the II value of 1. The best execution time was obtained by a replication factor of 32

with 80% logic utilization at $150Mhz$. This version is the most optimal and takes advantage of the automatic cache mechanism. Having the same features of this version with a larger frequency, using F_{max} attribute, was not possible for the offline compiler. It increased the number of constraints at the place and route time.

For this algorithm on Arria 10 device, the maximum possible replication factor is 40. Once greater than 40 we exceed 100% of the device capabilities. This version is sub-optimal not only because of the low frequency, but because the number of iterations is not divisible by 40. There will be an epilogue to handle, which will substantially increase the depth of the pipeline and lead to poor performances. It is imperative, as far as possible, to know the number of iterations of a given loop at compile time to have a bitstream that best fit the algorithm.

Algorithm 1 Kernel back-projector OpenCL

```
for all z_n, y_n, x_n do
  voxel_sum ← 0
  #pragma unroll factor
  for all φ do
    Compute(u_n, v_n)
    voxel_sum+ = sinogram[u_n, v_n, φ]
  end for
  volume[x_n, y_n, z_n] = voxel_sum
end for
```

A. FPGAs Arria 10 vs DE1-SoC

By comparing results shown in Table I with those obtained previously in [1], we observe that DE1-SoC SWI implementations were limited by its logical resources, as the tool was unable to generate an efficient pipeline kernel. Arria 10 implementations deliver better performance than both SWI and NDRange versions on DE1-SoC with 12.6 and 3.1 speedup respectively. It should be noted that interpolation was replaced by the nearest neighbor method on the DE1-SoC and the Arria 10 because it required more resources and computation. Besides, on DE1-SoC device the accumulation of a variable cost six clock cycle instead of one cycle for Arria 10. In terms of efficiency, the DE1-SoC is more efficient than the Arria 10 device when considering the number of logic elements.

B. FPGA, CPU, GPU

The GPUs are far ahead regarding performance (Table II), mostly thanks to their high core count and frequency, which makes them more suited for massive data-parallel computing. However, FPGA implementation provides better execution time than an optimized multi-core CPU version despite its low operating frequency. Overall, GPUs still consume less energy than FPGAs despite requiring more instantaneous power.

Device	Power(W)	Energy(mWh)	Execution time(s)
CPU E5-2667	47	862	66
Titan X Pascal	237	0,92	0.014
Jetson TX2	12,9	0,91	0,253
Arria 10	9,9	14,65	5,34

TABLE II

VI. Conclusions

In this work, we presented a pipeline implementation of a 3D back-projector on an Intel FPGA Arria 10. FPGA achieved better performance than CPU, but it does not perform as well as the optimized versions on GPUs. Considering the specifics of the FPGAs and metrics presented, the offline compiler can generate a pipeline architecture similar, however less effective, to those generated through HDL languages.

In SWI implementation, parallelism is to be extracted in the body of the function, and optimizations must be FPGA-centric to take full advantage of this architecture. The FPGA leading manufacturers recommend this approach, but if the application does not match this type of parallelism, it is a good idea to favor data parallelism. Tomography algorithms are better suited for massive data computing, which explains why GPUs are currently favored in this field.

To stand up to GPUs in tomography, FPGAs need to improve their ability to express data parallelism better, like having adapted memory buses or more DSP.

We would like to thank Daniel Charlet from the LAL (Orsay, France) for granting us access to an Arria 10 board.

[1] M. Martelli *et al.*, "3D Tomography back-projection parallelization on Intel FPGAs using OpenCL," *Journal of Signal Processing Systems*, 2018. https://hal.archives-ouvertes.fr/hal-01831884
[2] A. A. Purkayastha *et al.*, "Exploring the efficiency of OpenCL pipe for hiding memory latency on cloud FPGAs," in *IEEE HPEC*, 2019.
[3] S. Hc *et al.*, "FPGA accelerated PET image reconstruction," *Engineering and Technology*, 2019.
[4] M. Ravi *et al.*, "FPGA as a hardware accelerator for computation intensive mlem medical image reconstruction," *IEEE Access*, 2019.
[5] F. Siddiqui *et al.*, "FPGA-based processor acceleration for image processing applications," *Journal of Imaging*, 2019.
[6] M. A. Mansoori *et al.*, "Efficient FPGA implementation of PCA algorithm for large data using hls," in *PRIME Conference*, 2019.
[7] A. A. Purkayastha *et al.*, "LLVM-based automation of memory decoupling for OpenCL applications on FPGAs," *MICPRO*, 2019.
[8] M. Knaust *et al.*, "OpenMP to FPGA offloading prototype using OpenCL SDK," in *IEEE IPDPS*, 2019.
[9] N. Gac *et al.*, "High Speed 3D Tomography on CPU, GPU, and FPGA," *EURASIP Emb Sys*, 2008.
[10] F. Pfanner *et al.*, "High performance parallel backprojection on FPGA," in *Fully3D*, Potsdam, Germany, Jul. 2011.
[11] M. Leeser *et al.*, "Parallel-beam backprojection: an FPGA implementation optimized for medical imaging," ACM/SIGDA, 2002.
[12] J. K. Kim *et al.*, "Forward-Projection Architecture for Fast Iterative Image Reconstruction in CT," *IEEE Trans Sign Process*, 2012.

PGM based TV L0 regularized linear tomosynthesis reconstruction

Daniel Hadhazi, Gabor Horvath

***Abstract*—Tomosynthesis reconstruction algorithms estimate attenuation coefficients from a set of low-dose X-ray projection images, which are acquired over a limited total scan angle. Due to this limitation, the problem is highly underdetermined, which cause tomographic blur in the reconstructions along the in-depth direction. As the consequence of this artifact, the resolution of these reconstructions is poor, and strong regularization is required in order to increase the quality of the slice images.**

This paper proposes an L0 total variation (L0-TV) regularized algorithm to the reconstruction problem of linear tomosynthesis modality. The underlying optimization problem is solved by Proximal Gradient Descent Method, which can handle such problems efficiently.

The proposed reconstruction method is compared to other, well-known approaches (L2-TV, Thikhonov-TV regularization) based on simulations and theoretical discussion.

***Index Terms*—Linear CT reconstruction, Maximum a posteriori estimation, tomographic blur, total variation L0.**

I. Introduction

LINEAR tomosynthesis [1] estimates the X-ray attenuation coefficients of the examined volume from a set of (typically 40-60) low dose X-ray projection images, which are acquired by a flat-panel detector over a limited total scan angle. This angle usually cannot be more than 80° in practice, therefore high amount of tomographic blur degrades the quality of the reconstruction along the in-depth direction (also referred by z hereafter). The geometric arrangement of the modality in the case of chest imaging is illustrated in Fig. 1.

During the reconstruction, the X-ray attenuation coefficients of the examined volume are estimated by minimizing the cost function:

$$C(\mathbf{I}) = \varphi(\mathbf{I}) + r(\mathbf{I}) \tag{1}$$

where $\mathbf{I}$ is a vector, stacked from the estimated attenuation coefficients, $\varphi(\cdot)$ is the fidelity term, which penalizes the inconsistency between the estimated coefficients and the input projection images. $r(\cdot)$ is the regularizer term, which can be used to complement the measurements with a priori information about the preferred reconstructions. In X-ray tomography reconstruction problems (e.g. in tomosynthesis modality), the fidelity term is usually defined by:

$$\varphi(\mathbf{I}) = \left\| \mathbf{G} - H\{\mathbf{I}\} \right\|_{\mathbf{W}}^2 / 2 \tag{2}$$

here $\mathbf{G}$ is a vector stacked from the preprocessed (logarithmized) projection images, $H\{\cdot\}$ denotes the projection operator, $\mathbf{W}$ is the penalty weight of the inconsistencies (which can be utilized effectively e.g. modelling photon starvation).

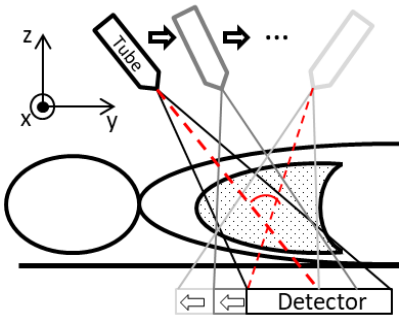


Fig. 1. Geometric arrangement of linear tomosynthesis. The figure illustrates a sagittal cross section. The trajectories of the detector and the X-ray source are parallel to the y axis. During the acquisition they move continuously in the opposite direction in limited range, therefore the total scan angle, which is marked by the red arc, is limited.

The utilization of regularization ($r(\cdot)$ in (1)) is necessary in linear tomosynthesis, because the acquired projection images ($\mathbf{G}$) do not contain information about large part of the spectrum of the sagittal slices. Therefore, the whole reconstruction problem is underdetermined (and also inconsistent because of the noise). This is the direct consequence of the limited total scan angle. The regularizer term should be defined in order to complement these measurements without inducing high bias towards non-acceptable reconstructions. The typical choice of the regularization term is the sum of power of the norm of the gradient vectors of the attenuation coefficients. In section II, the most frequently used regularization terms are examined and compared theoretically. In section III, a possible implementation of the TV-L0 regularized reconstruction problem is described, while in section IV, the empirical results are described, the conclusions

Manuscript received Januar 19, 2020, revised April 30, 2020. This work was supported by the Hungarian National Scientific Research Fund (OTKA 119866), BME-Biotechnology FIKP grant of EMMI (BME FIKP-BIO), by the National Research, Development and Innovation Fund (TUDFO/51757/2019-ITM and by the Nvidia Gpu Grant. *(Corresponding author: Daniel Hadhazi)*.

D. Hadhazi is with the Department of Measurement and Information Systems, Budapest University of Technology and Economics, Budapest, H-1117, Hungary (e-mail: hadhazi@mit.bme.hu).

G. Horvath, deceased, was with the Department of Measurement and Information Systems, Budapest University of Technology and Economics, Budapest, H-1117, Hungary.

and the discussions are drawn in section V.

II. Regularization in linear tomosynthesis

If only the interactions of the primary photons are modelled (according to the Beer-Lambert's law), the projection images can be approximated by parallel beam projections [2] of the pseudo sagittal slices of the examined volume:

$$\mathbf{S}^{(x)}_{(z,y)} = \mathbf{I}^{(z)}_{(x',y')} \tag{3}$$

where $\mathbf{I}^{(z)}_{(x,y)}$ is the (x,y)-th voxel of the z-th coronal slice of the examined volume. The other two indexes are defined by:

$$[x',y'] = [x,y]\cdot(z-d)/d \tag{4}$$

where d is the distance between the X-ray source and the detector. In a typical implementation of a linear tomosynthesis system (such the one illustrated in Fig. 1.), the thickness of the examined volume is much smaller than the distance between the detector and the X-ray source, while the distance between the detector and the examined volume is small (usually there is only a desk between them). This is the reason, why the maximal deviation of the angle of the projection rays of $H\{\cdot\}$ is small (not more than 1.7° in our geometric arrangement).

Based on this approximation and the Fourier Slice theorem, the input projections ($\mathbf{G}$ in (2)) contain information about only less than third of the spectrum of the pseudo sagittal slices. This can be also noted from the Modulation Transfer Function (MTF) of the $H^{\dagger}\{H\{\cdot\}\}$, which is shown in Fig. 2. Therefore, edges perpendicular to the projecting rays are blurry in the reconstructions.

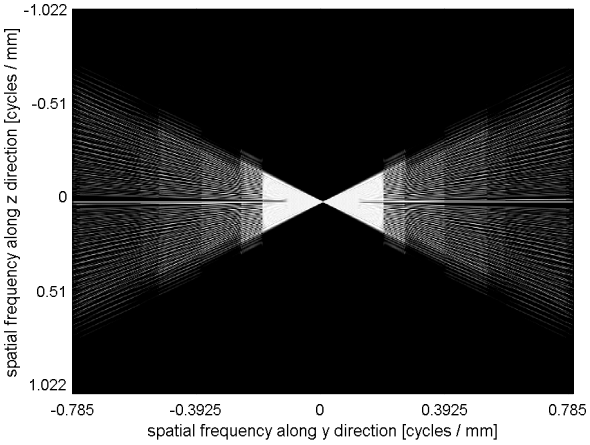


Fig. 2. MTF between the pseudo sagittal slices and their reconstructions.

In order to complete the measurements, utilization of strong regularization is required. In this section, we compare the most frequently utilized, gradient vector based regularizations of the reconstruction problem. Therefore, $r(\cdot)$ in (1) is concretized:

$$r(\mathbf{I}) = p(\nabla\mathbf{I}) \tag{5}$$

In order to analyze the posterior of different regularizations, the proximity operator [3] of $p(\cdot)$ is examined:

$$prox_{\lambda p}(\mathbf{b}) = \arg\min_{\mathbf{a}}\left\{(1/2)\|\mathbf{a}-\mathbf{b}\|_2^2 + \lambda p(\mathbf{a})\right\} \tag{6}$$

The motivation of this analysis is based on the proximal gradient method (PGM) [3], in which the regularization takes effect by the proximity operator in every iteration.

A. Minimizing the energy of the gradient vectors

Tikhonov-TV regularization penalizes the energy of the gradient vectors:

$$p(\nabla\mathbf{I}) = \sum_{x,y,z}\left\|\nabla\mathbf{I}_{(x,y,z)}\right\|_2^2 \tag{7}$$

The proximity operator of $p(\cdot)$ is:

$$prox_{\lambda p}(\mathbf{b}) = \mathbf{b}/(1+2\lambda) \tag{8}$$

Therefore, this regularization reduces the amplitudes of gradients multiplicatively, which induces that it directly degrades the edges along the contours of the objects.

B. Minimizing the L2 norm of the gradient vectors (L2-TV):

This regularization is usually referred by TV [4]. In this case, the amplitude of the gradients are penalized:

$$p(\nabla\mathbf{I}) = \|\mathbf{I}\|_{TV} = \sum_{x,y,z}\left\|\nabla\mathbf{I}_{(x,y,z)}\right\|_2 \tag{9}$$

$$prox_{\lambda p}(\mathbf{b}) = \max\left(\|\mathbf{b}\|_2 - \lambda, 0\right)\cdot\mathbf{b}/\|\mathbf{b}\|_2 \tag{10}$$

The corresponding proximity operator is the so-called shrinkage / soft threshold, which is applied in most of the TV minimization methods. This regularization additively reduces the amplitude of the gradient vectors, therefore it reduces the contrast of the reconstructed objects. The anisotropic TV (ATV) method underweights the penalty of difference of attenuation coefficients along the z axis [5]. As consequence, it tends to generate false edges perpendicular to the z axis, while it additively reduces the contrast along the x and y axes.

C. Minimizing the number of non-zero gradients (L0-TV)

Similar to the L2-TV and ATV, this regularization is based on sparsity assumption of the gradient vectors and the theory of Compressed Sensing (CS). The corresponding penalty term and proximity functions are:

$$p(\nabla\mathbf{I}) = \sum_{x,y,z}\operatorname{sgn}\left(\left\|\nabla\mathbf{I}_{(x,y,z)}\right\|_2\right) \tag{11}$$

$$prox_{\lambda p}(\mathbf{b}) = \begin{cases}\mathbf{b} & \|\mathbf{b}\|_2 > \sqrt{2\lambda}\\ \mathbf{0} & \|\mathbf{b}\|_2 \le \sqrt{2\lambda}\end{cases} \tag{12}$$

The L0 based regularization preserves objects, which have large contrasts, while it eliminates small changes in the estimated attenuation coefficients. Therefore, utilizing this regularization tends to produce reconstructions with sparse gradients, which is desirable in the sense of assuming piecewise homogeny attenuation coefficients. This is the reason, why this kind of regularization is suggested. Please note that (1) in this case is highly non-convex, and have many local extrema. Therefore, the direct utilization of this regularizer is impractical and makes the reconstruction highly sensitive to its initialization. This is the reason, why we propose the utilization of *iteratively re-weighted L1 (IRWL1)* [6] instead, which approximates the penalty function by:

$$p(\nabla\mathbf{I}) = \sum_{x,y,z}\log\left(\left\|\nabla\mathbf{I}_{(x,y,z)}\right\|_2 + \varepsilon\right) \tag{13}$$

In the implementation of the corresponding optimization, this regularizer term is linearly majorized in every iteration:

$$p^{(t)}(\nabla \mathbf{I}) = \sum_{x,y,z} \nabla \mathbf{I}_{(x,y,z)} \Big/ \left(\left\| \nabla \mathbf{I}^{(t-1)}_{(x,y,z)} \right\|_2 + \varepsilon \right) \tag{14}$$

$\nabla \mathbf{I}^{(t-1)}_{(x,y,z)}$ is the gradient vector of the estimated attenuation coefficients at the end of the $(t-1)$-th iteration, while $p^{(t)}(\cdot)$ is the regularizer penalty term at the t-th iteration. ε is a hyper-parameter. It controls the behavior of the IRWL1 based regularization between the L2-TV and the pure L0-TV ones.

III. Proposed Algorithm

Based on the previous discussion, we suggest the utilization of IRWL1-TV regularizer to the reconstruction problem of linear tomosynthesis. We propose the utilization of Proximal Gradient Descent Method (PGM) [3] to minimize (1). The corresponding method is described by Algorithm 1.

Algorithm 1 PGM based reconstruction

Initialize $\mathbf{I}^{(0)} = \mathbf{0}$
for $j=0,\ldots, N_j$

$$\mathbf{T}^{(j)} = \mathbf{I}^{(j)} - \alpha \left(H^* \left\{ diag(\mathbf{W}) \cdot H \left\{ \mathbf{I}^{(j)} \right\} \right\} - H^* \left\{ diag(\mathbf{W}) \cdot \mathbf{G} \right\} \right)$$

$\mathbf{R}^{(0)} = \mathbf{1}$
for k=0,…, N_k

$$\mathbf{L}^{(k)} = \arg\min_{\mathbf{L}} \left\{ (1/2) \left\| \mathbf{L} - \mathbf{T}^{(j)} \right\|_2^2 + \alpha \cdot \lambda \cdot \|\mathbf{L}\|_{TV,\mathbf{R}^{(k)}} \right\}$$

$$\mathbf{R}^{(k+1)}_{(x,y,z)} = 1 \Big/ \left(\left\| \nabla \mathbf{L}^{(k)}_{(x,y,z)} \right\|_2 + \varepsilon \right)$$

end for
$\mathbf{I}^{(j+1)} = \mathbf{L}^{(N_k)}$
end for

$H^*\{\cdot\}$ is the operation of backprojection (conjugate of $H\{\cdot\}$), $\mathbf{W}$ is the pixel-wise penalty weight of the fidelity term (2). The weighted TV norm is defined by:

$$\|\mathbf{L}\|_{TV,\mathbf{R}} = \sum_{(x,y,z)} \mathbf{R}_{(x,y,z)} \cdot \left\| \nabla \mathbf{L}_{(x,y,z)} \right\|_2 \tag{15}$$

The optimization problem of $\mathbf{L}^{(k)}$ is solved by 6 iterations of monotone FISTA algorithm [7]. α is the Lipschitz constant of (2). The tunable parameters of the algorithm are N_j, N_k, λ, ε and the number of mFISTA iterations (marked by N_f hereafter). The role of λ parameter is to control the balance between the fidelity and the regularizer terms. The value of ε controls the sparsity of the gradient of the reconstruction. As its value decreases, the estimated volume converges to the results of the L0-TV regularized ones, which is sparser but also much more instable. While as the value of ε increases, the reconstruction converges to the results of L2-TV regularized ones (which is a convex optimization problem). The values of the iteration numbers control the dynamics of the algorithm. Based on our examinations, the method is moderately sensitive to these values apart from extremity. The value of N_f controls the trade-off between the required number of floating point operations and the stability of the regularization. As the value of N_k increases, the gradient vector field of the resulting reconstruction becomes sparser, but the computational time also increases. The empirically suggested values of these hyper-parameters are: $\lambda = 0.002$, $\varepsilon = 0.01$, $N_f = 6$, $N_k = 5$, $N_j = 1000$.

We also examined line search methods to tune the value of α adaptively during the iterations, but it increased the reconstruction error if the number of floating point operations was not changed. Additionally, it can be efficiently estimated based on the reformulation of the projection operator (described by (3)). Different re-initialization strategies of $\mathbf{R}$ were also examined. We found that if it is not reinitialized in different iterations over j, then the reconstruction converges faster, but to a significantly worse convergence point (the resulting reconstructions were blurred because of the initialization of $\mathbf{I}^{(0)}$). The effect of different initializations (e.g. SART reconstruction, FBP or other, ML or MAP estimation based methods are under examination).

IV. Empirical Results

In this section, we examine the results of the proposed PGM based method (referred by L0-TV in this section) and compare it to the results of Tikhonov-TV regularization and L2-TV regularization. For this purpose, we examine reconstructions calculated from simulated noiseless projections of high-resolution chest CT scans of the LIDC-IDRI dataset (no. 0066, 0140, 0438, 0441) [8]. This dataset contains reconstructions of human patients, with axial slice thickness not more than 0.6 mm. Because of the proposed method is based on MAP estimation, examining the bias introduced by the regularization is critical. In our simulations, the distance between the X-ray source and the flat panel detector is 1000 mm, the total scan angle is 75° and the size of the simulated flat panel detector is 430 mm × 430 mm, which acquires 50 images in 512×512 pixel. The pivot beam in every projection contains a fix, focus point, which is located at $x = 0$, $y = 0$, $z = 175\,mm$.

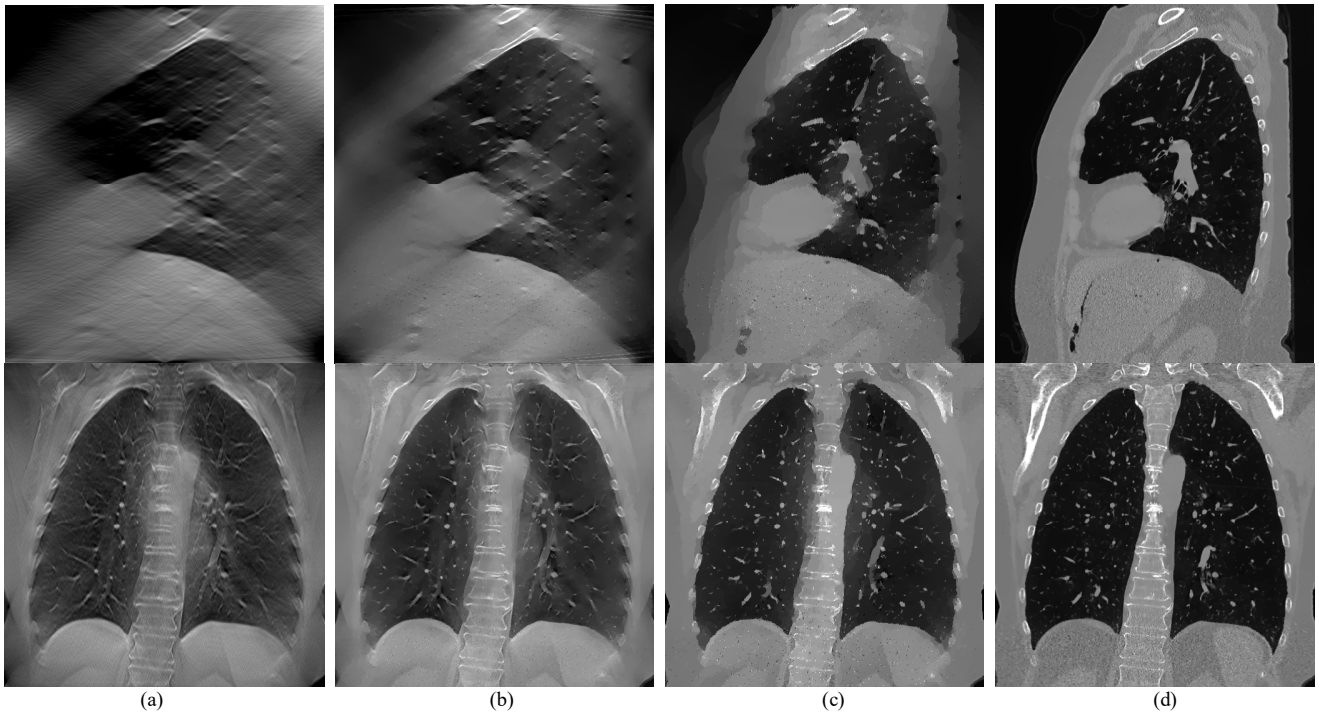
In order to examine the reconstruction quantitatively, we used the RMSE (root mean squared error) and the MAE (mean absolute error) measures:

$$\begin{aligned} RMSE &= \sqrt{\|\mathbf{I} - \mathbf{I}_r\|_2^2 / P} \\ MAE &= \|\mathbf{I} - \mathbf{I}_r\|_2^2 / P \end{aligned} \tag{16}$$

P denotes the number of voxels, $\mathbf{I}_r$ is the reconstructed volume and $\mathbf{I}$ is the ground truth reconstruction. The CT scans of the dataset are degraded by high amount of reconstruction noise, which makes the quantitative comparison pointless. In order to overcome this problem, the CT scans are preprocessed – first, the reconstructions are resampled (by cubic interpolation) to isotropic resolution, what is followed by a 3D isotropic total variation (TV) regularized deburring:

$$\mathbf{I}' = \arg\min_{\mathbf{K} \in \mathbb{R}_+^N} \left\{ (1/2) \|\mathbf{K} - \mathbf{I}\|_2^2 + \mu \|\mathbf{K}\|_{TV} \right\} \tag{17}$$

μ is the weight of the regularization. The 3D version of the FISTA was utilized to solve this optimization problem. For the simulations, $\mu = 0.015$ was applied. The parameters of the compared methods are tuned empirically by minimizing the mean MAE in the dataset. In the case of Tikhonov-TV regularization $\lambda = 0.015$, while for L2-TV $\lambda = 0.005$ was



(a) (b) (c) (d)

Fig. 3. Reconstructed slices of linear tomosynthesis reconstruction of LIDC 0066 scan: (a) Tikhonov regularized, (b) L2-TV regularized, (c) L0-TV regularized and (d) ground truth slices. The same sagittal slices are in the upper and the same coronal slices are in the bottom row.

chosen in this sense. In order to make a correct comparison, all three methods are compared after 700 TFlop of computation.

Fig. 3. shows sagittal and coronal slices of reconstructions calculated from the simulated projections of LIDC 0066 scan: Tikhonov regularized (a), the L2-TV regularized (b) and the proposed, PGM based L0-TV regularized (c) one, while the ground truth in column (d). From the results, it can be noted that the L0-TV can effectively reduce the blur along the in-depth direction, however, it tends to generate impulse noise (especially in large, homogenous but noisy regions, e.g. in the abdominal one). As it is also visible, the proposed reconstruction does not generate artificial objects in the reconstructed slice images. Based on our experiments, the proposed L0-TV regularized reconstruction preserves every real object, what is reconstructed by either of the other two regularization methods. The reduction of the in-depth blur is more spectacular in the sagittal cross sections and causes thinner coronal slices (as a consequence, the amount of out of plane structures is reduced in the slices of the L0-TV regularized reconstructions).
The mean reconstruction errors are summarized in Table I. Based on these results, the proposed, PGM based L0-TV method produces the lowest reconstruction error. The relative reduction of the MAE is bigger than the RMSE.

Table I: Mean reconstruction errors (in Hounsfield Unit)

Algorithm	*mean MAE*	*mean RMSE*
Tikhonov-TV	198.5	257.3
L2-TV	181.3	239.7
L0-TV	150.5	215.7

V. Conclusion and Discussion

This paper proposes an L0-TV regularized, PGM based algorithm to the reconstruction problem of linear tomosynthesis. We also examined the theoretical background of the problem and compared different approaches to handle the lack of information in the projection images (which is the consequence of the limited total scan angle). Based on this comparison, the utilization of the proposed, PGM based L0-TV regularized reconstruction algorithm is suggested for limited angle reconstruction problems (e.g. linear tomosynthesis).

References

[1] J. T. Dobbins III, “Tomosynthesis imaging: At a translational crossroads,” Medical Physics, vol. 36, no. 6Part1, pp. 1956–1967, May 2009.

[2] D. Hadházi, G. Horváth, “Partial volume artifact in MITS reconstructed digital chest tomosynthesis,” Proceedings of the 23rd PhD mini-symposium, pp. 14-18, 2016.

[3] N. Parikh, “Proximal Algorithms,” Foundations and Trends in Optimization, vol. 1, no. 3, pp. 127–239, 2014.

[4] T. Goldstein and S. Osher, “The Split Bregman Method for L1-Regularized Problems,” SIAM Journal on Imaging Sciences, vol. 2, no. 2, pp. 323–343, Jan. 2009.

[5] Z.-F. Pang, Y.-M. Zhou, T. Wu, and D.-J. Li, “Image denoising via a new anisotropic total-variation-based model,” Signal Processing: Image Communication, vol. 74, pp. 140–152, May 2019.

[6] E. J. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing Sparsity by Reweighted ℓ 1 Minimization,” Journal of Fourier Analysis and Applications, vol. 14, no. 5–6, pp. 877–905, Oct. 2008.

[7] A. Beck and M. Teboulle, “Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems,” IEEE Transactions on Image Processing, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[8] S. G. Armato III et al., “The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans,” Medical Physics, vol. 38, no. 2, pp. 915–931, Jan. 2011.

Iterative Reconstruction-Based Increase of Operating Pitch in Aviation Security CT

Michael H. Schmitt

***Abstract*— Worldwide, many CT are in service helping to screen passenger bags for explosive threats. The data acquisition systems of those systems were typically designed to provide minimally sufficient sampling for a specific FBP reconstruction algorithm. However, airports must continually scale up their baggage screening capabilities to keep up with an ever increasing flow of passengers. One possible approach is to increase belt speeds through existing systems, which increases the maximum theoretical throughput but potentially degrades image quality. We demonstrate that those existing systems can indeed be run beyond their design pitch by augmenting their existing FBP reconstruction with a second pass iterative reconstruction to control the resulting artifacts, achieving comparable performance to what would be expected if the CT had been originally designed for higher pitch. Using this approach we have attained, in at least one instance, the same level of threat detection certification on an explosive detection system (EDS) operating at a pitch 50% higher than it was designed for, requiring little more than a computing and software upgrade to achieve.**

***Index Terms*—iterative reconstruction, cone-beam tomography, high pitch CT, limited angle tomography, filtered-back projection, aviation security CT, explosive detection systems**

I. Introduction

Over the last 20+ years, thousands of computed tomography (CT) systems have been built and deployed across the world in service of screening aviation passenger bags for automated detection of threats such as dangerous explosives. In this work, we take a critical eye to these fleets – representing billions of dollars worth of capital investment in total. We show that these systems can, without significant modification, be driven to achieve substantially higher throughputs than they were originally designed for, while retaining the full detection capability required by security regulators.

Whereas a handful of modern systems were born of iterative reconstruction [1], the majority of systems actually deployed at the time of this writing – by and large CT of the third generation – were designed with the sampling requirements of a particular filtered back projection (FBP) algorithm in mind. Typical constraints may include PI-line completeness [2][3]; fully sampling the Katsevich envelope [4][5][6], or Tam-Danielsson [7][8], or Tuy-Smith [9][10]; or otherwise ensuring, for example, a minimum of "180 degrees of ray coverage" over some plane or another containing the region of interest, such as might be found in any of [11][12][13][14][15] as a very few examples.

For newly designed systems, recent advancements in iterative reconstruction techniques suggest that many of the aforementioned requirements may not be as strong as once thought. For example, a constraint – perhaps including even PI-line completeness – may be relaxed for a subset of the slices; such as described in [16] for example. However, as already noted, many aviation security CT scanners have already been built and deployed. Practically speaking, those systems were designed to operate at a single, fixed pitch, consistent with a common concept of operations for all screened bags, and with detector arrays designed to provide the minimally sufficient samples necessary for reconstruction.

Meanwhile, air traffic continues to rise; and the demands placed on airports and responsible agencies such as the TSA for screening that increased traffic rises with it. Given the need to accommodate this ever increasing flow of passengers, combined with other practical constraints – such as limited floorspace and finite budgets – there is significant value to *in situ* upgrades that neither consume additional footprint nor require lengthy operations disruptions for upgrade or replacement. We therefore seek a method to increase the pitch at which existing systems may be operated, without significant modifications and without loss of detection-relevant image quality.

II. Methods

Assuming that a given system design provides minimal support for a chosen FBP algorithm at some design pitch, running at a faster belt speed results in a form of limited angle tomography. For example, consider the illustration of a hypothetical system design in Fig. 1.

For the general class of limited angle tomography problems, iterative reconstruction techniques are known to be useful [17][18]; but despite dramatic advances in computational speed over the last decade (in particular, the emergence of inexpensive GPU cards as a viable tool), it is still largely impractical to entirely employ iterative methods for the full real-time reconstruction problem such as described here. In security CT, the system is continuously reconstructing at feed rates as high as 0.5 m/s. For automated detection, the image must be produced with minimal latency in a period of time

M. Schmitt is with Leidos Security Detection & Automation, St. Petersburg, FL 33701 USA (e-mail: michael.schmitt@leidos.com).

comparable to the length of time it takes to cross the scan plane – usually no more than a few seconds – with a goal to render a decision before the bag exits the system.

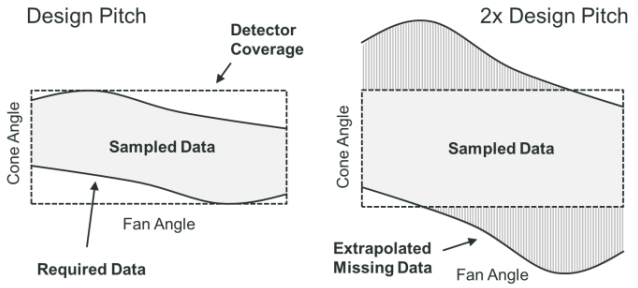


Fig. 1. A CT with support for a given pitch (left), driven to higher pitch results in areas of missing data which must be extrapolated (right).

As already stated, we are motivated to make minimal changes relative to the original image. Although the original FBP-based reconstruction algorithm – operating in the limited angle regime – lacks support; it may still be used. To do this, we have the option to treat the limited angle data as truncated or (preferably) to extrapolate the projection using any appropriate method. The resulting artifacts are likely to be similar in some respects to those encountered in "long objects" of inexact helical CT [19] or "large patients" exceeding the allowable ROI [20]. For pitch factors up to about 50% above the designed operating point, we may expect the resulting image, despite flaws, to provide an adequate initialization for an appropriate iterative reconstruction algorithm, which can then provide further refinement (as shown in this work, this expectation is indeed met). We may also apply regularization with the iterative reconstruction; for example, total variation (TV) minimization [21] to improve the image quality and improve convergence.

Generally, the direct (FBP) method is the one chosen in the original design and the iterative method is any suitable – for example, SIRT [22] may be a reasonable choice.

III. Iterative Process

The input image – made inexact at least by the limited angles resulting from being operated beyond its design pitch – is used to seed an iterative reconstruction; with the iterative loop being designed primarily to reduce artifacts. As compared to notable work such as [23], the artifacts which are being corrected are due entirely to operating the CT at a pitch beyond that which it was designed for (for example, lacking even PI-line completeness).

A simplified process diagram is shown in Fig. 2. Here, a buffer is used wherein slices from the FBP reconstruction are fed into one end of the buffer and corrected slices are output from the other end, resulting in a fixed slice processing pipeline. As the limited angle FBP slices are received, a set of projections from the helix – centered on a slice dragging behind the leading edge of the image by a fixed distance determined by the number of slices involved in any simultaneous update – are selected and utilized in an iterative correction loop. Whereas extrapolations are used for the FBP step, only well-measured samples are included in the iterative reconstruction. Already corrected slices which fall outside the update region are "complete" and are output from the reconstruction process.

Experimentally, we find that this approach – combined with clamping, limiting the gain, and rolling off the update as a function of distance in the slice domain from the center of the update – provides the best results. Several iterative updates, even utilizing less than all of the full available projection data, is found to be sufficient to eliminate the bulk of the limited angle artifacts; while further iteration may drive the characteristics of the image further away from the baseline.

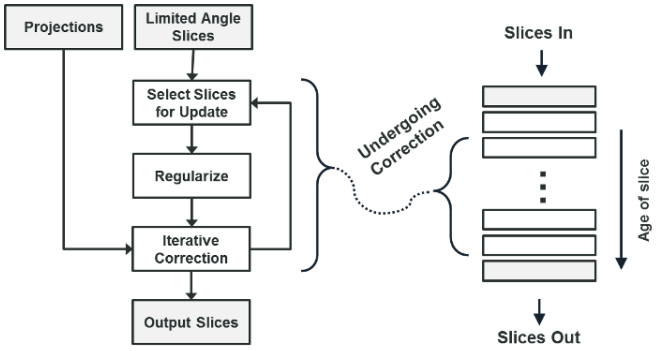


Fig. 2. Simplified iterative correction process diagram (left) and corresponding illustration of the buffer (right). A feature of the buffer is a fixed latency between the input slice, the range of slices being updated, and slices available for output.

III. Results

Simulated results are provided in Fig. 3. Here, we show a slice near one end of a 3D Shepp-Logan phantom reconstructed with an exact FBP algorithm both at a nominal design pitch and at 1.5x the design pitch, along with the corresponding image after iterative correction. A slice near the end of the phantom is chosen because extrapolation errors affecting the FBP reconstruction are significant due to the geometric shape of the phantom in this region.

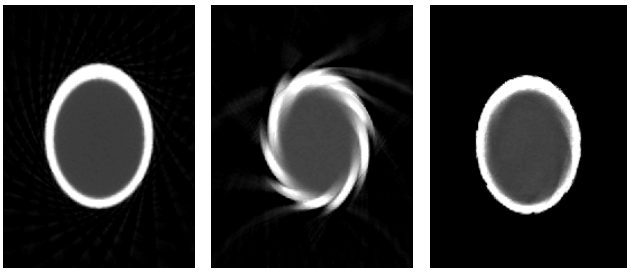


Fig. 3. Simulated slice reconstructed at design pitch (left), 1.5x design pitch (center), and at 1.5x design pitch with iterative corrections applied (right).

This approach has been reduced to practice and at least one EDS system has achieved certification status at the same detection level at both its design pitch and 1.5x its design pitch. We may quantitatively evaluate the image quality of each mode, for example using the ANSI N42.45 procedures [24]; and while we cannot disclose absolute performance numbers, we are able to present differences between the two modes. In Table I, we present a few select results from a test

conducted at design pitch and one conducted at 50% higher than the design pitch. Note that, for the compared resolution-based quantities, while the ANSI N42.45 procedure doesn't directly output resolution it does generate estimates of point-spread functions (PSF) and modulation transfer functions (MTF), from which it is straightforward to compute a full-width half-maximum (FWHM) of PSF. A spatially averaged resolution can be further estimated by converting the respective point-spreads in 3D into Gaussian equivalents and summing in quadrature.

A comparison of bag images using methods described herein are provided in Fig. 4.

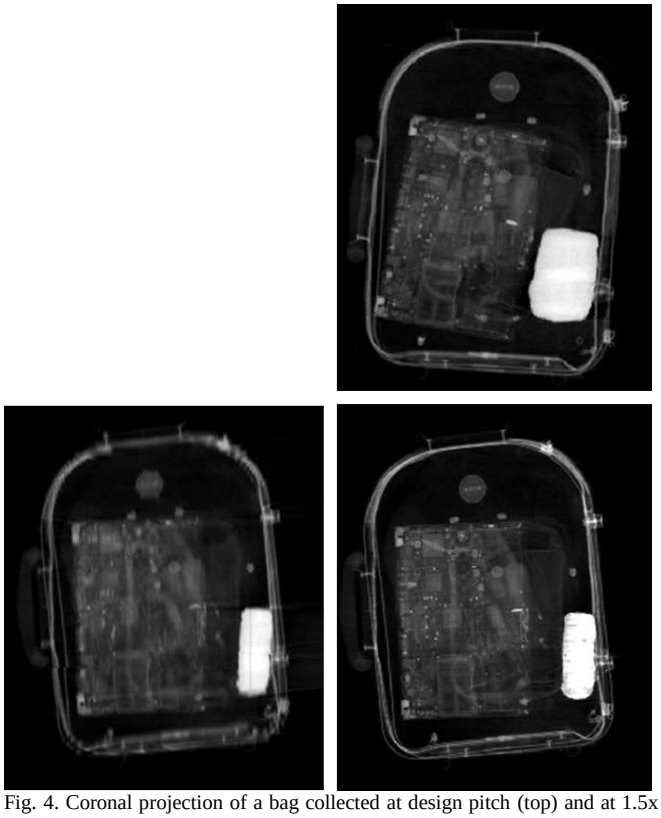


Fig. 4. Coronal projection of a bag collected at design pitch (top) and at 1.5x design pitch (bottom row) before (left) and after iterative correction (right). Images may have been altered to obscure capabilities.

IV. Discussion

As Fig. 3., Fig. 4., and Table I show, it is possible to make the image collected at a pitch exceeding the design values by 50% roughly equivalent to the original, at least within the context of the demands of security screening, in at least one embodiment. Experimentally, we find that the practical upper limit may be around 1.66x of the design pitch. On the example embodiment, it was found that pitch artifacts could be rapidly reduced up to about 1.5x of pitch, but required more iteration to eliminate as the pitch increased. Above 1.66x, no practical amount of iteration would eliminate pitch artifacts, so additional methods outside of the scope of this report may be required. It is presently unknown how well this finding would generalize to other CT as it is likely to be at least partially dependent on the original CT design and the choices taken along the way; for example, how much of the detector data is already well utilized by the original reconstruction algorithm [25][26].

TABLE I
Select Quantitative Comparisons

Measurement	$(V_{HP} - V_{DP}) / V_{DP}$
Resolution, transverse (FWHM of PSF†)	+9.7%
Resolution, along slice (FWHM of SSP†)	-13.5%
Resolution, full 3D (spatially averaged†)	0.0%
Streaks; std. along a line (Acetal, control)	+11.0%
Streaks; std. along a line (Acetal, near pins)	-21.4%
Uniformity, mean of CT (Acetal)	+0.6%
Uniformity, std. of CT (Acetal)	+5.4%
Object Length Accuracy	0.0%

Sample quantitative image quality test values obtained using ANSI N42.45 procedures on the high pitch configuration (V_{HP}), computed relative to values collected on the baseline configuration operating at design pitch (V_{DP}).

†Resolution is not specifically calculated as output of ANSI test but may be derived from them, as indicated in the body

With regards to the quantitative results summarized in Table I, in our work – to the extent possible, consistent with our goals – the iterative reconstruction parameters were tuned to minimize overall differences across all the various quantifiable measurements. However, this tuning process was also done with important external constraints, such as limits on the maximum slice latency, the maximum amount of computational power allotted, and limited development time.

As may be expected, we find that resolution and CT noise show the largest disagreements. Changes in reconstruction behavior around streak-inducing metal pins were also evident. However, we simultaneously found that many such differences were offsetting (e.g., the full 3D spatially averaged resolution – clearly an important imaging parameter as it relates to randomly oriented objects within bags – is compatible). Furthermore with modifications – for example, to the regularization – it is possible to reverse the sign of any given comparison. It seems likely that with further work it would be possible to reduce all quantifiable differences between the two modes to essentially zero; however as mentioned previously, a Leidos scanner has already demonstrated regulatory certification to the same standard of automated detection using this imaging technique. In that instance, we found that the imaging achieved was close enough to the existing baseline that we were able to adapt existing threat detection algorithms, designed for the "low" (design) pitch version of the scanner, to achieve new certification for the high pitch version without undue effort.

V. Conclusion

Herein we have reported on our work [27] increasing the operating pitch of security CT using iterative reconstruction techniques as a secondary correction process to an underlying FBP image reconstruction. We have shown that it is possible to increase the pitch by 50% or more while maintaining roughly consistent image quality; and in fact have reduced this to practice by certifying an EDS at a pitch 50% higher than its design. This approach has potential application, minimally, to the thousands of other aviation security CT already built.

EXPORT STATEMENT

This technical data is released to the public as "published" information in accordance with Export Administration Regulations (EAR) Part 734.7(a)(1).

REFERENCES

[1] A. Foland, "Design of the Discrete Skew Geometry and Iterative Reconstruction of the MV3D Scanner", *Proc. CT-Meeting*, pp. 301-304, 2012

[2] Y. Ye, J. Zhu, and G. Wang, "Minimum detection windows, PI-line existence and uniqueness for helical cone-beam scanning of variable pitch", *Med. Phys.*, vol. 31, no. 3, pp. 566-572, 2004

[3] Y. Zou, X. Pan, D. Xia, and G. Wang, "PI-line-based image reconstruction in helical cone-beam computed tomography with a variable pitch", *Med. Phys.*, vol. 32, no. 8, pp. 2639-2648, 2005

[4] A. Katsevich, "Analysis of an exact inversion algorithm for spiral cone-beam CT", *Phys. Med. Biol.*, vol. 47, pp. 2583-2597, 2002

[5] A. Katsevich, "An improved exact filtered backprojection algorithm for spiral computed tomography", *Adv. Appl. Math.*, vol. 32., pp. 2012-2026, 2002

[6] A. Katsevich, S. Basu, and J. Hsieh, "Exact filtered backprojection reconstruction for dynamic pitch helical cone beam computed tomography", *Phys. Med. Biol.*, vol. 49, no. 14, p. 3089, 2004

[7] K. Tam, S. Samarasekera, and F. Sauer, "Exact cone-beam CT with a spiral scan", *Phys. Med. Biol.*, vol. 43, pp. 1015-1024, 1998

[8] P-E. Danielsson, P. Edholm, J. Eriksson, and S. Magnusson, "Towards exact reconstruction for helical cone-beam scanning of long objects", *Proc. Fully 3D*, pp. 141-144, 1997

[9] H. Tuy, "An Inversion Formula for Cone-Beam Reconstruction", *SIAM J. on Applied Mathematics*, vol. 43, no. 3, pp. 546-552, 1983

[10] B. Smith, "Image reconstruction from cone-beam projections: necessary and sufficient conditions and reconstruction methods", *IEEE Trans Med Imaging*, vol. 4., no. 1., pp. 14-25, 1985

[11] S. Orlov, "Theory of three-dimensional reconstruction. Conditions for a complete set of projections", *Sov. Phys. Crys.*, vol. 20, pp. 312-314, 1975

[12] R. Denton, B. Friedlander, and A. Rockmore, "Direct three-dimensional image reconstruction from divergent rays", *IEEE Trans. Nucl. Sci.*, vol. 26, pp. 4695-4703, 1979

[13] K. Smith, "Reconstruction formulas in computed tomography", *Proc. Symp. Appl. Math.*, vol. 27, pp. 7-23, 1982

[14] L. Feldkamp, L. Davis, and J. Kress, "Practical cone-beam algorithm", *J. Opt. Soc. Am.*, vol. 1.6, pp. 612-619, 1984

[15] M. Kachelrieß, S. Schaller, and W. Kalender, "Advanced single-slice rebinning in cone-beam spiral CT", *Med. Phys.*, vol. 27, no. 4, pp. 754-766, 2000

[16] H. Xue, L. Zhang, Z. Chen, and X. Jin, "An iterative reconstruction method for high-pitch helical luggage CT", *Proc. SPIE 8506, Dev. X-Ray Tom. VIII*, p. 8506, 2012

[17] E. Sidky, C. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT", *J. XRay Sci. Tech.*, vol. 14, pp. 119-139, 2006

[18] Z. Chen, X. Jin, L. Li, and G. Wang, "A limited-angle CT reconstruction method based on anisotropic TV minimization", *Phys. Med. Biol.*, vol. 58, no. 7, pp. 2119-2141, 2013

[19] M. Magnusson, P-E. Danielsson, and J. Sunnegärdh, "Handling of long objects in iterative improvement of nonexact reconstruction in helical cone-beam CT", *IEEE Tran. Med. Imag.*, vol. 25, no. 7, 2006

[20] J. Hsieh, E. Chao, J. Thibault, B. Grekowicz, A. Horst, S. McOlash, and T. Myers, "A novel reconstruction algorithm to extend the CT scan field-of-view", *Med. Phys.*, vol. 31, p. 2385, 2004

[21] T. Chan and S. Esedoglu, "Aspects of total variation regularized L1 function approximation", *SIAM J. Appl. Math.*, vol. 65, no. 5, pp. 1817-1837, 2005

[22] J. Gregor, "Algorithmic Improvement to SIRT with Application to X-Ray CT of Luggage", *Proc. CT-Meeting*, pp. 121-124, 2014

[23] J. Sunnegärdh, P-E. Danielsson, and M. Magnusson Seger, "Iterative Improvement of Non-Exact Reconstruction in Cone-Beam CT", *Proc. Fully 3D*, pp. 279-284, 2005

[24] American National Standard Institute, "American National Standard for Evaluating the Image Quality of X-ray Computed Tomography (CT) Security-Screening Systems," ANSI N42.45-2011, 2011

[25] M. Ellenbogen and R. Biijani, "Folded Array CT Baggage Scanner", US Patent 2006/0274879, 2006

[26] M. Kachelrieß, "Interesting detector shapes for third generation CT", *Med. Phys.*, vol. 40., no. 3, p. 31101, 2013

[27] M. Schmitt, A. Foland, and B. Oreper, "Systems and Methods for Image Reconstruction at High Computed Tomography Pitch", US Patent 10119923, 2018

Reconstruction of Voxels with Position- and Angle-Dependent Weightings

Lina Felsner, Tobias Würfl, Christopher Syben, Philipp Roser, Alexander Preuhs, Andreas Maier, Christian Riess

***Abstract*—The reconstruction problem of voxels with individual weightings can be modeled a position- and angle- dependent function in the forward-projection. This changes the system matrix and prohibits to use standard filtered backprojection.**

In this work we first formulate this reconstruction problem in terms of a system matrix and weighting part. We compute the pseudoinverse and show that the solution is rank-deficient and hence very ill posed. This is a fundamental limitation for reconstruction. We then derive an iterative solution and experimentally show its superiority to any closed-form solution.

I. Introduction

THE theory of X-ray computed tomography is well established [1], [2]. The reconstruction problem is oftentimes expressed in matrix notation as a linear system of equations

$$\mathbf{p} = \mathbf{A}\mathbf{x} \ , \tag{1}$$

where $\mathbf{p}$ are the projections, $\mathbf{x}$ the unknown voxel content, and $\mathbf{A}$ the system matrix. Each element $a_{i,j}$ of the system matrix $\mathbf{A}$ describes the contribution of a particular voxel to a particular projection. $\mathbf{A}$ is typically full-rank, and tall (i.e., it has more rows than columns), which allows to solve Eqn. 1 via the left-hand sided pseudoinverse $\mathbf{A}^{\ddagger}$ of $\mathbf{A}$ (cf. Tab. II). This leads to the solution

$$\mathbf{x} = \left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top\mathbf{p} \ , \tag{2}$$

which is commonly known as backprojection filtering where $\left(\mathbf{A}^\top\mathbf{A}\right)^{-1}$ represents the filter and $\mathbf{A}^\top$ the backprojection.

One basic assumption of the reconstruction problem is that the unknown quantity $\mathbf{x}$ is constant under all projections. However, there exist setups with more complicated signal formation, such that this assumption does not hold. One example is the sensitivity of the phase shift in a Talbot-Lau grating interferometer [3], [4]. The measured diffraction is a linear ramp that depends on the distance between the voxel and two gratings that bound the measurement space. Hence, parts of a large object that are closer to the source contribute a much stronger signal than parts that are closer to the detector. This ramp has to be considered as a position-dependent function in single projections [3], [4] and as a position- and angle-dependent function in tomography [5]. Another example is the X-ray dark-field signal [6], [7]. Here, each voxel is measured an angle-dependent linear combination of isotropic and anisotropic scattering, which can also be represented as an angle-dependent weighting in the reconstruction [8].

Such voxel characteristics can also be formulated as a component of the system itself. Since the system matrix already describes a position- and angle-dependent mapping, the integration of such specific weights in the forward-projector via multiplication is straightforward.

However, the modification of the system matrix prohibits the use of standard reconstruction algorithms, since the filter-kernel and adjoint operator are not known. Therefore, we propose to decompose the weighted system matrix into a pure system matrix and a weighting matrix. This allows use the known filter and backprojection with additional operations defined by the weights.

In this paper, we formulate this reconstruction problem and show in Sec. II why the solution to the decomposed system is ill-conditioned. We carry out a simple experiment to explicitly calculate the different possible pseudoinverses. This shows that an exact solution cannot be obtained unless the weighted system matrix is small enough to directly calculate the pseudoinverse. In a second step, we derive in Sec. III the iterative update formula and show that the solution can be iteratively approximated. We report experiments to compare the iterative scheme and its L1-weighted variant to backprojection filtering and an approximate solution. Finally, we conclude our findings and give a brief outlook in Sec. IV.

II. Calculation of the Pseudoinverse: Origin and Impact of its Ill-conditioning

We investigate the inclusion of a position- and angle-dependent function into the reconstruction problem. Inclusion of the weights in the forward-projector can be formulated as

$$p(t,\theta) = \int f(x,y)\cdot g(x,y,\theta)\ \mathrm{d}\mathbf{r} \ , \tag{3}$$

where $p(t,\theta)$ is the projection at detector position t under rotation θ along the projection ray $\mathbf{r}$. The object is given as function $f(x,y)$. The function $g(x,y,\theta)$ is a position- and angle-dependent multiplicative factor. In matrix notation, the equivalent formulation is

$$\mathbf{p} = \tilde{\mathbf{A}}\mathbf{x} \ , \tag{4}$$

where $\tilde{\mathbf{A}}$ denotes the system matrix including the weights given by $g(x,y,\theta)$. According to Eqn. 2, the solution can be calculated with the pseudoinverse $\tilde{\mathbf{A}}^{+}$. The standard system matrix $\mathbf{A}$ allows to derive the adjoint operator and filter function with the Fourier-Slice theorem. However, an analogous

All authors are with the Pattern Recognition Lab, University of Erlangen-Nuremberg, Germany.
{lina.felsner, christian.riess}@fau.de
L. Felsner, A. Maier are with the International Max Planck Research School - Physics of Light

interpretation for $\left(\tilde{\mathbf{A}}^\top\tilde{\mathbf{A}}\right)^{-1}$ and $\tilde{\mathbf{A}}^\top$ is not available and hence unknown.

Nevertheless, the weighted reconstruction problem can be analytically examined by deconstructing $\tilde{\mathbf{A}} = \mathbf{BW}$ into a pure system matrix $\mathbf{B}$ and a weight matrix $\mathbf{W}$. The weights $\mathbf{W}$ shall be individually chosen per voxel x_i and per rotation angle θ_j. Hence, $\mathbf{W}$ consists for each angle θ_j of a diagonal submatrix $\mathbf{W}_{\theta_j} = \text{diag}\left(w_{j,1}, w_{j,2}, \ldots, w_{j,V}\right)$. The full matrix $\mathbf{W}$ is given on the right side of Eqn. 5. The angle-dependency in $\mathbf{W}$ requires to construct the system matrix $\mathbf{B}$ slightly different from traditional system matrices $\mathbf{A}$: $\mathbf{B}$ is a block-diagonal matrix, where each block contains the traditional system matrix under one specific angle θ_j. Hence,

$$\mathbf{B} = \begin{bmatrix} \mathbf{A}_{\theta_1} & & & \\ & \mathbf{A}_{\theta_2} & & \\ & & \ddots & \\ & & & \mathbf{A}_{\theta_\Omega} \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} \mathbf{W}_{\theta_1} \\ \mathbf{W}_{\theta_2} \\ \vdots \\ \mathbf{W}_{\theta_\Omega} \end{bmatrix} , \tag{5}$$

and $\tilde{\mathbf{A}} = \mathbf{BW}$ as stated above. $\mathbf{W}$ has full rank due to its design from diagonal matrices. $\mathbf{B}$ has full rank due to its block-diagonal structure and because every row within one block models an individual ray with individual voxel contributions. The full reconstruction equation is

$$\mathbf{p} = \mathbf{BWx} \ . \tag{6}$$

As in traditional reconstruction, solving Eqn. 6 for $\mathbf{x}$ requires to multiply the pseudoinverse from the left. To this end, two approaches could be considered, namely either

$$\mathbf{W}^+\mathbf{B}^+\mathbf{p} = \mathbf{W}^+\mathbf{B}^+\mathbf{BWx} \tag{7}$$

or

$$(\mathbf{WB})^+\mathbf{p} = (\mathbf{BW})^+(\mathbf{BW})\mathbf{x} \ . \tag{8}$$

However, it is important to realize that for the dimensions of $\mathbf{B}$ and $\mathbf{W}$, the individual inversion of these matrices in Eqn. 7 does not isolate $\mathbf{x}$, and can hence not be used. This follows from the fact that $\mathbf{B}$ is a broad matrix with dimensions $N\Omega \times V\Omega$ with $N < V$, since there are by definition more voxels V than detector pixels N. In this case, the right-hand sided Moore-Penrose pseudoinverse $\mathbf{B}^\dagger$ has to be used (see Tab. II). However, left-multiplying the right-hand side pseudoinverse $\mathbf{B}^\dagger$ leads to the term $\mathbf{B}^\top\left(\mathbf{BB}^\top\right)^{-1}\mathbf{B} \neq \mathbf{I}$ and hence does not isolate $\mathbf{x}$ as intended. As a consequence, it is necessary to calculate the general pseudoinverse from Eqn. 8. Note that this insight also implies that

$$(\mathbf{BW})^+ \neq \mathbf{W}^+\mathbf{B}^+ \ . \tag{9}$$

Lemma. The generalized pseudoinverse of the matrices $\mathbf{B}$ and $\mathbf{W}$

$$(\mathbf{BW})^+ = \left(\mathbf{B}^+\mathbf{BW}\right)^+\left(\mathbf{BWW}^+\right)^+ \ . \tag{10}$$

does not have full rank. Thus, we have more degrees of freedom than linearly independent equations, which makes the system of equations ill-conditioned.

Table I: List of symbols

Symbol	Dimension	Interpretation
V		Number of voxels
Ω		Number of projection angles
θ_i		One projection angle ($\forall i \in \{1, ..., \Omega\}$)
N		Number of detector pixels
$\mathbf{x}$	$\mathbb{R}^{V\times 1}$	Vector representation of the object
$\mathbf{p}$	$\mathbb{R}^{N\Omega\times 1}$	Vector containing the projections
$\mathbf{A}$	$\mathbb{R}^{N\Omega\times V}$	System matrix
$\tilde{\mathbf{A}}$	$\mathbb{R}^{N\Omega\times V}$	System matrix with weights
$\mathbf{B}$	$\mathbb{R}^{N\Omega\times V\Omega}$	Block matrix where each block contains the system matrix for one angle
$\mathbf{W}$	$\mathbb{R}^{V\Omega\times V}$	Weight matrix
$\cdot_i$		Element i of vector $\cdot$
$\cdot_{i,j}$		Element i, j of matrix $\cdot$
$\cdot^+$		Pseudoinverse

Table II: Moore-Penrose pseudoinverses for a matrix $\mathbf{A}$ with full rank [9]

Dimension	Name	Rank	Pseudoinverse
$n \times n$	Square	n	$\mathbf{A}^+ = \mathbf{A}^{-1}$
$n \times m$	Broad	n	$\mathbf{A}^+ = \mathbf{A}^\dagger = \mathbf{A}^\top\left(\mathbf{AA}^\top\right)^{-1}$
$n \times m$	Tall	m	$\mathbf{A}^+ = \mathbf{A}^\ddagger = \left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top$

Proof. We show that the rows in $(\mathbf{BW})^+$ are linearly dependent, from which the remaining statements follow. The pseudoinverse of the matrix product $(\mathbf{BW})^+$ is defined as [9]

$$(\mathbf{BW})^+ = \left(\mathbf{B}^+\mathbf{BW}\right)^+\left(\mathbf{BWW}^+\right)^+ \tag{11}$$

$$= \left(\mathbf{B}^\dagger\mathbf{BW}\right)^+\left(\mathbf{BWW}^\ddagger\right)^+ \tag{12}$$

$$= \left(\mathbf{B}^\top\left(\mathbf{BB}^\top\right)^{-1}\mathbf{BW}\right)^+ \cdot \left(\mathbf{BW}\left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\right)^+, \tag{13}$$

where we inserted between Eqn. 11 to 13 the respective left- and right-hand sided Moore-Penrose pseudoinverses.

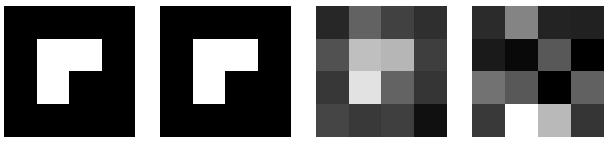
The linear dependency occurs in the product $\mathbf{WW}^\ddagger = \mathbf{W}\left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top$. In order to see this, we first note that $\left(\mathbf{W}^\top\mathbf{W}\right)^{-1}$ is a diagonal matrix due to the diagonal structure of the submatrices of $\mathbf{W}$.

$$\left(\mathbf{W}^\top\mathbf{W}\right)^{-1} = \left(\begin{bmatrix} \text{diag}\left(w_{1,1}, \ldots, w_{1,V}\right) \\ \vdots \\ \text{diag}\left(w_{\Omega,1}, \ldots, w_{\Omega,V}\right) \end{bmatrix}^\top \begin{bmatrix} \text{diag}\left(w_{1,1}, \ldots, w_{1,V}\right) \\ \vdots \\ \text{diag}\left(w_{\Omega,1}, \ldots, w_{\Omega,V}\right) \end{bmatrix}\right)^{-1} \tag{14}$$

$$= \text{diag}\left(1/\sum_i w_{i,1}^2, 1/\sum_i w_{i,2}^2, \ldots, 1/\sum_i w_{i,V}^2\right) \tag{15}$$

Due to the commutativity of diagonal matrices we have the equality $\mathbf{W}\left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top = \left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{WW}^\top$. Here,

$$\mathbf{WW}^\top = \begin{bmatrix} \mathbf{W}_{\theta_1}\mathbf{W}_{\theta_1} & \mathbf{W}_{\theta_1}\mathbf{W}_{\theta_2} & \ldots & \mathbf{W}_{\theta_1}\mathbf{W}_{\theta_\Omega} \\ \mathbf{W}_{\theta_2}\mathbf{W}_{\theta_1} & \mathbf{W}_{\theta_2}\mathbf{W}_{\theta_2} & \ldots & \mathbf{W}_{\theta_2}\mathbf{W}_{\theta_\Omega} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{\theta_\Omega}\mathbf{W}_{\theta_1} & \mathbf{W}_{\theta_\Omega}\mathbf{W}_{\theta_2} & \ldots & \mathbf{W}_{\theta_\Omega}\mathbf{W}_{\theta_\Omega} \end{bmatrix} . \tag{16}$$



(a) GT $\mathbf{x}$ (b) $\tilde{\mathbf{A}}^{+}\mathbf{p}$ (c) $\mathbf{W}^{+}\mathbf{B}^{+}\mathbf{p}$ (d) $(\mathbf{BW})^{+}\,\mathbf{p}$

Figure 1: Results of the pseudoinverse. Fig. a, b are windowed between $0-1$ [a.u.]. Fig. c is windowed between $0-0.5$ [a.u.]. Fig. d is windowed between $-2.5-12$ [a.u.].

Each block in this matrix is again a diagonal matrix $\mathbf{W}_{\theta_i}\mathbf{W}_{\theta_j} = \text{diag}\left(w_{i,1}w_{j,1}, w_{i,2}w_{j,2}, \ldots, w_{i,V}w_{j,V}\right) \forall i,j \in \{1,\ldots\Omega\}$. The rows and columns between the block matrices are linear dependent. Rows i and j only differ by a factor w_{ν_i}/w_{ν_j}, and the same holds for the columns. Thus, the matrix $\mathbf{W}\mathbf{W}^{\top} \in \mathbb{R}^{V\cdot\Omega\times V\cdot\Omega}$ has only rank V, which makes the system of equations highly under-determined and ill-conditioned. □

Experiment. We illustrate these findings with an example that is sufficiently small to also calculate the pseudoinverse of the matrices directly. We use a 4×4 volume, with a triangular structure of value 1 in the center as our object (Fig. 1a). The system matrix uses 4 projection angels with a total of 22 projections. We then calculate the solutions for three different pseudoinverses. This includes first the theoretically correct pseudoinverse of $\tilde{\mathbf{A}}$, which can be directly calculated due to the small problem size. Second, the theoretically wrong solution $\mathbf{W}^{+}\mathbf{B}^{+}\mathbf{p}$, where $\text{rank}(\mathbf{W}^{+}) = V$ and $\text{rank}(\mathbf{B}^{+}) = N\Omega$. Third, the theoretically correct, but ill-conditioned $(\mathbf{BW})^{+}\,\mathbf{p}$.

The results are shown in Fig. 1b – 1d, respectively. The pseudoinverse of the weighted system matrix $\tilde{\mathbf{A}}$ delivers an almost perfect result (Fig. 1b). The L2–distance to the ground truth $\mathbf{x}$ is $1.68E-14$. By using the theoretically incorrect pseudoinverse $\mathbf{W}^{+}\mathbf{B}^{+}$ the overall object shape is recovered, but the values are not correct (Fig. 1c), with a L2–distance to the ground truth of 1.15. The estimation of the theoretically correct but ill-conditioned pseudoinverse completely fails to represent the object (Fig. 1d), with a L2–distance to the ground truth of 16.92. It is surprising that the mathematically incorrect solution still provides a rough estimate of the original object, while the mathematically correct, yet ill-conditioned, solution completely fails to recover the object.

III. Iterative Reconstruction

The first part shows that inclusion of individual weights prevents a closed-form reconstruction. However, the original linear systems of equations in Eqn. 6 has full rank, which suggests that an iterative solution can still be found. To this end, we derive a gradient-descent algorithm and perform a proof-of-concept experiment where we compare the proposed method with and without additional TV regularization to backprojection filtering and a solution with the theoretically incorrect inversion.

Since the Moore-Penrose pseudoinverse estimates the optimal solution in a least-squares sense, we formulate the reconstruction as an iterative least-squares reconstruction task with the objective function

$$\min \frac{1}{2}||\mathbf{BWx} - \mathbf{p}||_2^2 \ . \tag{17}$$

Equation 17 can be iteratively solved via gradient descent. The update-rule for the iterative algorithm is derived below following the Kaczmarz method [10].

One projection in the decomposed system from Eqn. 6 is given as

$$p_i = \mathbf{b}_i\mathbf{Wx} \ , \tag{18}$$

which can be interpreted as a line in the V-dimensional space. The orthogonal projection of $\mathbf{x}$ on this line is given as

$$\mathbf{x}' = \mathbf{x} + \lambda\left(\mathbf{b}_i\mathbf{W}\right)^{\top} \tag{19}$$

$$= \mathbf{x} + \lambda\mathbf{W}^{\top}\mathbf{b}_i^{\top} \ . \tag{20}$$

Since the projected point $\mathbf{x}'$ is a point on the line, Eqn. 18 must also hold

$$p_i = \mathbf{b}_i\mathbf{W}\mathbf{x}' \ . \tag{21}$$

Inserting Eqn. 20 into Eqn. 21 leads to

$$p_i = \mathbf{b}_i\mathbf{W}\left(\mathbf{x} + \lambda\mathbf{W}^{\top}\mathbf{b}_i^{\top}\right) \ . \tag{22}$$

Isolating the parameter λ in Eqn. 22 leads to

$$\lambda = \frac{(p_i - \mathbf{b}_i\mathbf{Wx})}{\mathbf{b}_i\mathbf{W}\mathbf{W}^{\top}\mathbf{b}_i^{\top}} \ . \tag{23}$$

This can be inserted back into Eq. 21, leading to the final update rule

$$\mathbf{x}' = \mathbf{x} + \frac{p_i - \mathbf{b}_i\mathbf{Wx}}{\mathbf{b}_i\mathbf{W}\mathbf{W}^{\top}\mathbf{b}_i^{\top}}\mathbf{W}^{\top}\mathbf{b}_i^{\top} \ , \tag{24}$$

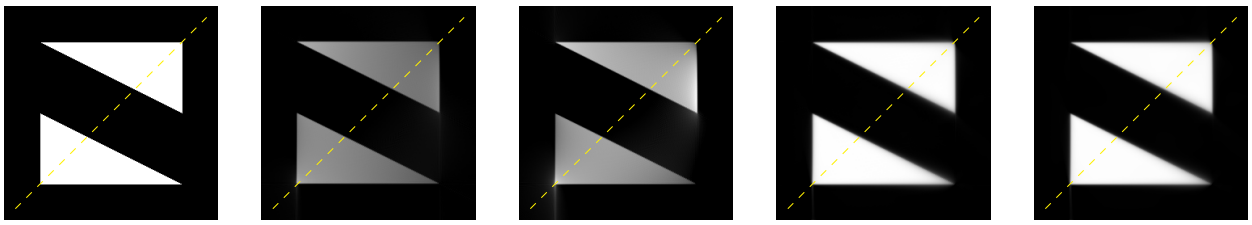
where $\mathbf{b}_i\mathbf{Wx}$ is the weighted forward-projection, $p_i - \mathbf{b}_i\mathbf{Wx}$ is the gradient, and $\mathbf{b}_i\mathbf{W}\mathbf{W}^{\top}\mathbf{b}_i^{\top}$ is a normalization factor that contains the sum of squared weights along the ray.

Experiment. In this experiment we use a more complex sample to investigate the results of the iterative reconstruction.

The weighting function is the sensitivity ramp as it occurs in X-ray Talbot-Lau grating interferometers [3], [4], [5]). The function is a linear ramp between a minimal and maximal sensitivity value, which rotates along with the source and detector. We use an experimental setting that is similar as in Chabior *et al.* [5], with a phantom that consists of two wedges in a parallel-beam geometry. We use a 256×256 voxel volume with 360 projections over an angular range of π.

We reconstruct the object with various approaches. First, a backprojection and then filtering from Eqn. 2 that ignores the weights ("BPF"). Second, the incorrect pseudoinverse $\mathbf{W}^{+}\mathbf{B}^{+}$ from Eqn. 7 ("Wrong PsInv"). Third, the iterative update rule derived in this section ("Iterative"), and fourth the iterative update rule with additional TV regularization ("Iterative + TV"). The mathematical representation of the algorithms is listed in Tab. III. For the iterative reconstruction, a step size of 0.5 is used. An adaptive step size is used for the TV regularization. Both iterative algorithms use 100 iterations.

Figure 2 shows the ground truth and the reconstruction results. The line plots along the yellow dashed line are shown



(a) GT (b) BPF (c) Wrong PsInv (d) Iterative (e) Iterative + TV

Figure 2: Results of the reconstructions. The line plots along the yellow dashed lines are shown in Fig. 3. Fig. c is windowed between $0 - 0.01$ [a.u.]. All other figures are windowed between $0 - 1$ [a.u.].

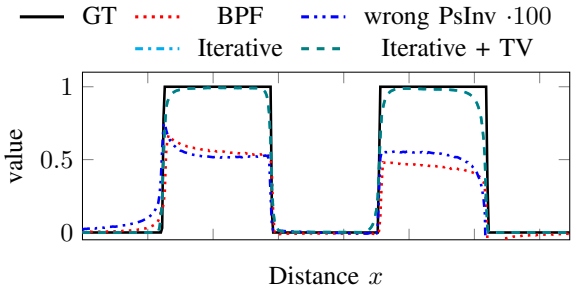


Figure 3: Line plots along the yellow dashed lines in the reconstructions in Fig. 2.

in Fig. 3. The root mean square error (RMSE) and structural similarity (SSIM) are reported in the right columns of Tab. III. Severe artifacts can be observed in the reconstruction with the backprojection and filtering that ignores the weighting function. It exhibits a global offset in the reconstructed values, and the object is not homogeneous. The reconstruction with the incorrect pseudoinverse shows similar artifacts. Also here, the scaling of the reconstructed values is only half of the actual values. The RMSE is twice as large as for backprojection and filtering, and the SSIM is decreased by several orders of magnitude. The iterative reconstruction is visually and quantitatively very good. The TV regularization changes the quantitative result only at the fifth decimal position. Visually, both reconstructions are identical with a high similarity to the ground truth.

This only very small performance difference between iterative reconstruction and its regularized version can be attributed to two reasons. The experiment is neither under-determined, nor is any noise included in this experiment. We expect a larger improvement of TV over the pure iterative algorithm in the presence of this cases.

IV. Conclusion and Outlook

We investigated the inclusion of a per-voxel position- and angle-dependent weighting function in tomographic reconstruction. First, we describe the problem and decompose it into a system matrix and weighting matrix. Calculating the pseudoinverse shows in a rank analysis that the solution is very ill-conditioned. In a second part, we derive an iterative solution and experimentally show its superiority to any possible algebraic solution. We conclude that it is not possible to obtain an analytic solution for the reconstruction of voxels with general position- and angle-dependent weightings, while experimental results for iterative reconstruction well approximate the ground truth solution.

In future work, it will be interesting to investigate special cases of weight functions, and in particular whether for these cases filters can be derived analytically or learned with a Neural-Network [11].

Table III: RMSE and SSIM of the results in Fig. 2

Method	**Equation**	**RMSE**	**SSIM**
BPF	$\mathbf{A}^{\ddagger}\mathbf{p}$	0.24	0.61
Wrong PsInv	$\mathbf{W}^{\ddagger}\mathbf{B}^{\dagger}\mathbf{p}$	0.47	$1.26E-4$
Iterative	$\min \frac{1}{2}\|\|\mathbf{BWx}-\mathbf{p}\|\|_2^2$	0.06	0.99
Iterative + TV	$\min \frac{1}{2}\|\|\mathbf{BWx}-\mathbf{p}\|\|_2^2 + \lambda \mathrm{TV}$.	0.06	0.99

References

[1] F. Natterer, "Computerized tomography," in *The Mathematics of Computerized Tomography*. Springer, 1986, pp. 1–8.

[2] G. L. Zeng, *Medical image reconstruction: a conceptual tutorial.* Springer, 2010.

[3] M. Engelhardt, J. Baumann, M. Schuster, C. Kottler, F. Pfeiffer, O. Bunk, and C. David, "High-resolution differential phase contrast imaging using a magnifying projection geometry with a microfocus x-ray source," *Applied physics letters*, vol. 90, no. 22, p. 224101, 2007.

[4] T. Donath, M. Chabior, F. Pfeiffer, O. Bunk, E. Reznikova, J. Mohr, E. Hempel, S. Popescu, M. Hoheisel, M. Schuster *et al.*, "Inverse geometry for grating-based x-ray phase-contrast imaging," *Journal of Applied Physics*, vol. 106, no. 5, p. 054703, 2009.

[5] M. Chabior, M. Schuster, C. Schroer, and F. Pfeiffer, "Grating-based phase-contrast computed tomography of thick samples," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 693, pp. 138–142, 2012.

[6] T. H. Jensen, M. Bech, I. Zanette, T. Weitkamp, C. David, H. Deyhle, S. Rutishauser, E. Reznikova, J. Mohr, R. Feidenhansl, and F. Pfeiffer, "Directional X-ray dark-field imaging of strongly ordered systems," *Physical Review B*, vol. 82, no. 21, p. 214103, 2010.

[7] V. Revol, C. Kottler, R. Kaufmann, A. Neels, and A. Dommann, "Orientation-selective X-ray dark field imaging of ordered systems," *Journal of Applied Physics*, vol. 112, no. 11, p. 114903, Dec. 2012.

[8] F. L. Bayer, S. Hu, A. Maier, T. Weber, G. Anton, T. Michel, and C. P. Riess, "Reconstruction of scalar and vectorial components in x-ray dark-field tomography," *Proceedings of the National Academy of Sciences*, vol. 111, no. 35, pp. 12 699–12 704, 2014.

[9] K. B. Petersen and M. S. Pedersen, "The Matrix Cookbook," p. 22, 2012.

[10] S. Kaczmarz, "Angenäherte auflösung von systemen linearer gleichungen," *Bull. Int. Acad. Pol. Sci. Lett. A*, vol. 6–8A, pp. 355–357, 1937.

[11] C. Syben, B. Stimpel, J. Lommen, T. Würfl, A. Dörfler, and A. Maier, "Deriving neural network architectures using precision learning: parallel-to-fan beam conversion," in *German Conference on Pattern Recognition*. Springer, 2018, pp. 503–517.

Investigation of Base Materials for Accurate Dual-Energy Computed Tomography. Comparison between the Alvarez-Macovski method and DIRA

Maria Magnusson, Gudrun Alm Carlsson, Michael Sandborg, Åsa Carlsson Tedgren and Alexandr Malusek

***Abstract*—The choice of the material base to which the material decomposition is performed in dual-energy computed tomography may affect the quality of reconstructed images. The aim of this work is to investigate how the commonly used bases (water, bone), (water, iodine) and (photoelectric effect, Compton scattering) affect the reconstructed linear attenuation coefficient in case of the Alvarez-Macovski method. The performance of this method is also compared to the performance of the dual-energy iterative reconstruction algorithm DIRA. The results show that the Alvarez-Macovski method produced artefacts when iodine was present in the phantom together with human tissues, since this method can only work with one doublet. It was shown that these artefacts could be avoided with DIRA using a (water, lipid, protein) triplet for tissues and the (water, iodine) doublet for the iodine solution.**

I. Introduction

THE Alvarez-Macovski method (AM) [1] produces two images of the base material weight coefficients by material decomposition of dual-energy CT data in the projection domain to one base material doublet. From these coefficients, monoenergetic images at any energy can be computed. This way beam-hardening can be eliminated.

The Dual-energy Iterative Reconstruction Algorithm (DIRA) [2] is an algorithm that performs the decomposition in the image domain. Calculated forward projections together with the original projections are used to update the image. This way, beam-hardening can be iteratively removed. It is possible to use different base material doublets at different spatial positions in the image. DIRA does not require geometrically consistent projections and has been extended to the 3D helical geometry [3].

The linear attenuation coefficient (LAC), $\mu(E)$, depends on the photon energy and, in the medical diagnostics energy range of 20–150 keV, includes photoelectric absorption, Compton scattering, and coherent scattering:

$$\mu(E) = \mu_p(E) + \mu_C(E) + \mu_{coh}(E), \tag{1}$$

where $\mu_p(E)$ is approximately proportional to E^{-3}, $\mu_C(E)$ is proportional to the Klein-Nishina cross section, and $\mu_{coh}(E)$ makes a relatively small contribution.

Maria Magnusson[1,2,3] (e-mail: maria.magnusson@liu.se), Gudrun Alm Carlsson[2,3], Michael Sandborg[2,3], Åsa Carlsson Tedgren[2,3,4] and Alexandr Malusek[2,3] are with [1]Dept. of Electrical Engineering, [2]Department of Health, Medicine and Caring Sciences, Division of Diagnostics and Specialist Medicine, [3]Center for Medical Image Science and Visualization (CMIV), Linköping University, SE-581 83 Linköping, Sweden and [4]Dept. of Medical Radiation Physics and Nuclear Medicine, Karolinska University Hospital, SE-171 76 Stockholm, Sweden.

With a base material doublet, $\mu(E)$ can be mathematically decomposed as

$$\mu(E) = w_1 \cdot \mu_1(E) + w_2 \cdot \mu_2(E), \tag{2}$$

where w_1 and w_2 are weight coefficients, and $\mu_1(E)$ and $\mu_2(E)$ are the LACs for the two base materials. The original base material doublet suggested in [1], from now on called the photo-Compton doublet (PC), used $\mu_p(E) \sim E^{-3}$, though some other authors also used $\mu_p(E) \sim E^{-2.8}$ [4].

Another common base material doublet is the water-bone doublet (WB), see for example [5]. Yet another one is the water-iodine doublet (WI), used by General Electric, see [6].

Of interest is the intrinsic dimensionality of the linear attenuation coefficient of biological tissues in the energy range 20–150 keV, which is the energy range for medical CT. Williamson et al. [7] investigated biological tissues with effective atomic numbers $Z = 2, \ldots, 20$ in the energy range 20–140 keV. For determination of individual (mass macroscopic) cross sections, $\mu_p(E)/\rho$ and $\mu_C(E)/\rho$, where ρ is the mass density, they recommend a Water-Polystyrene doublet for $Z = 1, \ldots, 8$ and a Water-'Calcium chloride solution' for $Z = 8, \ldots, 20$. They claim, however, that one base material doublet is sufficient to represent LAC values. Bornefalk [8] applied the Principal Component Analysis (PCA) to $Z = 1, \ldots, 20$ in the energy range 25–120 keV and discovered an intrinsic dimensionality of 3 to 4. Alvarez [9] stated that "A two function basis set is sufficient to approximate the attenuation coefficients of biological materials. However, a three or higher dimension basis set is needed if an externally administered high atomic number contrast agent is used."

Of interest is how the choice of different material bases affects the accuracy of reconstructions algorithms in DECT and what limits the intrinsic dimensionality of the cross sections in practice.

II. Methods

A. Visual investigation of the dimensionality of medical CT materials

Material decomposition to the water-bone doublet was performed for elements with $Z = 1, \ldots, 20$, water, lipid, protein, adipose, muscle, compact bone, femora spongiosa, Ti, Zn, I, Ba, and Ce for the following reasons. Most of the elements comprising human tissues have $Z \leq 20$. Soft tissues consist of mainly water, lipids and proteins, and bones consists of compact bone and spongiosa. Ti may be used

for implants, Zn can be found in small amounts in prostate and I, Ba, and Ce can be used as contrast agents. True mass attenuation coefficients for these materials were either taken directly from the EPDL97 library [10] or derived from elemental compositions taken from [2].

Weight coefficients w_1 and w_2 of the material decomposition for the two base materials were obtained by solving an equation system consisting of two equations (2), one at $E = 50$ keV and the other at $E = 88$ keV.

B. X-ray Spectra and Phantoms

An 80 kV spectrum and an 140 kV spectrum filtered with Sn were used, see Fig. 1.

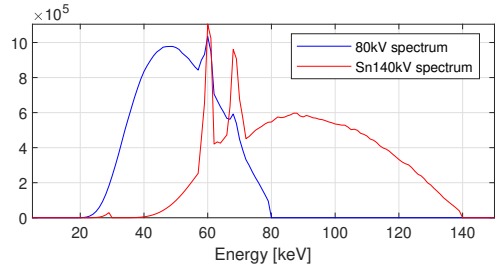


Fig. 1. Non-normalized energy distributions of photons in X-ray spectra.

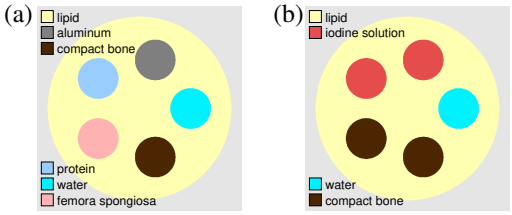


Fig. 2. Schematic drawing of the phantom without (a) and with (b) the iodine-water solution. The lipid cylinders with the diameter of 316 mm contained 5 rod inserts with the diameter of 70 mm containing different materials.

Two cylindrical lipid phantoms containing five rod inserts were used, see Fig. 2. Rod inserts of the first phantom were made of water, protein, compact bone, femura spongiosa and aluminum. Rod inserts of the second phantom were made of water, compact bone and iodine-water solution.

C. Setup for DIRA

Projections of the phantom were calculated with Drasim [11] in a fan beam geometry described in [2] and rebinned to parallel projections; this approach was used for the AM too. In the case of DIRA, only the phantom with iodine was processed. A threshold segmentation was used to determine regions with iodine, bone, and soft tissues. In the iodine and bone regions, two-material decompositions to the (water, iodine) and (compact bone, bone marrow) doublets, respectively, were used. In the soft tissue region, the three-material decomposition to the (water, lipid, protein) triplet was used, see [2] for more information.

D. Setup for the AM algorithm

Projections were calculated as in section II-C. The AM algorithm was implemented according to Ying et al [12]; the photo-Compton, water-bone, and water-iodine base material doublets were used. The AM reconstruction resulted in two images of the base material weight coefficients w_1 and w_2. Averages of w_1 and w_2 were taken in regions of interest (ROIs) inside the rod inserts and the water cylinder. From these values, LACs as functions of energy in the range 20–150 keV were obtained using Eq. (2). Also, three mono-energetic images for 30, 50, and 88 keV were calculated.

III. Results

The ability of the water-bone doublet to represent the material of interest is represented by the $\mu(E)/\mu_{tab}(E)$ ratio. This ratio, obtained by direct application of material decomposition via Eq. (2), is plotted in Figs. 3 and 4 for elements with $Z = 1, \ldots, 20$ and selected common human tissues, respectively.

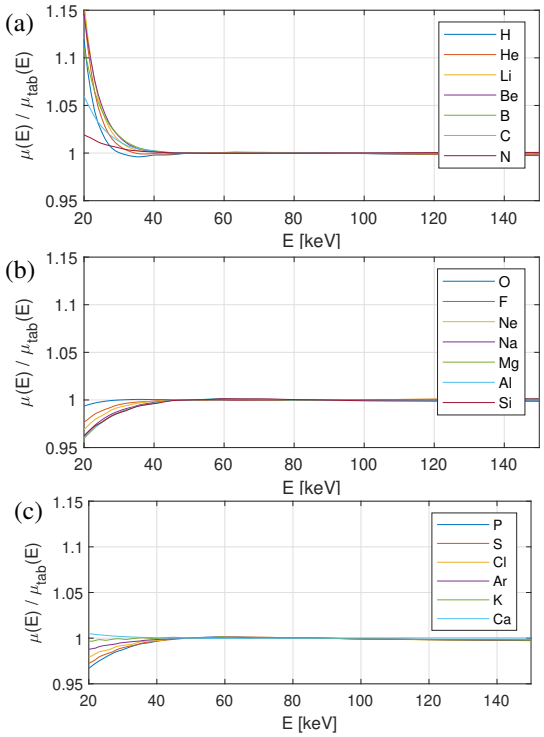


Fig. 3. The LAC approximated by the water-bone doublet relative to the tabulated LAC, $\mu(E)/\mu_{tab}(E)$, as a function of energy for elements with $Z = 1, \ldots, 20$.

Note that the elements are well represented for energies above $\approx$ 35 keV. For energies in the range 20–35 keV, the discrepancy is larger, especially for H, He, Li, Be, B. However, these substances, except of H, are typically not found in the human body. Cross sections of compounds containing H are typically dominated by the other elements. In Fig. 4, water, bone, lipid, protein, adipose, muscle and femora spongiosa are plotted. For those, the discrepancy in the range 20–35 keV is rather small. Moreover, the relative intensity of the X-ray spectrum (Fig. 1) is small in the range 20–35 keV.

Some other materials of interest for medical CT (Ti, Zn, I, Ba, and Ce) are plotted in Fig. 5. Note that Ti is represented as good as the human tissues by the water-bone doublet. Zn is not represented so well and the contrast agents I, Ba, and Ce are represented even worse. The K-edges around 40 keV are an additional complication.

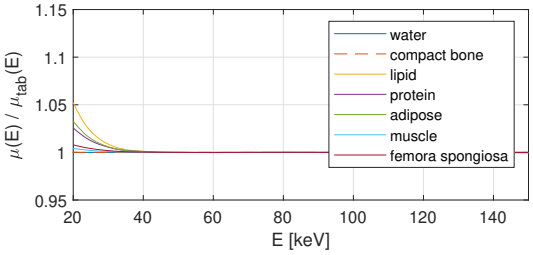


Fig. 4. Ratio $\mu(E)/\mu_{tab}(E)$ for common materials, $\mu_{tab}(E)$ is the tabulated LAC and $\mu(E)$ is the LAC approximated with the water-bone doublet.

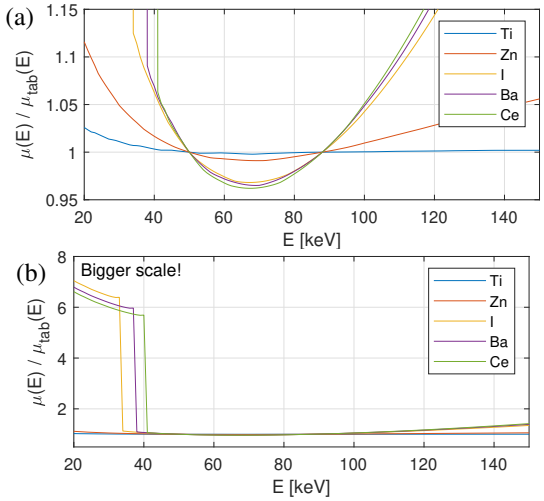


Fig. 5. The $\mu(E)/\mu_{tab}(E)$ ratio for some materials of interest in CT, where $\mu_{tab}(E)$ is the tabulated LAC and $\mu(E)$ is the LAC approximated with the water-bone doublet. Fine (a) and coarse (b) scales on the y-axis.

A. *Analysis of reconstructed data for phantom without iodine*

Fig. 6 shows images reconstructed by AM for the WB doublet at 30, 50, and 88 keV. There was a good agreement with tabulated values. The largest difference was for the aluminum rod at 30 keV, with a measured value of 299.6 m^{-1} and the tabulated 303.8 m^{-1}, giving a ratio of $299.6/303.8 \approx 0.986$. This value agrees with aluminum in Fig. 7 (top).

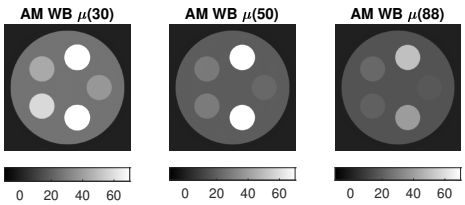


Fig. 6. LACs (in m^{-1}) for the phantom without iodine reconstructed by AM at 30, 50, and 88 keV, with the WB doublet.

Fig. 7 shows the $\mu(E)/\mu_{tab}(E)$ ratio obtained by AM for the WB, PC, and WI bases and the phantom without iodine. The WB and PC bases approximated the LAC well for $E > 40$ keV. In the range 20–35 keV, the relative difference between the LACs was larger, but it was still less than 5 %. In the case of WI, the K edge of iodine at 33.2 keV caused large discrepancies between the LACs in the range 20–40 keV. These discrepancies lead to a notable beam hardening artefact between the compact bone and aluminum inserts at 50 keV, see also Fig. 8.

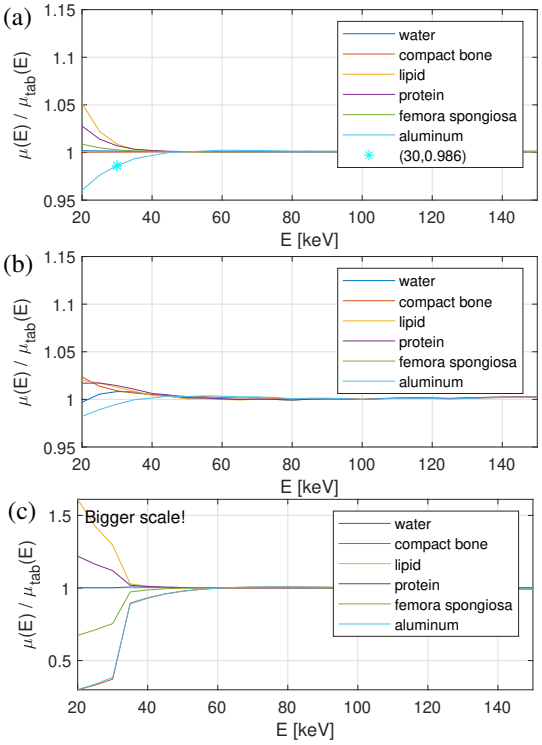


Fig. 7. The approximated LAC relative to the tabulated LAC, $\mu(E)/\mu_{tab}(E)$, as a function of energy calculated by the AM for the phantom without iodine and the water-bone (a), photo-Compton (b), and water-iodine (c) doublets.

Reconstructed images of the phantom without iodine displayed at a reduced LAC window are shown in Fig. 8. Contrary to the WB and PC doublets, the WI doublet produced a clearly visible beam hardening artefact.

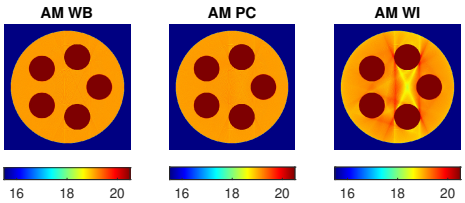


Fig. 8. LACs (in m^{-1}) reconstructed by AM at 50 keV for the WB, PC, and WI doublets and the phantom without iodine.

The PC doublet did not provide the correct fractions of the photoelectric and Compton LACs. Nevertheless, Fig. 9 shows that it agreed with the tabulated (photo+Compton+coherent) LAC when both components were added together. This shows that the PC doublet with the E^{-3} dependence can be used

in the AM for beam hardening removal but not for the determination of true photoelectric and Compton fractions.

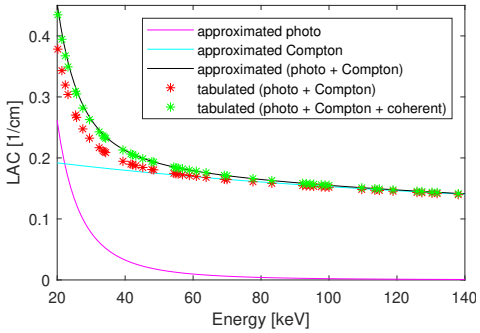


Fig. 9. Reconstructed approximated LAC and tabulated LAC for the lipid rod insert in the phantom without iodine as functions of photon energy. The approximated (photo+Compton) LAC agreed well with the tabulated (photo+Compton+coherent) LAC.

B. Analysis of reconstructed data for phantom with iodine

Images of the phantom with iodine were reconstructed by DIRA and by AM for the WB and WI doublets at 30, 50, and 88 keV. Reconstructed images displayed at a reduced LAC window are shown in Fig. 10. Both AM methods produced clearly visible beam hardening artefacts. DIRA reconstructed the phantom without such artefacts. More LAC values for 30 and 50 keV are given in Tab. I, where deviating values are highlighted. For 88 keV, the discrepancies were small.

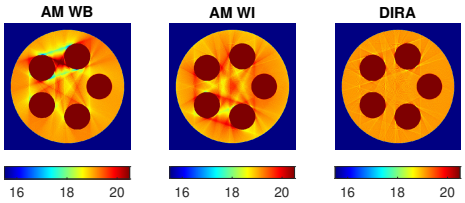


Fig. 10. LACs (in m^{-1}) for the phantom with iodine reconstructed at 50 keV by AM for the WB and WI doublets and by DIRA.

TABLE I
LACs (IN m^{-1}) FOR THE PHANTOM WITH IODINE. RECONSTRUCTIONS BY AM WITH WB, AM WITH WI, AND DIRA.

	30 keV				50 keV			
	true	WB	WI	DIRA	true	WB	WI	DIRA
water	37.6	37.1	37.6	37.6	22.7	22.6	22.8	22.7
bone	244.9	245.9	91.0	244.5	79.2	79.4	77.3	79.4
I-sol	90.3	412.1	90.5	89.4	100.0	105.5	100.3	100.3

IV. Discussion

Fig. 3 shows that it may be difficult to reach the dimensionality of 3–4 stated by Bornefalk [8] at the energy range of 35–150 keV for elements with $Z = 1, \ldots, 20$. In this energy range, the dimensionality of 2 is more likely, i.e. one doublet can predict the LAC values of the elements. All the major differences between the LACs of elements are in the energy range 20–35 keV. The relative number of photons in this part of the spectrum is small (Fig. 1). Moreover, in clinical applications most of these photons are absorbed by adult patient bodies.

The situation may be different for Spectral CT, however, where the low-energy channel may be in the range (20-40 keV). Additionally, for children, more photons of the initial spectrum in this low energy range will pass the body. Then the discrepancies found in energy range 20–35 keV may be more important.

V. Conclusion

The PC and WB doublets accurately approximated the LAC values for human tissues and elements with $Z = 1, \ldots, 20$, in the range 20-150 keV, though there was a small ($< 5\,\%$) discrepancy in the range 20–35 keV. The WI doublet did not represent the tissues as well as PC and WB; the largest discrepancies ($> 50\,\%$ in some cases) were in the range 20–40 keV.

LACs reconstructed with the AM and DIRA followed this trend. AM produced artefacts when iodine was present in the phantom together with human tissues, since AM can only work with one doublet. It was shown that these artefacts could be avoided with DIRA using a (water, lipid, protein) triplet for soft tissues, a (compact bone, bone marrow) doublet for bone and the WI doublet for the iodine solution.

Acknowledgment

David Ballestero is acknowledged for implementing AM.

References

[1] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography," *Physics in Medicine and Biology*, vol. 21, pp. 733–744, Sep. 1976.

[2] A. Malusek, M. Magnusson, M. Sandborg, and G. Alm Carlsson, "A model-based iterative reconstruction algorithm DIRA using patient-specific tissue classification via DECT for improved quantitative CT in dose planning," *Medical Physics*, vol. 44, no. 6, pp. 2345–2357, 2017.

[3] M. Magnusson, M. Björnfot, Å. C. Tedgren, G. A. Carlsson, M. Sandborg, and A. Malusek, "DIRA-3D—a model-based iterative algorithm for accurate dual-energy dual-source 3D helical CT," *Biomedical Physics & Engineering Express*, vol. 5, no. 6, p. 065005, Sep. 2019.

[4] J. B. Weaver and A. L. Huddleston, "Attenuation coefficients of body tissues using principal-components analysis," *Medical Physics*, vol. 12, no. 1, pp. 40–45, Jan. 1985.

[5] B. J. Heismann, B. T. Schmidt, and T. Flohr, *Spectral Computed Tomography*. Bellingham, Wash. (1000 20th St. Bellingham WA 98225-6705 USA): SPIE Press, 2012.

[6] P. R. S. Mendonca, P. Lamb, and D. V. Sahani, "A Flexible Method for Multi-Material Decomposition of Dual-Energy CT Images," *IEEE Transactions on Medical Imaging*, vol. 33, no. 1, pp. 99–116, Jan. 2014.

[7] J. F. Williamson, S. Li, S. Devic, B. R. Whiting, and F. A. Lerma, "On two-parameter models of photon cross sections: Application to dual-energy CT imaging: Two-parameter cross-section models," *Medical Physics*, vol. 33, no. 11, pp. 4115–4129, Oct. 2006.

[8] H. Bornefalk, "XCOM intrinsic dimensionality for low-Z elements at diagnostic energies," *Medical Physics*, vol. 39, no. 2, p. 654, 2012.

[9] R. E. Alvarez, "Dimensionality and noise in energy selective x-ray imaging," *Medical Physics*, vol. 40, no. 11, p. 111909, 2013.

[10] D. Cullen, J. Hubbell, and L. Kissel, "EPDL97: The Evaluated Photon Data Library,'97 Version, 1997," University of California, Lawrence Livermore National Laboratory, Livermore, CA 94550, Tech. Rep. UCRL-50400, 1997.

[11] K. Stierstorfer, "DRASIM: A CT-simulation tool," Siemens Medical Engineering, Internal Report, 2007.

[12] Z. Ying, R. Naidu, and C. R. Crawford, "Dual energy computed tomography for explosive detection," *Journal of X-ray Science and Technology*, vol. 14, no. 4, pp. 235–256, 2006.

KNN-based nanoparticle agent identification in PCD-CT

Sumin Baek and Okkyun Lee*

***Abstract*—One of the potential applications of photon counting detector (PCD)-based computed tomography (CT) is the identification of targeted nanoparticle contrast agents. However, conventional approaches, such as K-edge ratio and maximum likelihood-based methods, are vulnerable to noise. In this study, we propose a novel noise-robust K-nearest neighbors (KNN)-based method, which consists of calculating posterior sinograms using the KNN classifier, performing backprojection on the sinograms, and comparing the images of different classes. We show that the proposed method provides much clear identification of gold inserts for different concentrations and x-ray dose levels without background noise, compared to the conventional methods.**

***Index Terms*—Photon counting CT, nanoparticle contrast agents, K-edge ratio method, maximum likelihood, K nearest neighbors algorithm**

I. Introduction

One of the main features of the photon counting detector (PCD)-based computed tomography (CT) is the K-edge imaging applications by exploiting the multiple energy-dependent measurements from the PCD [1]. The K-edge imaging can be utilized for target localization combined with a new type of target-specific nanoparticle contrast agent. For example, in molecular and functional imaging, active targeting is extensively studied to visualize specific targets such as tumors and microcalcifications in small animal imaging using gold nanoparticle contrast agents [2, 3].

Conventional methods, such as maximum likelihood (ML)-based material decomposition, and K-edge ratio method can be used for the target localization by identifying the nanoparticle contrast agents. The K-edge ratio method identifies nanoparticle agents by comparing the bin-wise images between the K-edge pixel-by-pixel [4]. The ML-based method uses the noise statistics on the measurements to decompose the basis image of the nanoparticle contrast agents of interest [5]. However, both methods are reported to be sensitive to noise due to the lack of photon statistics [4, 6], and also increased dimension of the basis materials for the ML-based method [7].

In this study, we propose a noise-robust K-nearest neighbors (KNN)-based method for the nanoparticle contrast agent identification. We design the KNN method such that it classifies the PCD measurements into the class, including the nanoparticle agents, or the other class, without the agents. Then, the proposed method can be described in the following three steps: 1) Calculate posterior probabilities of each class for all detector channels and projection views using the KNN estimator, 2) backproject the posterior sinograms for each of the classes, and then 3) identify the targeted contrast agents by pixel-by-pixel comparison between the backprojection images. The proposed method is validated using simple and abdominal phantoms with gold nanoparticle contrast agents targeting localized regions and compared to those of the ML-based material decomposition and the K-edge ratio method.

In Section II, we summarize the conventional methods with background theory, describe the proposed method, and provide the simulation setups. Section III presents the main results, followed by the conclusion in Section IV.

S. Baek and O. Lee are with the Department of Robotics Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, 42988 Korea
*Corresponding author (contact: oklee@dgist.ac.kr)

This work was supported by the DGIST Start-up Fund Program of the Ministry of Science and ICT(2020010044).

II. Method

A. Material Decomposition

The line-integrals of an object, the energy-dependent linear attenuation coefficient, $\mu(E, \mathbf{r})[cm^{-1}]$, can be modeled by a linear combination of basis materials such as water, bone, and a nanoparticle contrast agent used for the K-edge imaging:

$$\int \mu(E, \mathbf{r}) d\mathbf{r} = \ell_w \mu_w(E) + \ell_b \mu_b(E) + \ell_k \mu_k(E), \tag{1}$$

where $\ell_w[cm]$, $\ell_b[cm]$, and $\ell_k[cm]$ are thicknesses of water, bone, and the contrast agent, respectively, along the x-ray path.

B. PCD Measurements Model

The number of measured photons at the j-th energy bin of the PCD, m_j, follows a Poisson distribution, and the mean value y_j can be modeled as follows:

$$y_j(\mathbf{A}_\ell) = \int_0^\infty S_j(E)\phi(E)e^{-\mathbf{A}_\ell^T \boldsymbol{\mu}_a(E)} dE, \quad for\ j = 1, 2, \cdots, J, \tag{2}$$

where J denotes the number of energy bins, $\mathbf{A}_\ell = [\ell_w, \ell_b, \ell_k]^T$, $\boldsymbol{\mu}_a(E) = [\mu_w(E), \mu_b(E), \mu_k(E)]^T$, and $\phi(E)$ [photons/keV] denotes an x-ray incident spectrum. The bin sensitivity function $S_j(E)$ addresses spectral distortions in the PCD [5].

C. Maximum Likelihood-based Estimator

The ML-based material decomposition estimates the basis line-integrals from the following optimization problem:

$$\hat{\mathbf{A}}_\ell = \arg\min_{\mathbf{A}_\ell} f(\mathbf{A}_\ell), \tag{3}$$

where $\hat{\mathbf{A}}_\ell$=$(\hat{\ell}_w, \hat{\ell}_b, \hat{\ell}_K)$. We assume that the measurements m_j $(j = 1, 2, \cdots, J)$ are independent Poisson random vari-

ables. The cost function in (3) is the negative log-likelihood from the noise statistics in the photon counts as follows [5]:

$$f(\mathbf{A}_\ell) = \sum_{j=1}^{J} [y_j(\mathbf{A}_\ell) - m_j \ln y_j(\mathbf{A}_\ell)]. \quad (4)$$

$\hat{\mathbf{A}}_\ell$ is calculated by minimizing (4) using given measurements and the system models such as $S_j(E)$ and $\phi(E)$. One can apply the filtered backprojection (FBP) to the estimated $\hat{\ell}_K$ (for all channels and views) to get the associated basis image for the identification of the nanoparticle agents.

D. K-edge Ratio Method

K-edge ratio method first calculates the following ratio of images (the division is done by pixel-by-pixel) [4],

$$R = \frac{Image(m_{j+1})}{Image(m_j)}, \quad (5)$$

where $Image(m_j)$ denotes the result of the FBP applied to the projection data for the j-th energy bin, $L(m_j)$:

$$L(m_j) = -\ln(\frac{m_j}{y_{j0}}), \quad (6)$$

where y_{j0} is the mean value when the object is absent for j-th energy bin. Due to the property of the K-edge in the attenuation curve, $Image(m_{j+1})$ (above the K-edge) has a larger pixel value than that of $Image(m_j)$ (below the K-edge) when the pixel contains the nanoparticle agents. On the other hand, $Image(m_{j+1})$ has a less pixel value than that of $Image(m_j)$ when the pixel is absent of the nanoparticle agents. Therefore, the method selects the region where $R > 1$ for the identification of the target [4].

E. Proposed Method

The main idea of the proposed method comes from the following assumption motivated from the nature of the active targeting: Targeted cells or tissues such as specific lesions where the nanoparticles accumulated are localized in the object, i.e., the PCD measurements can be classified into whether it includes the nanoparticles (C_1) or not (C_2), as illustrated in Fig. 1. One can apply a machine learning technique to classify

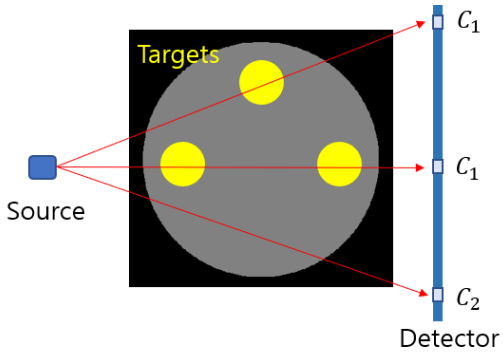


Fig. 1: The concept for the classification of the PCD measurements.

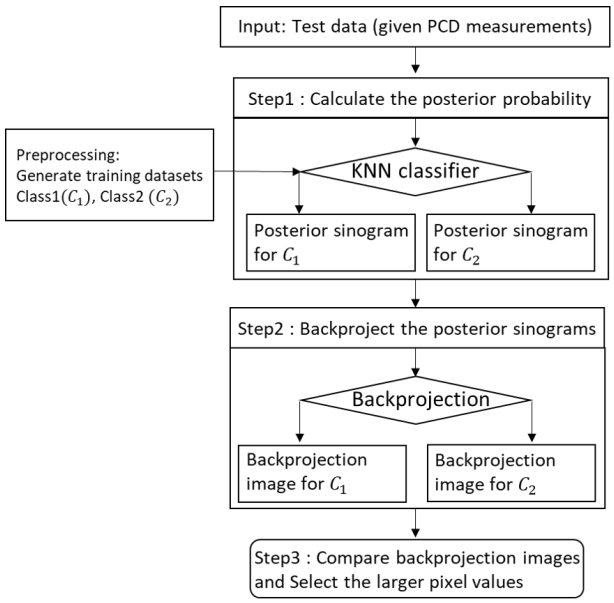


Fig. 2: The flowchart for the proposed method.

the PCD measurements into either C_1 or C_2. In this study, we use the KNN method for this purpose, and we calculate the posterior probabilities for each of the classes. The benefit of calculating the posterior probability is that one can apply backprojection to the posterior sinograms for each class such that it provides the reference to identify the targets in the image domain. The details of the method are described using a flowchart in Fig. 2. First, we generate the training datasets, which is the mean photon counts for various combination of the basis materials (detailed explanation is given in section 2.F). We use a given PCD measurements as the input in the algorithm. Then, the KNN classifier calculates the posterior probability for the input (for all channels and views) for each of the classes and generates the posterior sinograms (e.g., see Fig.3(a,b) and Fig.4(a,b)). Next, backproject is applied the posterior sinograms for each class (e.g., see Fig.3(c,d) and Fig.4(c,d)). Finally, we compare the backprojection images and select the pixel in C_1 when the value is larger than that of the C_2 for the purpose of the identification of the nanoparticle agents (e.g., see Fig.3(e) and Fig.4(e)).

F. Simulation Setup

The total number of x-ray photons was set as 10^6 (for high dose case) or 10^5 (for low dose case) for a 140kVp incident spectrum with aluminum pre-filtration. Four energy bins, 66, 80, 100, 114keV, were used (80keV is the K-edge for the gold), and the spectral distortion was modeled for a realistic simulation [5]. We used simple and abdominal phantoms [8]: A numerical water phantom with various materials as shown in Fig.3(f), including the inserts of gold nanoparticle agents of three different concentrations, and an abdominal phantom with gold nanoparticle agents of different shapes and concentrations as illustrated in Fig.4(f). The image size was 512 x 512: 0.5mm/pixel for the simple phantom and 0.75mm/pixel for the abdominal phantom. The number of detector (PCD)

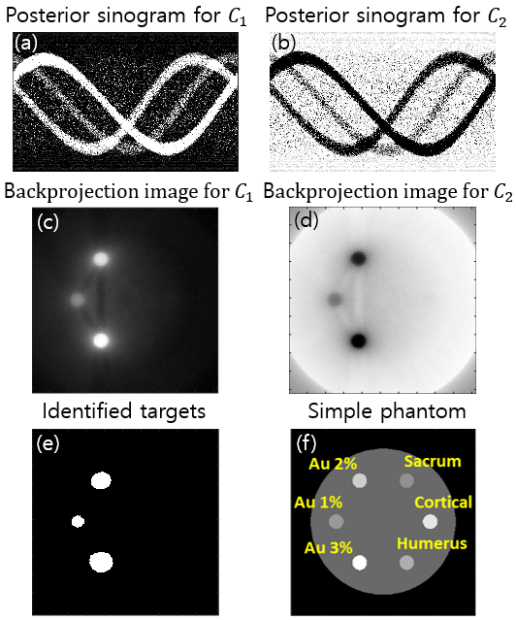


Fig. 3: The results of the proposed method for the simple phantom (high dose case): (a) posterior sinogram for C_1, (b) posterior sinogram for C_2, (c) backprojection image for C_1, (d) backprojection image for C_2, (e) identified targets, and (f) simple phantom.

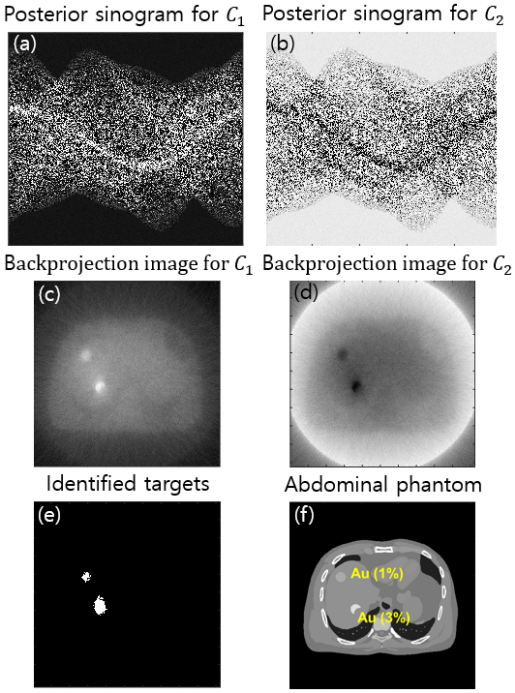


Fig. 4: The results of the proposed method for the abdominal phantom (low dose case): (a) posterior sinogram for C_1, (b) posterior sinogram for C_2, (c) backprojection image for C_1, (d) backprojection image for C_2, (e) identified targets, and (f) abdominal phantom.

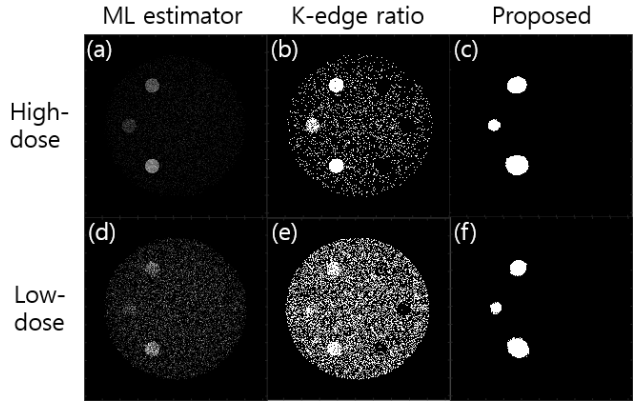


Fig. 5: Results of the nanoparticle agents identification for the simple phantom using various methods (first row: high dose, second row: low dose).

channels and the number of views per rotation were 601 and 1,000, respectively, with detector pitch of 1mm for the simple phantom and 1.5mm for the abdominal phantom. The equi-spaced 2D fan-beam geometry is used with the distance from the source to the iso-center, 50cm, and from the source to the detector, 100cm. We used ASTRA [9] for the simulation described above.

The training datasets for the KNN classifier were generated as follows: 1) Set a range for the three materials for the class C_1 such as water (0∼39cm), bone (-1∼8cm), and gold (1%; 0∼15cm), 2) select regular points on the range (we set the interval as 0.7 for water, 0.3 for bone and 0.6 for gold 1%), 3) generate the mean value y_j (j=1 to 4) for each point (in practice, this corresponds to the averaged measurements by multiple scanning for a calibration phantom), and 4) do the same for C_2 without the gold, and interval of the points were adjusted as 0.15 for water and 0.057 for bone so that the number of training data for all classes is about the same. The hyperparameter K was heuristically selected as 12 for the high dose case and 70 for the low dose case for both phantoms.

III. RESULTS

The result for the gold nanoparticle agents identification using various methods with simple phantom is illustrated in Fig. 5 for high and low dose cases (the intermediate results for the posterior sinograms and backprojection images for high dose are shown in Fig. 3). Note that both the K-edge ratio method and the proposed method provide thresholded (binary) images while the ML estimator offers a density image. How to threshold the image for the identification is challenging; thus, we show images (without thresholding; clipped from 0 to twice of the ground truth for Au 3%). In Fig. 5(a), the ML estimator separates three different concentrations of gold nanoparticles. In Fig. 5(b), the K-edge ratio method also separates the target, but it suffers from the background noise. However, in Fig. 5(c), the proposed method clearly localizes the three different concentrations of gold nanoparticles without background noise. For the low dose case, both the ML estimator and the K-edge ratio method suffer from severe background noise so that

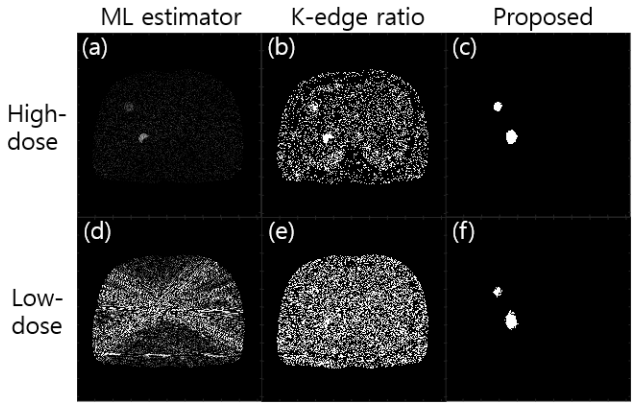


Fig. 6: Results of the nanoparticle agents identification for the abdominal phantom using various methods (first row: high dose, second row: low dose).

the targets with a low concentration (1% and 2%) are hard to be distinguished. On the other hand, the proposed method separates all the targets without background noise.

The result for the abdominal phantom using various methods is illustrated in Fig. 6 (the intermediate results for the posterior sinograms and backprojection images for low dose are shown in Fig. 4). For the high dose case, due to the increased size of the object, the target with 1% of Au is difficult to distinguish from the background noise while the proposed method clearly localizes both targets. For the low dose case, both conventional methods present a substantial amount of noise so that both targets are hard to bee seen there. On the other hand, the proposed method clearly distinguishes two targets without background noise.

IV. CONCLUSION

We developed the noise-robust K-nearest neighbors (KNN)-based method for nanoparticle contrast agents identification. The proposed method first calculates the posterior sinogram for each class using the KNN classifier, backproject the sinograms, and then identify the targeted contrast agents from the backprojection images. Compared to the conventional maximum likelihood-based estimator and the K-edge ratio method, the proposed method was able to localize the gold nanoparticles without background noise for both high and low dose cases.

REFERENCES

[1] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: single photon counting x-ray detectors in medical imaging," *Med. Phys.*, vol. 40, no. 10, pp. 100 901(1–19), October 2013.

[2] J. R. Ashton, J. L. West, and C. T. Badea, "In vivo small animal micro-CT using nanoparticle contrast agents," *Frontiers in Pharmacology*, vol. 6 (256), 2015.

[3] L. E. Cole, T. Vargo-Gogola, and R. K. Roeder, "Contrast-enhanced X-ray detection of breast microcalcifications in a murine model using targeted gold nanoparticles," *ACS Nano*, vol. 8, no. 7, pp. 7486–7496, July 2014.

[4] H. Ghadiri, M. R. Ay, M. B. Shiran, H. Soltanian-Zadeh, and H. Zaidi, "K-edge ratio method for identification of multiple nanoparticulate contrast agents by spectral CT imaging," *The British journal of radiology*, vol. 86, no. 1029, August 2013.

[5] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler, A. Livne, and R. Proksa, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys. Med. Biol.*, vol. 53, pp. 4031–4047, 2008.

[6] O. Lee, S. Kappler, C. Polster, and K. Taguchi, "Estimation of basis line-integrals in a spectral distortion-modeled photon counting detector using low-order polynomial approximation of x-ray transmittance," *IEEE Trans. on Medical Imaging*, vol. 36, no. 2, pp. 560–573, Feb. 2017.

[7] R. E. Alvarez, "Dimensionality and noise in energy selective x-ray imaging," *Med. Phys.*, vol. 40, no. 11, pp. 111 909(1–13), November 2013.

[8] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–3808, Aug. 2008.

[9] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible X-ray tomography using the ASTRA toolbox," *Opt. Express*, vol. 24, pp. 25 129–25 147, 2016.

ntCT: submicrometer laboratory X-Ray microtomography down to 0.2µm resolution

J. Graetz (Dittmann), D. Müller, A. Balles, T. Tuohimaa, T. Donath, C. Fella

***Abstract*—Following up on the development of the NanoTube X-ray source within the collaborative European NanoXCT project [1], a complete laboratory nano tomography setup (ntCT) further comprising a custom-built high energy photon counting detector has been realized. A preliminary quantification of the focal spot size is given here based on the structure size dependent contrast visibility of a siemens star, indicating a full width at half maximum of about 0.26µm. The ntCT system is now routinely operated at sub-micrometer tomographic resolutions in the range between 0.2–1µm. The translation of the quantified 2D imaging resolution into 3D tomographic applications is demonstrated on a micro-electronics sample exhibiting structures on the 10^{-7}m scale.**

I. Introduction

3D X-ray microscopy at the laboratory is a valuable tool both for life sciences and materials research. In contrast to other microscopy techniques, focusing is one of the central physical and technical challenges in the context of X-ray microscopy. Three general approaches are available, each of which has its strengths and weaknesses. Here, geometrically magnifying shadow microscopy using a highly focused X-ray source is considered, which can be easily combined with highly efficient – also for higher energies – direct converting X-ray detectors, making it highly attractive particularly for the imaging of technical samples based on heavier elements as compared to organic life sciences samples. Geometrically magnifying cone beam setups based on highly focused X-ray sources and large area detectors further provide a high degree of flexibility regarding resolution and sample size, being able to cover multiple orders of magnitude of sample sizes ranging from less than 0.1mm up to centimeters (depending on the absolute density of the considered sample).

Recent advances in the developments of X-ray sources allowed to push the limit of routinely achievable imaging resolution for this type of microscope to the 10^{-7}m scale. The ntCT integrates the Excillum NanoTube with a custom-built Dectris photon counting detector for higher energies up to the 100keV range. First results on NanoTube based imaging have been shown in [2], [3]. A quantification and comparison of tomograpic resolution has been given in [4]. Here, a preliminary quantification of the actual focal spot size shall be presented, as well as examples of high resolution tomographic images demonstrating the practical performance of the ntCT nano tomography system.

J. Graetz, A. Balles and C. Fella are with the Fraunhofer nano-tomography group of the magnetic resonance and X-ray imaging department MRB of the Fraunhofer IIS/EZRT in Würzburg, Germany.
(email: christian.fella@iis.fraunhofer.de)
D. Müller is with the Lehrstuhl für Röntgenmikroskopie at the University of Würzburg, Germany
T. Tuohimaa is with Excillum AB, Kista, Sweden
T. Donath is with Dectris, Baden-Daettwil, Switzerland

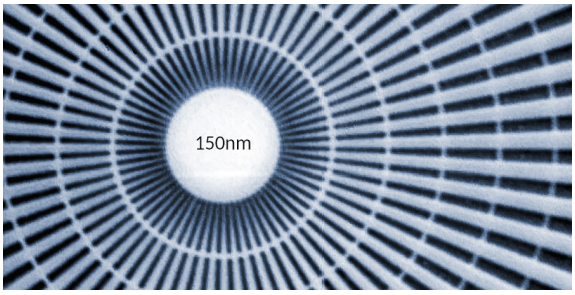


Figure 1. X-ray projection of a Siemenstar with 150nm line patterns at its innermost section. Magnification is achieved geometrically by means of a short source–object distance using a nanofocus X-ray source (Excillum NanoTube) and a photon counting direct detector (Dectris). The resolution in this geometry is dominated by the focal spot size (as opposed to the detector properties), which conversely can be deduced from the pattern contrast visibility.

Figure 2. The ntCT system, comprising an Excillum NanoTube and a custom built Dectris photon counting detector for energies up to the 100keV range.

II. Experimental Setup

The ntCT is a compact nano-tomography system based on a pre-series version of the Excillum NanoTube and a custom-built Dectris CdTe photon counting detector for high X-ray energies. With multi-axis control of both the detector

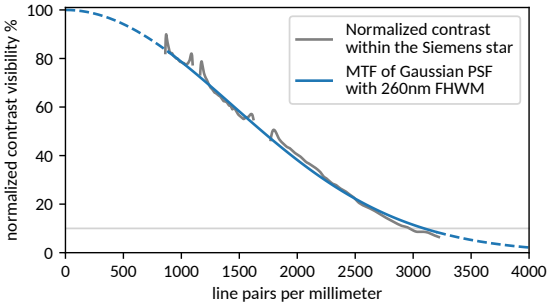


Figure 3. Evaluation of the average contrast visibility in dependence of the pattern frequency of the Siemens star shown in Fig. 1. The observed slope corresponds to the expected modulation transfer function of a Gaussian point spread of 260nm full width at half maximum, corresponding to 10% contrast at about 3100 line pairs per millimeter. This is consistent with the visual detectability of the 150nm structures within the Siemens star (Fig. 1)

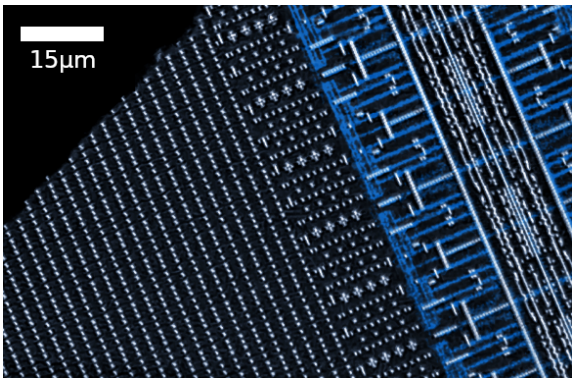


Figure 4. Subsection of a tomographic slice showing structures of an integrated circuit. The voxel sampling is about 0.2µm. The full field of view (not shown) is about 200µm wide. The smallest feature sizes visible in the image are on the voxel scale at 0.2µm.

and sample stage, high flexibility with respect to imaging geometries is provided. The system can be operated both in a high flux mode for micrometer resolutions as well as in a high resolution mode, that will be quantified in the following. Imaging is performed in geometric magnification with short source–object distances and magnification factors typically between 10^2 and 10^3. An acceleration voltage of 60keV is used. By working with large fan angles in the range above 30°, a large solid angle of the emitted X-rays is used for imaging. A custom auto-calibration technique for the projection geometry ensures optimal reconstructions [5], [6].

III. Results

In order to quantify the achievable resolution, a Siemens star featuring structures down to 150nm has been imaged at a magnification factor of about 1000 (Fig. 1). By polar unwrapping and Fourier analysis of the resulting line patterns, the modulation transfer function can be deduced. The result is shown in Figure 3. The three innermost sections of the Siemens star were fully contained in the projection image, covering a range of about 900 to 3300 line pairs per millimeter. The observed slope is consistent with a Gaussian point spread of 260nm full width at half maximum, corresponding to 10% contrast at about 3100 line pairs per millimeter, in accordance with the visual detectability of the smallest (150nm) features of the Siemens star in Fig. 1.

Figures 4 and 5 provide example micro-tomographic images acquired at magnification factors of 215 and 375 respectively to demonstrate the translation of this resolution into actual tomographic imaging applications on the ntCT. Figure 4 shows high contrast structures on the 10^{-7}m scale within an integrated circuit, representing also the maximum expectable resolution of the setup. Figure 5 shows a lower resolution scan of a radolaria fossile at a voxel scale of 0.35µm exhibiting in-line phase contrast.

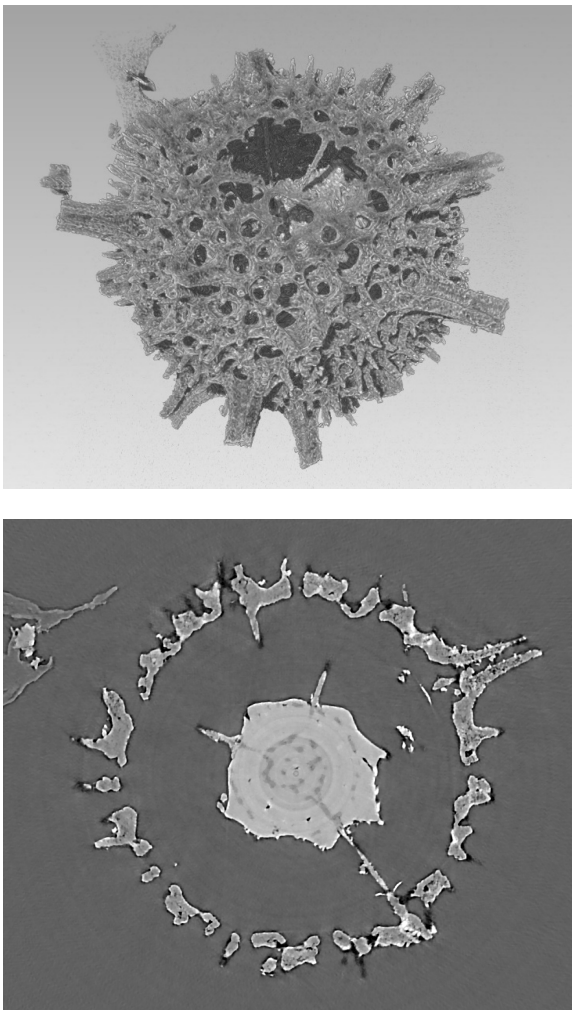


Figure 5. Volume visualization and tomographic slice of a radolaria of about 200µm diameter, scanned at a resolution of about 0.35µm. Strong phase contrast induced edge enhancement can be observed.

REFERENCES

[1] F. Nachtrab, M. Firsching, C. Speier, N. Uhlmann, P. Takman, T. Tuohimaa, C. Heinzl, J. Kastner, D. H. Larsson, A. Holmberg, G. Berti, M. Krumm, and S. C, "Nanoxct: development of a laboratory nano-ct system," *Proc SPIE Dev. in X-ray Tomography IX*, p. 92120L, 2014.

[2] M. Müller, I. S. de Oliveira, S. Allner, S. Ferstl, P. Bidola, K. Mechlem, A. Fehringer, L. Hehn, M. Dierlof, K. Achterhold, B. Gleich, J. Hammel, H. Jahn, G. Mayer, and F. Pfeiffer, "Myoanatomy of the velvet worm leg revealed by laboratory-based nanofocus x-ray source tomography," *PNAS*, vol. 21, pp. 12 378–12 383, 2017.

[3] C. Fella, J. Dittmann, D. Müller, T. Donath, T. Tuohimaa, and R. Hanke, "Implementation of a computed tomography system based on a laboratory-based nanofocus x-ray source," *Microscopy and Microanalysis*, vol. 24, no. S2, pp. 234–235, 2018.

[4] S. Zabler, M. Ullherr, C. Fella, R. Schielein, O. Focke, B. Zeller-Plumhoff, P. Lhuissier, W. DeBoever, and R. Hanke, "Comparing image quality in phase contrast sub x-ray tomography—a round-robin study," *NIMA*, vol. 951, p. 162992, 2020.

[5] J. Dittmann and S. Zabler, "Autocalibration of cone beam ct projection matrices based on arbitrary traceable features within a regular tomographic scan," *Proc. Intl. CT Meeting*, 2018.

[6] J. Dittmann, "Cone beam geometry calibration for micro computed tomography systems using arbitrary fiducial markers," *arXiv:1808.03337*, 2018.

ACKNOWLEDGEMENTS

The authors thank Prof. Dr. Randolf Hanke and Dr. Simon Zabler for facilitating the present work. Funding from the LeeBed EU project is acknowledged.

BeadNet: a network for automated spherical marker detection in radiographs for geometry calibration

V. Nguyen[1], J. De Beenhouwer[1], S. Bazrafkan[1], A-T. Hoang[2], S. Van Wassenbergh[3], J. Sijbers[1]
[1] imec - Vision Lab, Department of Physics, University of Antwerp, Belgium
[2] DiSTA, University of Insubria, Italy
[3] Functional Morphology Lab, University of Antwerp, Belgium

***Abstract*—Spherical markers are commonly used by phantom-based calibration methods for X-ray CT systems. Defining the position of the marker centers is therefore crucial to estimate the geometry parameters accurately. Although marker bearing structures are often built from materials of low X-ray attenuation, they still overlap with the marker in projection images. This complicates accurate determination of the marker centers.**
In this work, we explore the technique of Deep Learning to extract the marker center coordinates from the calibration projections. By training a Deep Learning network for each marker center coordinate, 2D positions of the marker are derived. With simulated as well as real experiments, it is shown that the trained Deep Learning networks can be used to accurately estimate the marker positions, and hence also the geometry of the X-ray CT system.

***Index Terms*—Deep Learning, calibration, cone-beam CT.**

I. Introduction

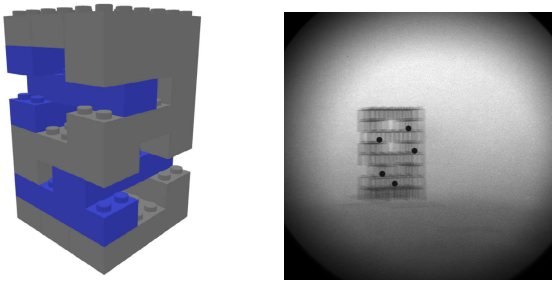
X-ray CT geometry calibration using phantoms often relies on spherical markers to calibrate the geometry parameters. In general, calibration procedures align the analytically calculated markers' orbits with the centers extracted from the calibration phantom radiographs, it is therefore important to define the centers of the markers on the calibration projection as accurately as possible. Several effects caused by the cone-beam X-ray geometry that hinder the extraction of the marker center positions were addressed in the work of Desbat et al. [1] including magnification, stretch, and asymmetrical factors. Desbat et al. [1] addressed three possible factors that generate asymmetrical marker projections in the X-ray radiographs. In that study, the calibration radiographs were first segmented to obtain the binary images containing the marker regions only. Marker centroids were calculated as center-of-masses of the marker regions. Sawyer et al. [2] located the marker centers by fitting a Gaussian and a projection of a sphere for corresponding marker projections. In both works, X-ray attenuation of the marker bearing structure on the projection data, was not taken into account. It was also difficult to accurately segment or to fit a fitting function for asymmetrical marker regions.
In the work presented by Liu et al. [3], the marker centers were extracted by firstly applying a grey-value threshold to remove non-marker structure as much as possible. Afterwards, several iterations of erosion followed by dilation filters were applied to remove narrow stripes of the supporting material. Finally, a circular Hough transform was used to locate the center of each marker on the X-ray radiograph. The procedure requires the diameters of the markers to be known.
In our previous work [4], we presented a technique based on the normalized cross correlation (NCC) between the markers and a template to define region-of-interests (ROIs) around the markers on the calibration data. This was followed by a subsample matching process to locate the center of the markers more precisely. The extracted center positions were then used to calibrate a real X-ray CT system [5].
In this paper, we propose Deep Learning to extract the calibration marker centers for which two neural networks are trained separately for each of 2D center coordinates on the projection. We validate our method on simulated and experimental calibration use cases and compare to the conventional NCC approach. The paper is structured as follows. Section II presents our methodology to generate the training dataset along with the process to extract the bead centers accurately using BeadNet. Section III discusses the simulation experiments that were performed to validate our proposed procedure as well as the experiments and results using real LEGO phantom datasets. Finally, further discussion and conclusions are presented in section IV.

II. Methodology

A. Geometry estimation

Fig. 1 shows the stereoscopic X-ray system used in this study: The 3-Dimensional DYnamic MOrphology using X-rays (3D^{2}YMOX) system that is designed for morphological and biomechanical research on animals [5]. Fig. 2 demonstrates the LEGO phantom (Fig. 2a) and its X-ray radiograph acquired with the 3D^{2}YMOX system (Fig. 2b). The geometry of this modular system can be described by six parameters to parameterize detector translations $\left\{\Delta x^d, \Delta y^d, \Delta z^d\right\}$ and detector orientations $\left\{\theta^d, \phi^d, \eta^d\right\}$ in 3D space.
In addition, six more parameters were needed to define the object's orientations $\{\theta^o, \phi^o, \eta^o\}$ and translations $\{\Delta x^o, \Delta y^o, \Delta z^o\}$ with respect to the rotation axis coordinate system. As has been discussed in [4], the geometry calibration procedure involved iteratively minimizing the total Euclidean norm between the reference and measured coordinates across all projections and marker centers. Knowledge of the position of the marker centers is key for accurate estimation of the X-ray system's geometry.

Fig. 1. The 3D^{2}YMOX system.



(a) (b)

Fig. 2. A real LEGO calibration phantom (a) and its X-ray projection from the 3D^{2}YMOX system (b).

B. BeadNet

In the conventional method [4], the center of each bead is estimated from the center-of-mass (COM) of its corresponding region of interest (ROI), which is extracted from the calibration projections. However, cone-beam effects and the overlapping projection of the holding structure on the ROIs complicate the COM calculation. It calls for a robust method that can handle different cone-beam geometry effects as well as asymmetric ROIs and derive center locations more accurately.

The center estimation procedure can be described as finding a mapping model F that takes the marker ROIs $\boldsymbol{x}$ as inputs with parameters $\boldsymbol{W}$ and returns the corresponding center coordinates. $\boldsymbol{W}$ is obtained through an optimization process that minimizes the difference between $F(\boldsymbol{x}, \{\boldsymbol{W}\})$ and ground-truth center coordinates (u^{gt}, v^{gt}).

$$\hat{\boldsymbol{W}} = \arg\min_{\boldsymbol{W}} \left\{ \left[F(\boldsymbol{x}, \{\boldsymbol{W}\}) - (u', v') \right]^2 \right\} \qquad (1)$$

In Eq. (1), F represents any Deep Learning model that learns abstract features from the input ROIs and maps them to the center coordinates of the bead in the ROI considering t being the hyperbolic tangent function and $t(u^{gt}) = u'$, $t(v^{gt}) = v'$. Two Deep Learning models (BeadNet) were trained separately for each center coordinate regression. The goal of BeadNet is to find a suitable abstract feature to map a marker ROI to its corresponding center coordinates. Choosing a feature learning model is important to have accurate center inference. Resnet-50 [6] emerges as a Deep Learning model that is trained on more than million images for object classifications. Resnet-50 is 50 layers deep and is divided into five convolution layers. We exploited the robustness of the pre-trained Resnet-50 model in learning object abstract features to continue training for our marker center extraction. To this end, two BeadNets were trained at a learning rate of 0.01 using an adaptive learning rate optimizer with a batch size of 40 ROIs.

Strategy to generate training dataset

The generation of training data is one of the key steps for applications of Deep Learning. The X-ray energy spectrum can be different for each acquisition. Hence, the training dataset included projections of the calibration phantom simulated with different X-ray source spectra. Moreover, as the cone-beam X-ray CT system was parameterized using 12 degrees-of-freedom, the training dataset needs to mimic a variety of possible geometry configurations, which are common settings of the 3D^{2}YMOX system. A set of 400 projections were simulated for each of 120 angles covering 360° rotation. For each set, the object and detector orientations and translations were modified by a random value generated from a uniform distribution in the intervals of $[-10, 10]^\circ$ and $[-30, 30]$ mm, respectively. The object yaw ϕ^o was generated randomly up to 200° simulating varied orientations of the calibration phantom. A 1100 mm source-detector distance (SDD) along with varied source-object distances (SODs) were simulated for typical positions of the sources, the object, and the detectors of the 3D^{2}YMOX system.

Reference marker center orbits that correspond to the simulated geometries were calculated analytically with which 25 ROIs were extracted around each marker from every simulated radiograph using reference marker positions. The training dataset contains the marker ROIs and their corresponding ground-truth center coordinates. In this work, the ROIs were at a size of 39×39 pixels that mainly cover center patches of the marker projections.

III. Experiments and Results

A. Generation of experimental datasets

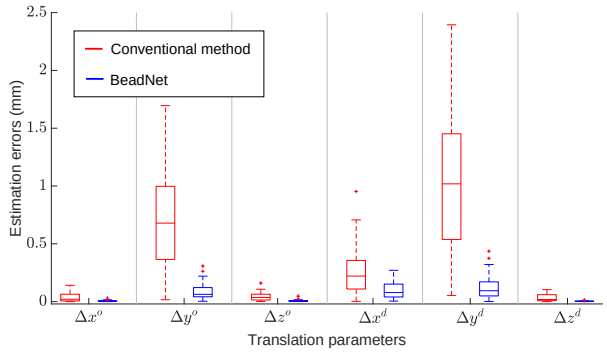
The training and simulated datasets were generated using the LEGO phantom STL models and the ASTRA CAD projector toolbox [7]. The system vector geometry was calculated with respect to the geometry misalignment parameters for every projection angle using the ASTRA Toolbox [8], [9]. Then, the ASTRA CAD projector simulated X-ray radiographs of the phantom with a 150 keV polychromatic spectrum, the predefined vector geometries, and a detector pixel size of 142 μm, which corresponds to the pixel size of the 3D^{2}YMOX system.

The validation dataset containing 44 simulated sets were generated mimicking different X-ray cone-beam geometries. Corresponding marker ROIs along with the initial marker center coordinates were extracted to feed the trained BeadNets so as to estimate the center coordinates more accurately. Two more simulated datasets were generated for 3D CT reconstruction evaluations including a set of the calibration phantom and a set of a test phantom projections. These two

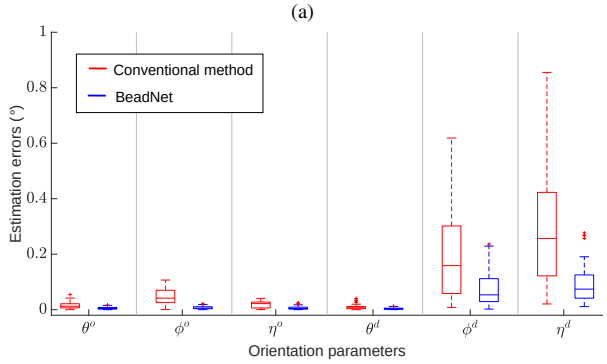
datasets were simulated with the same detector translation and orientation parameters but different randomly generated object positions and orientations.

Validation of BeadNet using simulated datasets

The bead centers were extracted from the validation



(a)



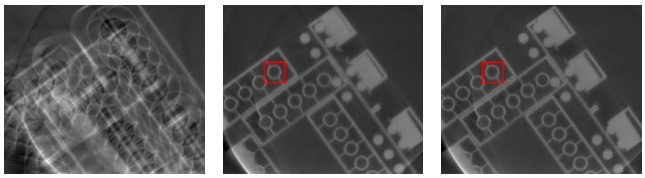
(b)

Fig. 3. Estimation errors of the translation (a) and the orientation (b) parameters using the conventional method (red) and BeadNet (blue) to extract the marker centers.

dataset with conventional NCC method and BeadNet. From the extracted bead trajectories, the X-ray system's geometry was then estimated. Fig. 3 shows the geometry estimation errors from the NCC and BeadNet method. As shown in this figure, the geometry parameters are more accurately estimated when relying on the BeadNet marker center estimates compared to NCC based estimation. The translation parameters are calibrated with average errors of 130 μm from BeadNet method, while conventional method yields mean error of up to 1000 μm. The orientation parameters are also estimated close to the GTs by BeadNet inferred marker centers with mean error of less than 0.1° while mean error from conventional method is of 0.25°. These estimates show that BeadNet yields accurate marker positions for geometry calibration.

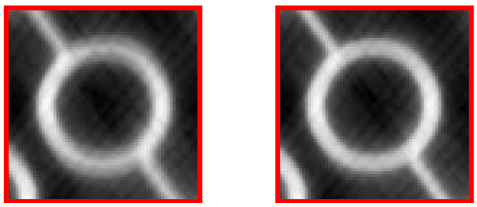
Reconstruction from simulated phantom projections

In practice, multiple objects are usually scanned in a same geometry setting. That is, the sources and detectors are placed in a fixed position in terms of the detector



(a) (b) (c)

Fig. 4. Cross-sections of the reconstructed volume using simulated phantom dataset before (a) and after geometry calibration with the conventional technique (b) and BeadNet (c). Zoomed versions from the red squares are shown in Fig. 5.



(a) (b)

Fig. 5. With the conventional method, misalignment artifacts still appear at the edges of the bricks in the reconstruction (a), while the artifacts are substantially reduced in the CT slice using BeadNet (b).

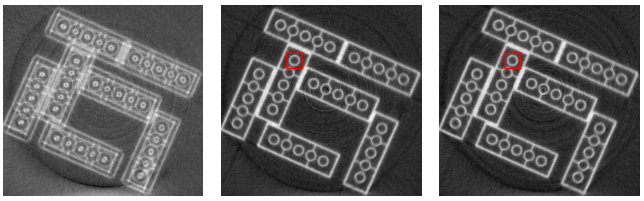
translations $\{\Delta x^d, \Delta y^d, \Delta z^d\}$ and orientations $\{\theta^d, \phi^d, \eta^d\}$ in the acquisition while the target objects are altered to acquire different datasets. The geometry was calibrated with the calibration phantom dataset for the reconstruction a test phantom, which was also built from LEGO bricks, to assess 3D CT quality and to anticipate the artifacts caused by the remaining geometry misalignments.

Fig. 4 shows the cross-sections of a real phantom reconstruction before (Fig. 4a) and after (Fig. 4b, Fig. 4c) geometry alignment. The vertical translation of the detector Δy^d was estimated more accurately using the marker orbits obtained with BeadNet. This is why the two axial slices shown in Fig. 4b and in 4c do not coincide but are shifted vertically in the 3D CT volumes. Apparent artifacts can be observed in the reconstructed slices before correcting the geometry (Fig. 4a), while the brick structure is clearly revealed in Fig. 4b and 4c. These images demonstrate that the misaligned geometry was substantially compensated in the reconstruction.

Reconstruction from real test phantom projections

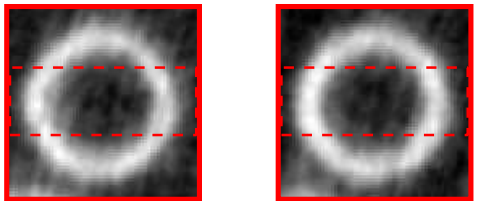
The geometry of the 3D^{2}YMOX system was calibrated with a real LEGO phantom. The calibration data were firstly flatfield and log corrected before they were undistorted to remove the pincushion distortion due to the intensifier curvature [10], and the sigmoidal distortion caused by the magnetic field generated during the stage rotation [11]. The bead center trajectories were extracted by the conventional method and BeadNet, and used to estimate the geometry parameters.

In this validation, the geometry was corrected for the misalignments before reconstructing a real test dataset obtained in the same geometric configuration with that the calibration projections were acquired. Fig. 6 shows the



(a) (b) (c)

Fig. 6. Cross-sections extracted from the reconstructions of a LEGO test phantom before (a) and after geometry calibration using conventional method (b) and BeadNet (c) to extract the marker centers. Without compensating the geometry misalignment, the internal brick structures are distorted (a). In (b, c), the artifacts are considerably reduced revealing sharp edges and apparent shapes of the LEGO bricks. Moreover, the artifacts are better suppressed in (c) than in (b) as highlighted in red and shown in Fig. 7b and Fig. 7a, respectively.



(a) (b)

Fig. 7. Misalignment artifacts still appear at the edges of the LEGO bricks and distort the shapes in the conventional method (a), while the edges are better recovered and sharper in the CT slice using BeadNet (b).

reconstructed slices of the test phantom before (Fig. 6a) and after (Fig. 6b, Fig. 6c) calibration. The reconstructed volume before compensating the geometry misalignment suffered from severe artifacts in the form of blurred edges of the LEGO bricks. In Fig. 6b and Fig. 6c, however, the shapes and edges of the bricks are well recovered in the reconstruction. This demonstrates that our estimation algorithm is capable of calibrating a real X-ray CT system. Moreover, the artifacts are better suppressed in Fig. 6c than in Fig. 6b as highlighted in red and displayed in Fig. 7a and Fig. 7b, respectively. Additionally, Fig. 8 shows the accumulated intensity profiles that were plotted through the center rows (dashed red) in the ROIs Fig. 7a and Fig. 7b from the conventional NCC method (orange) and BeadNet (blue). The line plots indicate that the contrast was slightly improved in the reconstruction from BeadNet method.

IV. Discussion & conclusion

In this paper, we presented a robust method to extract the centers of the spherical calibration markers on the X-ray radiographs using Deep Learning. Simulation experiments demonstrated that BeadNet can deliver better estimation of the marker centers and consequently the geometry parameters were estimated more accurately. After geometry misalignment correction, the misalignment artifacts were significantly suppressed from the 3D CT volume. From the marker center orbits extracted by BeadNet the geometry parameters are calibrated, and hence we obtain better quality of the CT reconstructions compared to the results with conventional NCC method. The reconstruction from the real datasets acquired from the

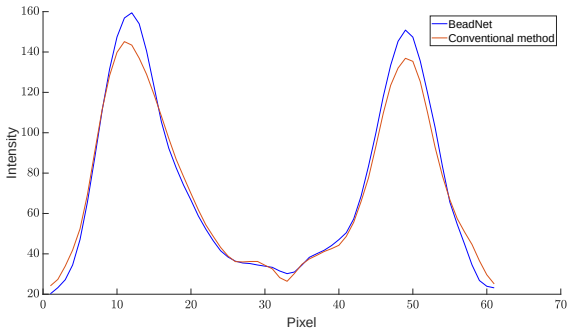


Fig. 8. Intensity profiles plotted through the rows (dashed boxes) in Fig. 7a and Fig. 7b.

3D^{2}YMOX system demonstrated that BeadNet can accurately extract the center of the markers on the real X-ray projections.

Acknowledgements

This work has been supported by the University Research Fund, UAntwerp BOF-GOA 2016 33927, the Research Foundation - Flanders (FWO) SBO project MetroFlex (S004217N), and the Flemish Government (AI Vlaanderen programme).

References

[1] L. Desbat, R. Clackdoyle, L. Grezes-Besset, C. Mennessier, and I. Bricault, “Cone-beam imaging of delta functions,” in *IEEE Nuclear Science Symposium Conference Record*, vol. 5. IEEE, 2006, pp. 2859–2863.

[2] S. A. Sawyer and E. C. Frey, “Practical improvement of geometric parameter estimation for cone beam microCT imaging,” in *IEEE Nuclear Science Symposium Conference Record*, vol. 4. IEEE, 2005, pp. 2315–2319.

[3] Y. Liu, “Improve industrial cone-beam computed tomography by integrating prior information,” Ph.D. dissertation, ETH Zurich, 2001.

[4] V. Nguyen, J. De Beenhouwer, J. Sanctorum, S. Van Wassenbergh, P. Aerts, J. J. J. Dirckx, and J. Sijbers, “A low-cost and easy-to-use phantom for cone-beam geometry calibration of a tomographic X-ray system,” *9th Conference on Industrial Computed Tomography, (iCT 2019)*, pp. 1–8, 2019.

[5] 3D2YMOX, 2018-06-19, 3-Dimensional DYnamic MOrphology using X-rays. [Online]. Available: https://www.uantwerpen.be/en/research-groups/3d2ymox/

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, 2016, pp. 770–778.

[7] A. Marinovszki, J. De Beenhouwer, and J. Sijbers, “An efficient CAD projector for X-ray projection based 3D inspection with the ASTRA Toolbox,” in *8th Conference on Industrial Computed Tomography*, 2018.

[8] W. Van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, “Fast and flexible X-ray tomography using the ASTRA toolbox,” *Optics Express*, vol. 24, no. 22, pp. 25 129–25 147, 2016.

[9] W. Van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, “The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography,” *Ultramicroscopy*, vol. 157, no. 2015, pp. 35–47, 2015.

[10] J. G. Sanctorum, D. Adriaens, J. J. J. Dirckx, J. Sijbers, C. Van Ginneken, P. Aerts, and S. Van Wassenbergh, “Methods for characterization and optimisation of measuring performance of stereoscopic x-ray systems with image intensifiers,” *Measurement Science and Technology*, vol. 30, no. 10, p. 105701, 2019.

[11] J. G. Sanctorum, V. Nguyen, J. Sijbers, S. Van Wassenbergh, and J. J. J. Dirckx, “Projection angle adapted distortion correction in high-speed image-intensifier based tomography,” in *6th International Conference on Image Formation in X-Ray Computed Tomography*, 2020.

Spectral calibration and non-linear intensity correction of photon-counting detectors for spectral CT

Emil Y. Sidky[1], Emily R. Paul[1], Taly Gilat-Schmidt[2], and Xiaochuan Pan[1]

***Abstract*—This work develops a calibration model for photon counting detectors for the purpose of providing accurate projection data for spectral computed tomography (CT) imaging. The calibration model fits the spectral response and transmitted intensity dependence for all energy binned windows simultaneously. The fitting is performed one detector pixel at a time using transmission data from a known object of known composition, a step-wedge phantom composed of different thicknesses of Aluminum, a bone equivalent, and Polymethyl methacrylate (PMMA), a soft-tissue equivalent. The spectral calibration and non-linear intensity correction is validated on transmission measurements through a third material, Teflon, at different X-ray photon flux levels.**

I. Introduction

Photon-counting detectors are a promising technology for application in spectral computed tomography (CT), because they have the ability to simultaneously acquire multiple projections with different X-ray spectral sensitivities. One of the potential applications for photon-counting based spectral CT is quantitative imaging of subjects composed of more than two types of tissues. This situation comes up in imaging patients with metal or when using K-edge contrast agents [1].

An important aspect of performing quantitative imaging in spectral CT with photon-counting detectors is calibrating the spectral response of the system, which includes the X-ray source spectrum, detector response, and additional physics that may impact the relationship between incident and detected numbers of photons. One approach to spectral CT calibration is to estimate a spectral response from transmission measurements through a known object of known composition [2,3], but this model assumes a linear relationship between detected and incident photon numbers.

[1]The University of Chicago, Department of Radiology MC-2026, 5841 S. Maryland Avenue, Chicago IL, 60637.

[2]Marquette University, Department of Biomedical Engineering, PO Box 1881, Milwaukee WI, 53201.

In this work, we aim to extend the spectrum estimation to include modeling the non-linear relationship between incident and detected photon numbers caused by detector physical factors such as pulse pile-up [4]. Dickmann *et al.* [5] extend spectrum calibration from transmission measurements by allowing for a spectral sensitivity function that depends on the photon flux of the incident X-ray beam. The approach we pursue here is to augment the X-ray transmission model, which includes the spectral response, with a polynomial function that relates the incident and detected X-ray photon flux. The coefficients of this polynomial along with spectral response parameters are estimated from transmission measurements through a step-wedge phantom containing discrete, known thicknesses of Aluminum and PMMA. In Sec. II the ideal transmission model and the spectrum estimation model with non-linear intensity correction are presented. In Sec. III experimental results for the calibration of the photon-counting detector are shown. Validation results are also presented, which consist of prediction of X-ray transmission through a Teflon slab at different flux levels.

II. Methods

For the photon-counting detector calibration set up, an X-ray transmission model along a single ray between the X-ray source and a single detector bin is needed. The continuous, ideal transmission model that includes X-ray spectral information is written

$$I_{w\ell} = \int S_w(E) \exp\left[-\int_\ell \mu(E, \vec{r}(t))dt\right] dE, \quad (1)$$

where $I_{w\ell}$ is the transmitted X-ray photon fluence along ray ℓ in energy window w; t is a parameter indicating location along ℓ; $S_w(E)$ is the spectral response; and $\mu(E, \vec{r}(t))$ is the energy and spatially dependent linear X-ray attenuation coefficient.

Employing a standard material-expansion decomposition, the spatially dependent attenuation map becomes

$$\mu(E, \vec{r}(t)) = \sum_m \left(\frac{\mu_m(E)}{\rho_m}\right) \rho_m f_m(\vec{r}[t]), \quad (2)$$

where ρ_m is the density of material m; $\mu_m(E)/\rho_m$ is the mass attenuation coefficient of material m; and $f_m(\vec{r})$ is the spatial map for material m. The spectrally-resolved transmission model is used for a calibration experiment with a step-wedge phantom composed of discrete values of material thicknesses. With this in mind and exploiting the materials decomposition, the discretized ideal transmission model becomes

$$I_{w\ell} = N_w \sum_i s_{wi} \exp\left[-\sum_m \mu_m t_{m\ell}\right] \Delta E, \quad (3)$$

where ray index ℓ now indicates discrete thickness combinations of the materials index by m; N_w is the incident photon number on energy window w and s_{wi} is the normalized spectral response discretized over the energy index i, i.e.

$$s_w(E) = \frac{S_w(E)}{\int S_w(E)dE} \quad (4)$$
$$s_{wi} = s_w(E_i);$$

and ΔE is the energy discretization width. The basic spectrum estimation problem assumes knowledge of attenuation coefficients μ_m and material thicknesses $t_{m\ell}$. Spectral sensitivity s_{wi} is estimated from the transmission measurements $I_{w\ell}$.

The proposed calibration model includes four main components: (1) a prior estimate of the X-ray source spectrum, (2) prior estimates of the binned energy window sensitivities, (3) an overall unknown function that modifies the combined spectral sensitivities, and (4) an unknown polynomial for each energy window mapping the incident transmission intensity to the detected transmission intensity. The coefficients of all of the polynomial functions are determined by fitting the transmission measurement model to the actual transmission measurements.

The spectral sensitivity is expressed as

$$S_w(E) = D_w(E)R(E)\exp\left(-\sum_{j=1}^{N_\beta}\beta_j(E/E_{\max})^j\right), \quad (5)$$

where $D_w(E)$ is the detector response model for window w; $R(E)$ is the X-ray source spectrum model; and the exponential factor with a polynomial function allows for overall modification of the spectral response as needed to fit the transmission measurements; i.e. coefficients β_j are unknown *a priori*. This particular functional form is used because it cannot be negative, and polynomial order N_β is chosen to be an even number so that a positive $\beta_{i=N_\beta}$ will result in a decaying function in both directions of energy E.

The corresponding normalized spectral response, s_{wi}, is obtained by substituting Eq. (5) into Eq. (4). To model the non-linear intensity transformation, we first define the transmission fraction

$$T_{w\ell} = \sum_i s_{wi}(\beta)\exp\left[-\sum_m \mu_m t_{m\ell}\right]\Delta E,$$

where $T_{w\ell} = I_{w\ell}/N_w$ and the dependence of s on unknown parameters β is explicitly indicated. The models for the measured and incident transmission fractions are related by another polynomial

$$T_{w\ell}^{\text{meas}}(\alpha,\beta) = T_{w\ell}^{\text{incid}}(\beta) + \sum_{k=0}^{N_\alpha}\alpha_k(T_{w\ell}^{\text{incid}}(\beta))^k. \quad (6)$$

This particular polynomial expression is chosen so that $T^{\text{meas}} = T^{\text{incid}}$ if all coefficients α_k are zero.

Coefficients α and β are determined by minimizing the Kullback-Leibler divergence between the measurement model $T_{w\ell}^{\text{meas}}(\alpha,\beta)$ and the actual measured transmission fractions for all energy windows w

$$\alpha^*,\beta^* = \arg\min_{\alpha,\beta}\sum_{w\ell} D_{\text{KL}}\left(T_{w\ell}^{\text{data}}, \max(T_{w\ell}^{\text{meas}}(\alpha,\beta), 10^{-5})\right).$$

The thresholding of the model to values greater than 10^{-5} is necessary to avoid computing logarithms of negative numbers, as the transmission model may go negative depending on the value of the coefficients α_k. This problem is non-convex, but relatively small scale. Due to the non-convexity, a good initial guess is needed and the model is designed so that $\alpha = \beta = 0$ is a good initial estimate and zero initialization is used in the presented results. To perform the optimization we use the Broyden Fletcher Goldfarb Shanno (BFGS) algorithm, which is a quasi-Newton method.

III. Results

Photon-counting spectral data was acquired on a bench-top spectral CT system with a CdTe photon counting detector (DxRay, Northridge, CA). The detector consisted of an array of 4×64, 1.4mm × 1.0 mm pixels with four comparator channels per pixel and a maximum count rate of 10^6 counts per second per pixel. Data were acquired with a tube voltage of 90 kV and with four energy windows with thresholds of 22, 45, 55, and 65 keV. The energy window data represented the number of photons counted between the energy thresholds. Calibration data were obtained by acquiring projection measurements through a step wedge phantom consisting of 25 combinations of PMMA and Aluminum thicknesses,

created by combining 0 to 4 PMMA slabs and 0 to 4 Aluminum slabs. The PMMA and Aluminum slabs had a thickness of 2.54cm and 0.64cm, respectively. A set of calibration training data was acquired at five flux levels corresponding to 9%, 14%, 27%, 40%, and 54% of the maximum detector count rate. The acquisition time was adjusted at each flux level such that the tube current time product (total fluence) remained at constant value of 0.456 mAs across all flux settings.

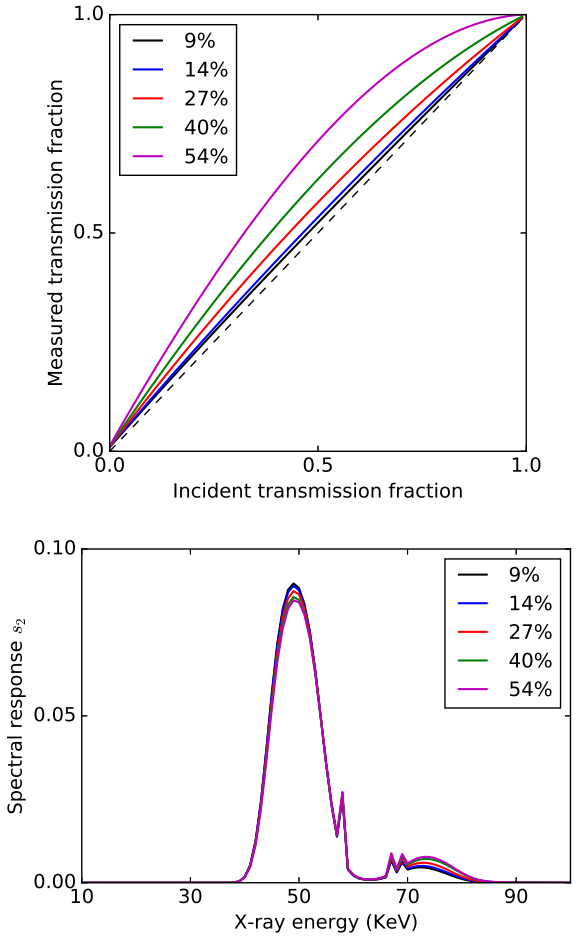


Fig. 1. Calibrated intensity (top panel) and spectral (bottom panel) responses for the second window of the first detector pixel at five different flux levels, which are reported as a percentage of the maximum count rate of 10^6 counts per second. The dashed line in the top panel indicates the identity function, where incident and measure transmission fractions are equal.

The calibration of all 64 pixels and five flux settings is performed with the same model parameter settings; namely the polynomial orders are selected as $N_\alpha = 3$ and $N_\beta = 10$.

For the first set of results, we focus on the second energy window of the first detector pixel. In the top panel of Fig. 1, the non-linear mapping of incident transmission fraction to the measured transmission fraction is plotted for the five available flux levels. Note that the non-linearity of this mapping becomes more pronounced as the flux increases as might be expected. The corresponding spectral response $s_{w=2}$ is shown in the bottom panel of Fig. 1. Interestingly, the overall spectral response remains relatively independent of the flux level.

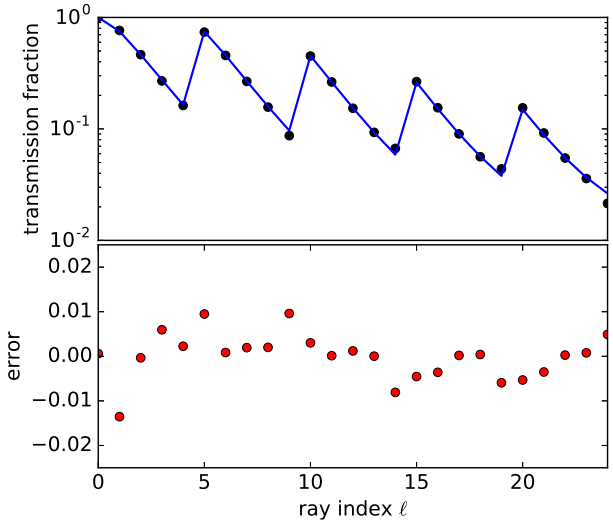


Fig. 2. Top panel: comparison of the transmission fraction measurements (black dots) and fit transmission fraction model (blue curve) for the second energy bin of the first detector pixel. Bottom panel: difference between the transmission measurements and model fit (red dots). The results are shown for the highest flux level of 54% of the maximum count rate.

The ability of the proposed transmission fraction model to match the calibration transmission data is displated in Fig. 2. Again, we show results for the second energy window of the first detector pixel. The transmission data are plotted on a semi-logarithm plot in order to capture the large range in variation of the transmission measurements. The results are shown for the highest flux setting to show that the fitting can be accurate when the non-linearity of the detector intensity response is most pronounced.

For an independent test of the calibration procedure, we used the calibration obtained with the proposed model and PMMA-Aluminum step wedge phantom to predict the transmission measurements through a 1 inch thick block of Teflon. Figure 3 shows this prediction and compares it with actual measurements of the transmission fraction through the Teflon block. We can see that the agreement is quite close for all detector pixels and energy windows. We also note that despite the non-convexity of the transmission model, the fitting is quite robust as it performs equally well over all the detector pixels without changing parameters of the model.

IV. Conclusion

We have developed a calibration procedure for photon counting detectors that calibrates all energy windows of a given detector pixel simultaneously.

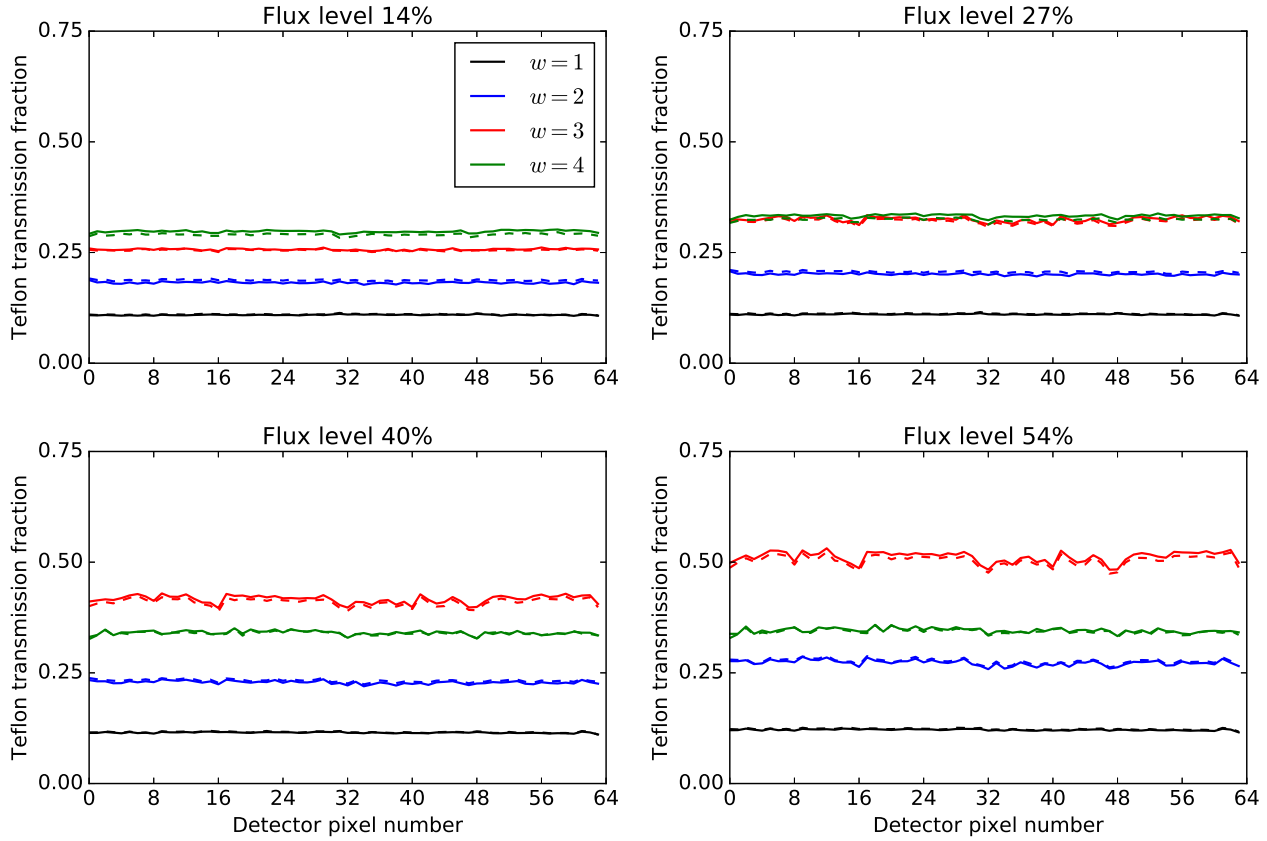


Fig. 3. Comparison of measured X-ray transmission through a 2.54cm slab of Teflon and the transmission fraction modeled using the calibrated spectrum and intensity response function. The model is indicated by the solid curves, and the corresponding measurements are displayed with dashed curves of the same color. Results for all 64 detector pixels and 4 energy windows are show at four of the five flux levels. The flux levels are reported as a percentage of the maximum count rate of 10^6 counts per second. Results for the 9% flux level are not shown to save space. They are similar to that of the 14% flux level.

The calibration model allows for non-linear mapping of the incident and measured intensity, which is necessary for acquisitions that involve count rates that are a significant fraction of the detector maximum count rate. The model is robust as it is able to fit all detector pixels without hand-tuning optimization parameters, and it appears to be able to predict transmission fractions for a material not in the calibration phantom quite accurately. At the meeting we will show results for spectra CT image reconstruction using data obtained by the presented calibration procedure.

V. Acknowledgment

This work is supported in part by NIH Grant Nos. R01-EB023968, and R01-EB026282. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G G. Zeitler, A. Livne, and R. Proksa, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys. Med. Biol.*, vol. 53, no. 15, pp. 4031–4048, 2008.

[2] E. Y. Sidky, L. Yu, X. Pan, Y. Zou, and M. Vannier, "A robust method of x-ray source spectrum estimation from transmission measurements: Demonstrated on computer simulated, scatter-free transmission data," *J. Applied Phys.*, vol. 97, no. 12, pp. 124701, 2005.

[3] W. Ha, E. Y. Sidky, R. F. Barber, T. G. Schmidt, and X. Pan, "Estimating the spectrum in computed tomography via kullback–leibler divergence constrained optimization," *Med. Phys.*, vol. 46, pp. 81–92, 2019.

[4] K. Taguchi, E. C. Frey, X. Wang, J. S. Iwanczyk, and W. C. Barber, "An analytical model of the effects of pulse pileup on the energy spectrum recorded by energy resolved photon counting x-ray detectors," *Med. Phys.*, vol. 37, pp. 3957–3969, 2010.

[5] J. Dickmann, J. Maier, S. Sawall, T. Thüring, S. Gkoumas, C. Brönnimann, and M. Kachelrieß, "A count rate-dependent method for spectral distortion correction in photon counting CT," in *2018 SPIE Proceedings: Physics of Medical Imaging*, 2018, vol. 10573, p. 1057311.

ROI CT reconstruction combining analytic inversion of the finite Hilbert transform and SVD

Aurélien Coussat*, Simon Rit*, Rolf Clackdoyle†, Michel Defrise‡, Laurent Desbat†, Jean Michel Létang*

***Abstract*—In computed tomography, a scan of the whole object may be impossible, leading to truncated projection data. Using differentiated backprojection, the reconstruction problem can be reduced to a set of independent and one-dimensional Hilbert transforms to invert. Depending on the truncation pattern, this inversion problem can either be "zero-", "one-" or "two-endpoint". The zero-endpoint case is known as the interior problem: the field-of-view is completely contained in the object and the reconstruction problem has no unique solution. The two-endpoint case possesses an analytic, numerically stable inverse. The one-endpoint case has a unique and mathematically stable inverse, but no analytic formula for its inverse has been derived so far. A field-of-view (FOV) which is not interior generally contains both one- and two-endpoint sub-regions and we propose here to combine them by using the analytic two-endpoint reconstruction as additional knowledge for the one-endpoint inversion in the rest of the FOV. We hence obtain two reconstructed regions, which we chose to slightly overlap to partially correct for a small residual error appearing in the one-endpoint reconstructions.**

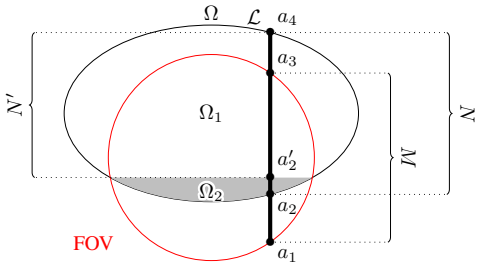


Fig. 1: Diagram summarizing our reconstruction set-up. The shape Ω is the object's convex hull. The gray area Ω_2 is the region that can be reconstructed using the two-endpoint Hilbert inversion formula [2]. The pair (a_1, a_3) (respectively (a_2, a_4)) bounds the FOV (respectively Ω) along the segment $\mathcal{L}$ (drawn here in the particular case where the Hilbert direction θ is equal to 0). The region Ω_1 is $\Omega \setminus \Omega_2$. We are only interested in reconstruction in FOV ∩ Ω.

I. Introduction

X-ray tomography conventionally requires the irradiation of an entire transaxial slice of the object to obtain an image of satisfactory quality. However, it can be undesirable or impossible to perform such an irradiation, and only a sub-region of the slice is irradiated. Conventional reconstruction approaches then become ineffective, but recent theoretical results show that accurate region-of-interest (ROI) reconstruction is possible for some patterns of projection truncation [1].

The context of this work is the two-dimensional (2D) parallel geometry, even though the results can possibly be generalized to other geometries. We assume that the convex hull of the object (denoted as Ω) is known. We define the field-of-view (FOV) of the scanner as the region where each point is irradiated over 180°. Any line segment contained in the union of the scanner FOV and Ω but whose two endpoints lie outside Ω has an analytic reconstruction formula [2] that we name the "two-endpoint Hilbert inversion". The region inside Ω that can be reconstructed with this approach is denoted Ω_2 (Fig. 1). Line segments which have a single endpoint lying outside Ω, such as $[a_1; a_4]$ in Fig. 1, admit a unique and mathematically stable reconstruction along the part of the line segment inside the FOV [3], although no analytic formula has yet been derived for it. Reconstruction along these segments allows reconstruction of the complete FOV for the configuration of Fig. 1. Called "one-endpoint Hilbert inversion", this one-dimensional reconstruction problem can be solved numerically by several techniques, such as iterative methods, e.g. projection onto convex sets [3], or singular value decomposition (SVD). Our technique for one-endpoint Hilbert inversion is partially based on SVD.

A standard regularization of SVD, truncated SVD (TSVD), was extended to include support information and applied to the one-endpoint problem in a previous work [4]. Here, we push this improvement further by first applying the two-endpoint inversion formula where possible, then incorporating some knowledge about this reconstruction into the previous "extended SVD" (XSVD) inversion. Finally, the two-endpoint reconstruction is used again to estimate and partially correct for a small residual error inherent to the one-endpoint problem.

II. Materials and Methods

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the 2D unknown function we want to recover from the measurements $p : [0; \pi[\times\mathbb{R} \to \mathbb{R}$ defined as

$$p(\phi, r) = p_\phi(r) = \int_{-\infty}^{+\infty} f\left(r\boldsymbol{\alpha}_\phi + s\boldsymbol{\beta}_\phi\right) \mathrm{d}s \qquad (1)$$

where $\boldsymbol{\alpha}_\phi = (\cos\phi, \sin\phi)$ and $\boldsymbol{\beta}_\phi = (-\sin\phi, \cos\phi)$. Here, in the context of ROI tomography, we focus on recovering f from p when the values of r span less than the extent of f, i.e., when the projections of f are truncated.

This work was partially supported by grant ANR-17-CE19-0006 (ROIdoré project) from the Agence Nationale de la Recherche (France).

*Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69373, Lyon, France
†Univ. Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, 38000 Grenoble, France
‡Department of Nuclear Medicine, Vrije Universiteit Brussel, Brussels, Belgium

The reconstruction technique introduced in this article is based on the link between the backprojection of the derivative of the measurements and the unknown function f [2]:

$$-\frac{1}{2\pi}\int_{\theta}^{\theta+\pi}\left.\frac{\partial p\left(\phi,r\right)}{\partial r}\right|_{r=\mathbf{x}\cdot\boldsymbol{\alpha}_{\phi}}\mathrm{d}\phi = H_{\theta}f(\mathbf{x}) \tag{2}$$

where $\mathbf{x} = (x_1, x_2)$ and $H_\theta f$ represents the directional Hilbert transform of f in the direction of the vector $\boldsymbol{\beta}_\theta = (-\sin\theta, \cos\theta)$:

$$H_{\theta}f(\mathbf{x}) = \fint_{-\infty}^{+\infty}\frac{f\left(\mathbf{x}-t\boldsymbol{\beta}_{\theta}\right)}{\pi t}\,\mathrm{d}t. \tag{3}$$

This procedure, Eq. (2), has been named differentiated backprojection (DBP), and we refer to θ as the Hilbert direction.

We assume that the samples of the DBP image are accessible at unit-spacing along each line segment $\mathcal{L}$, always oriented vertically in this work, as shown in Fig. 1. Along a segment $\mathcal{L}$, let $(a_1, a_3) \in \mathbb{N}^2$ be inner bounds of the FOV and $(a_2, a_4) \in \mathbb{N}^2$ be outer bounds of Ω with $a_1 < a_2 < a_3 < a_4$ for one-endpoint segments. Let $\mathbf{g} \in \mathbb{R}^M$ be the discretized DBP along this segment in the FOV with $M = a_3 - a_1 + 1$, and $\mathbf{f} \in \mathbb{R}^N$ the sought vector of samples of the unknown function with $N = a_4 - a_2 + 1$. By discretizing Eqs. (2) and (3), we obtain the discrete Hilbert transform operator $\mathbf{H}$ defined as the $M \times N$ matrix with coefficients

$$\mathrm{H}_{j-a_1+1,j'-a_2+1} = \frac{1}{\pi}\frac{1}{j-j'+\frac{1}{2}}, \tag{4}$$

$a_1 \le j \le a_3$ and $a_2 \le j' \le a_4$ such that $\mathbf{g} = \mathbf{H}\mathbf{f}$. The Hilbert transform is applied with a half-pixel shift to improve the reconstructed resolution [5].

Performing reconstruction along segment $\mathcal{L}$ can be done by finding a pseudo-inverse $N \times M$ matrix $\mathbf{H}^\dagger$ of $\mathbf{H}$ such that $\mathbf{H}^\dagger\mathbf{g} = \mathbf{f}^\dagger$. This can be achieved by SVD: if $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ with $\mathbf{U}$ a $M \times M$ matrix, $\boldsymbol{\Sigma}$ a $M \times N$ matrix and $\mathbf{V}$ a $N \times N$ matrix, then $\mathbf{H}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^\dagger\mathbf{U}^\top$ with $\boldsymbol{\Sigma}^\dagger \equiv 0$ except for coefficients $\Sigma^\dagger_{k,k} = 1/\Sigma_{k,k}$ for $k \in \big(1, \ldots, \mathrm{rank}(\mathbf{H})\big)$. In practice, this inversion is ill-conditioned, hence its regularization by TSVD: a cutoff k_c is chosen such that $0 < k_c \le \mathrm{rank}(\mathbf{H})$ after which the small and numerically unstable singular values $\sigma_k = \Sigma_{k,k}$ are discarded (assuming that the singular values are ranked in decreasing order). The reconstruction formula is

$$\mathbf{f}^{\text{TSVD}} = \sum_{k=1}^{k_c} c_k\mathbf{v}_k, \quad c_k = \langle\mathbf{g}, \mathbf{u}_k\rangle/\sigma_k \tag{5}$$

with $\mathbf{v}_k$ (respectively $\mathbf{u}_k$) the k^{th} column of $\mathbf{V}$ (respectively $\mathbf{U}$). This formula is applied to each line segment of the FOV measured in the Hilbert direction.

We previously proposed an improved TSVD reconstruction for the one-endpoint problem by incorporating p_θ, the measured projection information in the Hilbert direction [4]. Assume segment $\mathcal{L}$ is at distance $r_\mathcal{L}$ to the origin. We define $\bar{\mathbf{f}} \in \mathbb{R}^N$ as

$$\bar{\mathrm{f}}_i = \frac{1}{N}\,p(\theta, r_\mathcal{L}) \quad \forall i \in (1, \ldots, N)\,. \tag{6}$$

This vector corresponds to the discrete backprojection of a single ray and can be used as a rough estimation of $\mathbf{f}$ in order to approximate discarded singular components of Eq. (5). Equation (5) is therefore extended as follows:

$$\mathbf{f}^{\text{XSVD}} = \mathbf{f}^{\text{TSVD}} + \sum_{k=k_c+1}^{N}\bar{c}_k\mathbf{v}_k \tag{7}$$

with $\bar{c}_k = \langle\bar{\mathbf{f}}, \mathbf{v}_k\rangle$. We refer to Eq. (7) as XSVD standing for extended SVD. Both the TSVD and the XSVD techniques reconstruct the full-FOV. XSVD partially corrects a residual error observed with the TSVD which increases when approaching the border of the FOV due to the poorer stability of the inversion [3].

Region Ω_2 of the FOV can be analytically reconstructed by the stable two-endpoint Hilbert inversion formula [2] (using horizontal Hilbert line-segments in Fig. 1). We define $\Omega_1 = \Omega \setminus \Omega_2$. If a'_2 is the point at the intersection of segment $\mathcal{L}$ with the boundary between Ω_1 and Ω_2 (as in Fig. 1), the segment $[a_2; a'_2[$ can be first reconstructed using two-endpoint Hilbert inversion. Denote $\mathbf{f}^{\Omega_2} \in \mathbb{R}^N$ the vector containing this reconstruction in $[1; a'_2 - a_2]$ and zeros elsewhere. Let $N' = a_4 - a'_2 + 1$ be the length of the unknown segment $[a'_2; a_4]$ and $\bar{\mathbf{f}}' \in \mathbb{R}^{N'}$ such that

$$\bar{\mathrm{f}}'_i = \frac{1}{N'}\left(p(\theta, r_\mathcal{L}) - \sum_{j=1}^{a'_2-a_2}\mathrm{f}^{\Omega_2}_j\right) \quad \forall i \in \big(1, \ldots, N'\big)\,. \tag{8}$$

The estimate $\bar{\mathbf{f}}'$ corresponds to a single backprojection of the integral sum of f in Ω_1 along $\mathcal{L}$. The DBP $\mathbf{g}' \in \mathbb{R}^M$ of the restriction of f to Ω_1 is given by

$$\mathbf{g}' = \mathbf{g} - \mathbf{H}\mathbf{f}^{\Omega_2}. \tag{9}$$

Finally, the $M \times N'$ matrix $\mathbf{H}'$, defined using a'_2 instead of a_2, is decomposed by SVD similarly to $\mathbf{H}$: $\mathbf{H}' = \mathbf{U}'\boldsymbol{\Sigma}'\mathbf{V}'^\top$. The reconstruction using cutoff k'_c can now be defined as

$$\mathrm{f}^{\text{XSVD-2}}_i = \begin{cases}\mathrm{f}^{\Omega_2}_i & 1 \le i \le N - N' \\ \mathrm{f}^{\Omega_1}_{i-(N-N')} & N - N' + 1 \le i \le N\end{cases} \tag{10}$$

where

$$\mathbf{f}^{\Omega_1} = \mathbf{f}^{\text{TSVD-1}} + \sum_{k=k'_c+1}^{N'}\bar{c}'_k\mathbf{v}'_k, \quad \bar{c}'_k = \langle\bar{\mathbf{f}}', \mathbf{v}'_k\rangle \tag{11}$$

and where $\mathbf{f}^{\text{TSVD-1}} \in \mathbb{R}^{N'}$, given by

$$\mathbf{f}^{\text{TSVD-1}} = \sum_{k=1}^{k'_c} c'_k\mathbf{v}'_k, \quad c'_k = \langle\mathbf{g}', \mathbf{u}'_k\rangle/\sigma'_k. \tag{12}$$

Finally, the TSVD procedure for reconstruction in Ω is

$$\mathrm{f}^{\text{TSVD-2}}_i = \begin{cases}\mathrm{f}^{\Omega_2}_i & 1 \le i \le N - N' \\ \mathrm{f}^{\text{TSVD-1}}_{i-(N-N')} & N - N' + 1 \le i \le N\end{cases}. \tag{13}$$

Conceptually, these new procedures consist of "removing" the data corresponding to f^{Ω_2} from the one-endpoint problem, and reconstructing an object with support Ω_1 using Eq. (7) or (5). This reconstruction can then be combined with the two-endpoint reconstructed region Ω_2 to integrally reconstruct the full FOV.

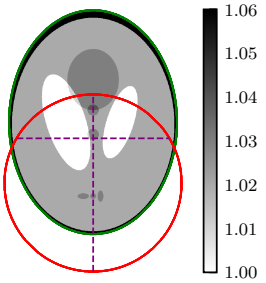


Fig. 2: Illustration of the simulation. The green oval represents the object's convex hull Ω and the red circle the scanner's FOV. The purple dashed lines are the profiles considered in Fig. 4.

TABLE I: Reconstructions considered in our simulations.

Reconstruction name	Two-endpoint reconstruction of Ω_2	Boundary correction	SVD extension (Eq. (7))
TSVD	No	N/A	No
TSVD-2	Yes	No	No
TSVD-2b	Yes	Yes	No
XSVD	No	N/A	Yes
XSVD-2	Yes	No	Yes
XSVD-2b	Yes	Yes	Yes

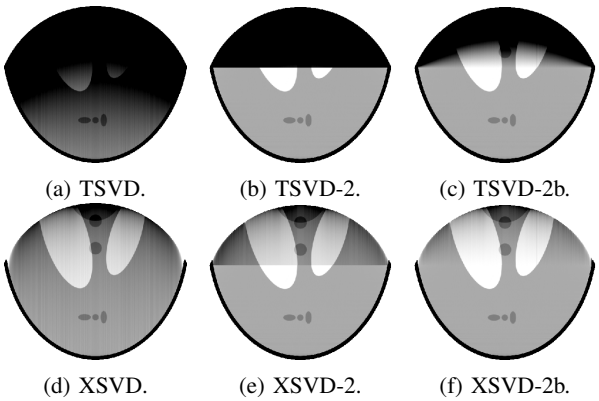
The XSVD-2 reconstruction can be further improved by enforcing continuity between the reconstructions at $a_2' - 1$ (two-endpoint Hilbert inversion) and at a_2' (one-endpoint Hilbert inversion). Our approach was to reconstruct $N' + 1$ samples using SVD with an additional unknown sample corresponding to location $a_2' - 1$ along $\mathcal{L}$ and an amended definition of $\mathbf{g}'$, and to add the difference ϵ between two-endpoint and SVD reconstructions at $a_2' - 1$ to the SVD reconstruction. This correction can be applied to both $\mathbf{f}^{\text{TSVD-2}}$ and $\mathbf{f}^{\text{XSVD-2}}$ and we named the new reconstructions $\mathbf{f}^{\text{TSVD-2b}}$ and $\mathbf{f}^{\text{XSVD-2b}}$.

The proposed methods were evaluated using computer simulations of the 2D Shepp-Logan phantom. A set of 720 parallel projections of 560 rays each were analytically computed over an arc of 180°. The pixel spacing of the projections and the reconstructions was approximately $0.26\,\text{mm}$. The set-up is illustrated in Fig. 2. The cutoffs k_c and k_c' were set to K and K', respectively, where $K = a_3 - a_2 + 1$ and $K' = a_3 - a_2' + 1$, as justified in a previous work [4]. Reconstructions were performed on a 1024×1024 pixel grid following various scenarios summarized in Table I. The Hilbert direction was chosen horizontal for the two-endpoint reconstruction, allowing the segment's two endpoints to lie outside the object, and vertical for the one-endpoint, so that in both cases the Hilbert line-segments were aligned with the pixel grid.

All simulations were implemented using Python 3.7.2, NumPy 1.15.4 and RTK 2.1.0 [6].

III. Results

Figure 3 shows the reconstruction results from the six scenarios of Table I. In Figs. 3a and 3d, reconstructions are entirely performed through one-endpoint Hilbert inversion (Eqs. (5) and (7)). Figures 3b and 3e show the two reconstructions without the Ω_2 to Ω_1 boundary correction at a_2', hence the horizontal transition artifact. In Figs. 3c and 3f, the one-endpoint reconstructions have been adjusted, enforcing vertical continuity of the reconstructions. Fig. 4 shows a vertical and a horizontal profile taken in the central column and above the a_2' threshold, respectively. In Figs. 3b, 3c, 3e and 3f, the two-endpoint reconstructions in Ω_2 are identical.



(a) TSVD. (b) TSVD-2. (c) TSVD-2b.

(d) XSVD. (e) XSVD-2. (f) XSVD-2b.

Fig. 3: Six different reconstructions of Fig. 2 from the scenarios of Table I. The grayscale is $[1; 1.06]$ as in Fig. 2.

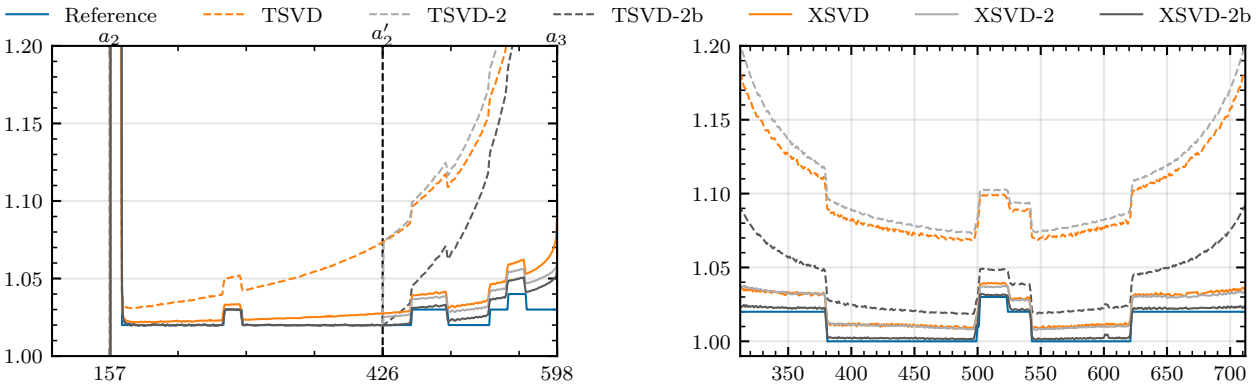
IV. Discussion and Conclusions

Figures 3a, 3d, and Fig. 4a show that in the one-endpoint setup, an offset in the reconstruction values occurs, with increasing intensity towards a_3, the "inner" boundary of the FOV along each vertical column. The effect is smaller than appears because both Figs. 3 and 4 display only a small range of intensity values, respectively 3% and 10% of the full range of values $[0, 2]$ in the phantom. This offset feature is attributed to the rapidly decreasing numerical stability known to occur in the one-endpoint Hilbert inversion model [3]. The fact that this same feature can be observed in other works using other approaches than SVD inversion [3], [7]–[9] suggests that the effect is intrinsic to this tomographic problem and not specific to the one-endpoint Hilbert inversion.

The main idea of the work here is to apply the stable two-endpoint method where possible, thereby reducing the load on the one-endpoint reconstructions. As expected, in the region Ω_2 where two-endpoint Hilbert inversion was applied, excellent quality reconstructions were observed; see Figs. 3b and 3e and especially Fig. 4a in the Ω_2 range $[a_2, a_2']$. The one-endpoint method now only needs to be applied starting at a_2' instead of a_2, resulting in a decrease from N unknowns to N'. However, the improvement was disappointingly small there, as seen in Figs. 3b, 3e and 4.

When the two-endpoint and one-endpoint reconstructions were combined, a visible line of discontinuity appeared in the images; see Figs. 3b and 3e. Eliminating this discontinuity by adding a constant (ϵ) to the one-endpoint region $[a_2', a_3]$ also had the benefit of reducing the offset feature, especially in the TSVD case; see Figs. 3c, 3f and 4.

To understand the disappointing outcome in the reduced region one-endpoint reconstruction, we examined the singular vectors $\mathbf{u}_k$ and $\mathbf{v}_k$ for the central vertical column ($x_1 = 512$) and for $k = K$ and $k = K + 1$ which are respectively the last components included, and first component excluded in series (Eq. (5)) used to build the reconstruction; see Figs. 5a and 5b.



(a) Vertical profile ($x_1 = 512$ pixels). (b) Horizontal profile ($x_2 = 460$ pixels).

Fig. 4: Two profiles of the purple dashed lines of Fig. 2 comparing the different reconstructions described in Table I.

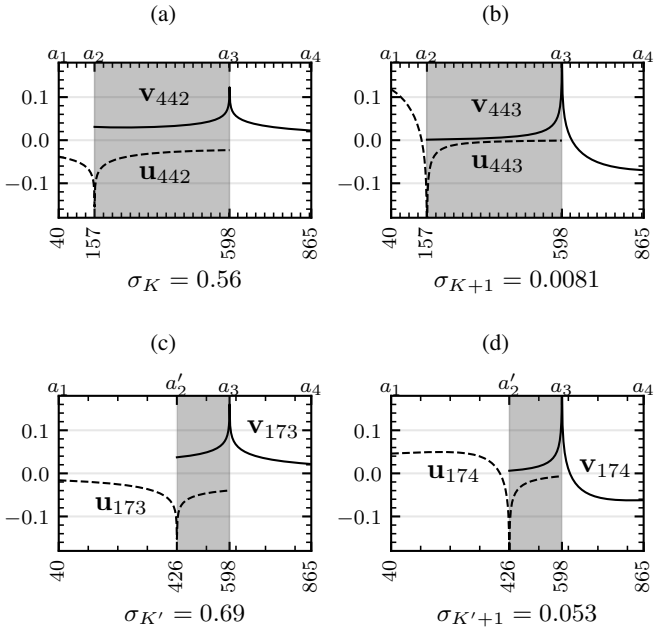


Fig. 5: Singular vectors (Eqs. (5) and (12)), mapped onto the spatial domain in the geometry of the central column of the Shepp-Logan phantom as in Fig. 2. Here, $(a_1, a_2, a_2', a_3, a_4) = (40, 157, 426, 598, 865)$ thus $K = 442$ and $K' = 173$. (Note that K and K' are also the number of elements in the grey region.)

We concentrate on $\mathbf{v}_{442}$ and $\mathbf{v}_{443}$, recalling that the desired function $\mathbf{f}$ is expressed as the series of 709 terms with basis vectors $\mathbf{v}_1$ to $\mathbf{v}_{709}$. We noted in [4] that the lost terms from $k = 443$ to $k = 709$ corresponded to vectors $\mathbf{v}_k$ which were very close to zero in the relevant (grey) region $[a_2, a_3]$ (inside the FOV). However, the first missing component, $\mathbf{v}_{443}$, when approaching a_3 inside the grey region, behaves very similarly to the offset feature in the one-endpoint reconstructions, so the absence of this term is the likely primary cause of this offset feature.

The same arguments apply for the reduced one-endpoint inversion, where the singular vectors shown in Figs. 5c and 5d are remarkably similar to those in Figs. 5a and 5b despite the substantially different number of unknowns ($N' = 440$ instead of $N = 709$) and different SVD cutoff indices ($K' = 173$ instead of $K = 442$). Surprisingly, the horizontal stretching and shrinking of the singular vectors did not appreciably change the shape of $\mathbf{v}_{174}$ ($K'+1$) compared to $\mathbf{v}_{443}$ ($K+1$) near a_3, which could explain the disappointingly small benefit of the reduced problem size.

In conclusion, we have proposed a method for region-of-interest reconstruction from truncated projections using Hilbert inversion methods. The two-endpoint inversion is used where possible to obtain accurate numerically stable reconstructions. The one-endpoint inversion is used elsewhere, and the characteristic offset feature has been reduced by a combination of a priori knowledge (described in [4]) and by matching reconstructed values at the boundary with the two-endpoint region. The resulting reconstructions, Fig. 3f, illustrate dramatic improvements compared to a direct TSVD approach.

References

[1] R. Clackdoyle and M. Defrise, "Tomographic Reconstruction in the 21st Century," *IEEE Sig. Proc. Mag.*, vol. 27, no. 4, pp. 60–80, Jul. 2010.

[2] F. Noo, R. Clackdoyle, and J. Pack, "A two-step Hilbert transform method for 2D image reconstruction," *Phys. Med. Biol.*, vol. 49, no. 17, pp. 3903–3923, Sep. 2004.

[3] M. Defrise, F. Noo, R. Clackdoyle, and H. Kudo, "Truncated Hilbert transform and image reconstruction from limited tomographic data," *Inverse Problems*, vol. 22, no. 3, pp. 1037–1053, Jun. 2006.

[4] A. Coussat, S. Rit, R. Clackdoyle, M. Defrise, L. Desbat, and J. M. Létang, "Region-of-Interest CT Reconstruction using Singular Value Decomposition of the Truncated Hilbert Transform," *IEEE NSS-MIC, Manchester*, Oct. 2019.

[5] F. Noo, J. Pack, and D. Heuscher, "Exact helical reconstruction using native cone-beam geometries," *Phys. Med. Biol.*, vol. 48, no. 23, pp. 3787–3818, Dec. 2003.

[6] S. Rit, M. Vila Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, and G. Sharp, "The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK)," *Journal of Physics: Conference Series*, vol. 489, p. 012079, Mar. 2014.

[7] R. Clackdoyle, F. Noo, Junyu Guo, and J. Roberts, "Quantitative reconstruction from truncated projections in classical tomography," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 5, pp. 2570–2578, Oct. 2004.

[8] E. Rashed, H. Kudo, and F. Noo, "Iterative Region-of-Interest Reconstruction From Truncated CT Projection Data Under Blind Object Support," in *IEEE NSS-MIC Conf. Rec.* IEEE, Oct. 2009, pp. 321–331.

[9] R. Clackdoyle, F. Noo, F. Momey, L. Desbat, and S. Rit, "Accurate Transaxial Region-of-Interest Reconstruction in Helical CT?" *IEEE Trans. Rad. Plasma Med. Sci.*, vol. 1, no. 4, pp. 334–345, Jul. 2017.

Automatic geometric calibration in 3D cone-beam geometry with sources on a line

Hung Nguyen, Laurent Desbat and Rolf Clackdoyle.

***Abstract*—In this paper, we consider the geometric calibration problem in 3D cone-beam geometry with sources on a line, as a series of many 2D calibration problems in fan-beam geometry. We give a method to identify the unknown source positions, directly from the projection data, knowing exactly three source positions, even though the detector positions each have unknown offsets. This method works with general objects, meaning they do not have symmetries. We recall two key results for our method: firstly, data consistency conditions for fan-beam geometry with sources on a line and secondly, the link between the center of mass (COM) of the projections and the projections of the COM of a weighted version of the object. We present numerical experiments based on random unit-density balls. These preliminary experiments show that the proposed method solves the source-position self-calibration problem in 3D cone-beam tomography with sources on a line, if three source positions are known.**

I. Introduction

Geometric calibration has always been of interest in medical imaging, because it is necessary to align the projection images before performing image reconstruction. One idea is to align all projection images according to their centers of mass. In parallel geometry, a well-known result can be stated that: the center of mass of a projection is the projection of the center of mass of the object. We can prove this result by simply considering projection moments or Helgason-Ludwig data consistency consitions (DCCs) of order 0 and 1, see [1], [2], [3]. In [4], Desbat and Clackdoyle produced a very similar result in fan-beam geometry: the center of mass of a fan-beam projection is now the fan-beam projection of the center of mass of a weighted version of the object. Moreover, in fan-beam geometry, if we define the moments of the projections, we can also obtain polynomial-type DCCs. With the same techniques, using the DCCs of order 0 and 1, we can prove the above result of the center of mass.

Projection moments (yielding the Helgason-Ludwig DCCs) have been used by Basu and Bresler for geometric self-calibration in 2D parallel tomography, see [5], [6]. Recent works proposed optimisation methods for geometric self-calibration in 3D cone-beam tomography also by using DCCs, see [7], [8]. In our work, we want to provide analytic formulae based on DCCs for the source positions in cone-beam with sources on a line. The moments and polynomial-type DCCs for fan-beam and cone-beam projections (with the sources on a line or a plane) have been summarized and presented in the context of 2D and 3D geometry in [9]. Furthermore, DCCs for fan-beam projections with the sources on a line have also been carefully studied by Clackdoyle in [10]. He gave two main DCCs for projections, when he considered one fan-beam projection in two different forms: either as an angular ray variable function or as a linear ray variable function. In our paper, we would like to make use of the DCCs for linear ray variable fan-beam projections to deal with the 3D cone-beam geometric calibration problem with the sources on a line. Our aim is to develop an analytic approach to identify the unknown source positions.

Physically, we are considering a horizontal conveyor belt containing our 3D object. It is placed between an x-ray source and a detector. The conveyor belt and the detector are moving together in a certain horizontal direction but at different speeds, to ensure that the projected object remains fairly central on the detector, while the x-ray source is fixed at its own position, see fig. 1. We assume that the detector is sufficiently large, so that none of the projections are truncated. This motion is equivalent to the 3D object being fixed, while the detector and the x-ray source are moving correspondingly to each other in two opposite directions, see fig. 2 (with the motion direction of the detector remaining unchanged compared to the previous model). The x-ray source emits the cone-beam through the 3D object and data are recorded on the detector as the projections. We assume that for some reason, after the first three projections, the system starts losing its track and this leads to the fact that among all of the projections, we only know exactly three relative positions of the x-ray source compared to the 3D object. Thus, we investigate an analytic approach to calibrate all of the others relative positions of the x-ray source.

Our mathematical model considers the 3D object being fixed at a certain location and that an x-ray source is moving along a line, producing a collection of many cone-beam projections of the 3D object, which have been taken on a detector plane parallel to the source line. If we know three source positions corresponding to three acquired projections, then our results claim that we can identify all of the others source positions indepentdently, provided the 3D object is general (with no particular symmetries).

II. 2D Model and Basic Materials

A. 2D model and definitions

In this section, we recall the classic 2D fan-beam problem with sources on a line. We have the standard 2D coordinate

This work was supported by the "Fonds unique interministériel" (FUI) and the European Union FEDER in Auvergne Rhône Alpes (3D4Carm project) and by the ANR (ROIdoré project ANR-17-CE19-0006-01)

H. Nguyen, L. Desbat and R. Clackdoyle are with Univ. Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, 38000 Grenoble, France. e-mail: Hung.Nguyen@univ-grenoble-alpes.fr.

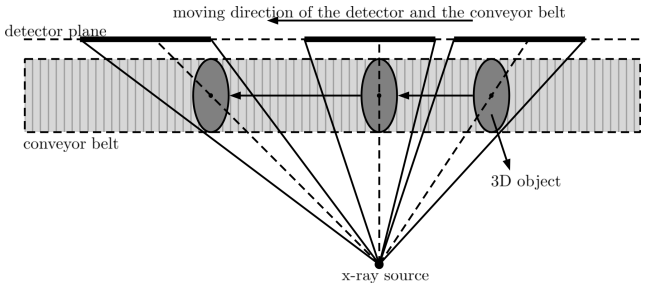


Fig. 1: Top view of the physical model: the x-ray source is fixed, the 3D object and the detector are moving respectively together in the same direction. The detector and conveyor movements are irregular and unknown, and different from each other. But the center of the object projects roughly to the center of the detector at all times, to avoid truncation of the projections.

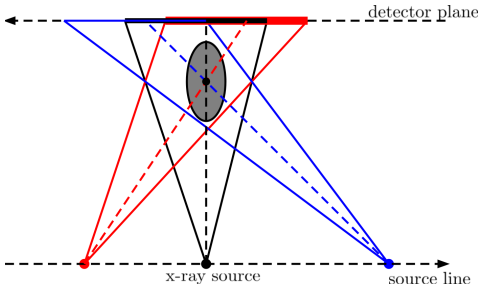


Fig. 2: Top view of the equivalent mathematical model: the 3D object is fixed, the x-ray source and the detector are moving respectively together in opposite directions.

system (x, y). The source is on the x-axis and the detector line is $y = D$ (with D is a positive constant). We call this detector line: the u-axis, which has the origin located at $(0, D)$, see Fig. 3. Our 2D object $f(x, y)$ is between the source line $y = 0$ and the detector line $y = D$.

For each u on detector line, the fan-beam projection with source at $(x, 0)$ is defined as:

$$\begin{aligned} p_x(u) &= \int f\left((x,0) + t\left[(u,D) - (x,0)\right]\right) dt \\ &= \int f\left(x + tu - tx, tD\right) dt. \end{aligned} \tag{1}$$

We also define the moment: for the source position $x \in \mathbb{R}$, the moment of projection p_x is: $M_n(p_x, x) = \int p_x(u) u^n du$.

B. Moment conditions (DCCs)

Using the definitions and a simple change of variables, it can be proven (see [10]) that

$$M_n(p_x, x) = \iint \frac{f(y_1, y_2)}{y_2^{n+1}} \left[Dy_1 + (y_2 - D)\, x\right]^n dy_1 dy_2.$$

Thus M_n is a polynomial of degree n in x, where x stands for the source position.

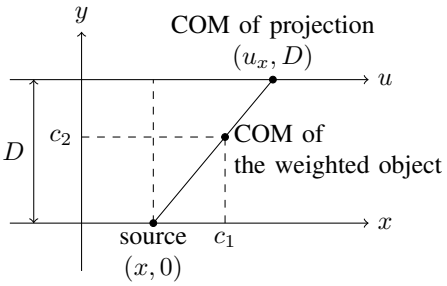


Fig. 3: The COM result can also be proven geometrically by using basic properties of congruent triangles.

C. Center of mass result

We call u_x the position of the center of mass (COM) of the projection associated to source position x. As mentioned in the introduction, using the above moment conditions of order 0 and 1, we can prove that:

$$u_x = x + \frac{D(c_1 - x)}{c_2}, \tag{2}$$

with $c_1 = C_{1,0}/C_{0,0}$ and $c_2 = C_{0,1}/C_{0,0}$ where

$$\begin{aligned} C_{0,0} &= \iint \frac{f(x_1, x_2)}{x_2^2} dx_1 dx_2, \\ C_{1,0} &= \iint \frac{f(x_1, x_2)\, x_1}{x_2^2} dx_1 dx_2, \\ C_{0,1} &= \iint \frac{f(x_1, x_2)\, x_2}{x_2^2} dx_1 dx_2. \end{aligned}$$

From the definitions above, (c_1, c_2) is the COM of the weighted object $g(x, y) = f(x, y)/y^2$. Eq. (2) implies that the source $(x, 0)$, the COM of the weighted object (c_1, c_2) and the COM of the projection (u_x, D) are collinear, meaning *the COM of the projection is the fan-beam projection of the COM of the weighted object*. Fig. 3 provides an illustration.

III. Fan-beam Calibration Problem and Algorithm

We assume that there exist unknown offsets in the detector positions, and the given data is the raw measurement functions. Our first step is to compute the COM of each raw function; then we shift these functions, so that the COM of each raw function is shifted to the origin of the detector. After this step, we obtain the COM-corrected measurement functions $m_x(u)$. Since $m_x(u)$ is also the corresponding projection to source position x, but has the COM of projection located at the origin of the detector, $m_x(u)$ is related to the theoretical projection $p_x(u)$ (defined at Eq. (1)) by the following formula:

$$m_x(u) = p_x(u + u_x), \tag{3}$$

where u_x is position of the COM of projection and is unknown. With the measurements $m_{x_i}(u)$, $i = 1, 2, 3, 4$ and three known source positions x_1, x_2, x_3, we are going to calibrate the unknown position x_4. All other unknown source positions are calibrated the same way.

Solving algorithm

From the four COM-corrected measurements, we compute their moments of order 2:

$$M_2(m_x, x) = \int m_x(u)u^2 du = \int p_x(u)(u - u_x)^2 du.$$

Applying the COM result (2), we can show that $M_2(m_x, x)$ is also a second-degree polynomial in x

$$M_2(m_x, x) = Ax^2 + Bx + C. \quad (4)$$

Thus if x_1, x_2, x_3 are known and two-by-two different, then we have sufficient information to find the three coefficients A, B, C. Since the unknown source position x_4 must satisfy (4), we end up a second-order equation to solve:

$$Ax_4^2 + Bx_4 + C = M_2(m_{x_4}, x_4).$$

In conclusion, we obtain at most two possible solutions of the unknown source position for each fan-beam problem.

IV. Cone-beam Calibration Problem

We now have a standard 3D coordinate system (x, y, z) and a 3D object F lies between the two planes $y = 0$ and $y = 1$. The x-ray source is moving along the x-axis and the detector plane is $y = 1$. The detector plane has its own origin located at position $(0, 1, 0)$. There are also two standard axes u and v on this plane, which respectively have the same directions as x and z-axis. Then for each point on the detector plane, the cone-beam projection is defined as:

$$\begin{aligned} P_x(u, v) &= \int F((x, 0, 0) + t[(u, 1, v) - (x, 0, 0)])\, dt \\ &= \int F(x + tu - tx, t, tv)\, dt. \end{aligned}$$

Fig. 4 illustrates the 3D model being considered.

Given four cone-beam projection images, with three different corresponding source positions known, we will find the other source position.

Solving algorithm

On detector plane, we consider the line $v = v_0$. Since it is parallel to the x-axis there is a plane containing them, which we called the oblique plane $(\Pi_{v_0}) : z = v_0 y$. The origin of (Π_{v_0}) has the same location as the origin of 3D space. The point $(x, y, v_0 y)$ in 3D space will have coordinates $\left(x, y\sqrt{1 + v_0^2}\right)$ on (Π_{v_0}). Assuming (Π_{v_0}) slices through the 3D object, the 2D slice on (Π_{v_0}) is defined as:

$$f_{v_0}\left(x, y\sqrt{1 + v_0^2}\right) = F(x, y, v_0 y).$$

Combining this and the cone-beam projection definition, we can see that for each point on the line $v = v_0$ on the detector

$$\begin{aligned} P_{x,v_0}(u) = P_x(u, v_0) &= \int F(x + tu - tx, t, v_0 t)\, dt \\ &= \int f_{v_0}\left(x + tu - tx, t\sqrt{1 + v_0^2}\right) dt \\ &= p_x(u)|_{D=\sqrt{1+v_0^2}, f=f_{v_0}}, \end{aligned}$$

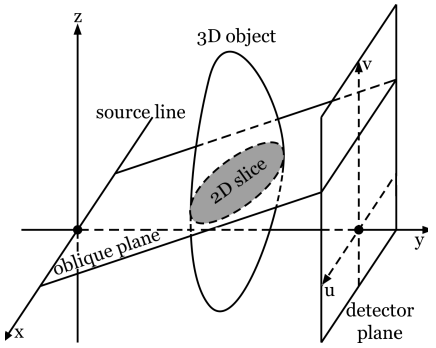


Fig. 4: 3D model in cone-beam geometry

see section II-A. In other words, on (Π_{v_0}), we have exactly the 2D fan-beam calibration problem, where the 2D object f_{v_0} is located between x-axis and the line $y = \sqrt{1 + v_0^2}$.

Therefore, according to section III, each oblique plane (Π_{v_0}) yields at most two possible solutions for the unknown source position. Then N oblique planes being considered will give us roughly N pairs of solutions. If the data are consistent, then these N pairs must have one solution in common, and that common solution is the true solution for the unknown source position. This is our approach to solve a geometric calibration problem in cone-beam with sources on a line; by repeating the same procedure to solve a series of many fan-beam calibration problems and finding the common solution of them.

V. Numerical Simulations

A. Experiment 1: Testing one random object

The 3D object we considered is the superposition of a set of fifteen unit-density balls, where the centers and radii were selected randomly with a uniform distribution. The centers were in the cube $[-0.2, 0.2] \times [0.4, 0.8] \times [-0.2, 0.2]$ and the radii were in the interval $[0.01, 0.15]$. We define

$$G_i = \begin{cases} 1, \text{if } (x - c_x(i))^2 + (y - c_y(i))^2 + (z - c_z(i))^2 \le r^2(i) \\ 0, \text{otherwise} \end{cases}$$

with $i = 1, 2, \ldots, 15$, where $(c_x(i), c_y(i), c_z(i))$ and $r(i)$ are the center coordinates and radius of the ith ball. Then the 3D object is $F = \sum_{i=1}^{15} G_i$. Fig. 5 shows 2D slices obtained from four oblique slice planes through the object.

We have developed two main programs, called *simulator* and *solver*. For each oblique plane (Π_{v_0}), *simulator* receives four numbers: the first three numbers are the known source positions, here $x_1 = -0.228$, $x_2 = -0.475$, $x_3 = 0.069$ and the last one $x_4 = 0.312$ is the unknown source position that we need to calibrate. Then it simulates the projections $p_x(u)$ on 1000 equidistant samples of u in $[-2, 2]$ (see Eq. (1)), and shifts them to their COMs to produce the COM-corrected measurements $m_x(u)$ (see Eq. (3)). After that, *solver* receives these measurements and the three known source positions, then estimates the unknown source position. This procedure is repeated for all the oblique planes. We have taken 200 oblique planes of the form $z = v_0 y$, with v_0 from -0.3 to 0.3. After running the whole program, we plot all these pairs of solutions

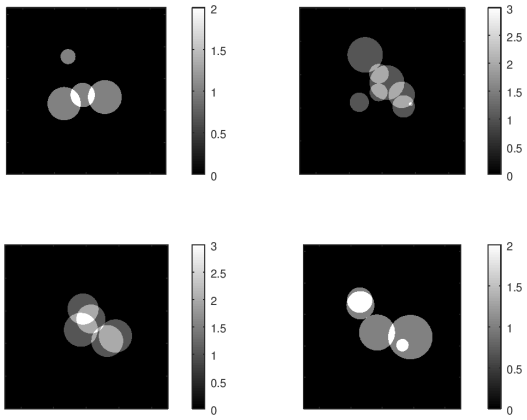


Fig. 5: Four slices of the random object taken from four oblique planes: $z = -0.3y$ (top-left), $z = -0.1y$ (top-right), $z = 0.1y$ (bottom-left) and $z = 0.3y$ (bottom-right)

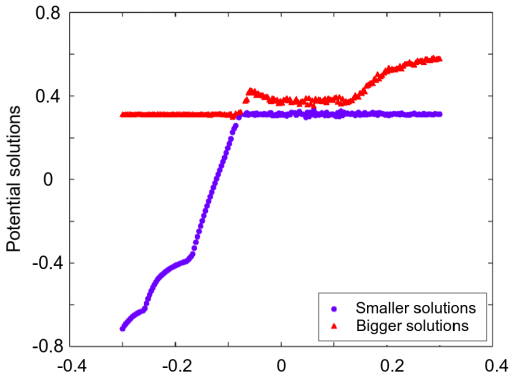


Fig. 6: Solutions on all oblique planes $z = v_0 y$; each value of v_0 (each oblique plane) contain at most two solutions

on a same figure. We presume that there will be a horizontal set of collinear points, where each point is one of each pair of solutions. Fig. 6 shows the result we obtained. Although we can visually identify the correct solution from Fig. 6, we provide the following algorithm to automatically extract the estimated source location.

Computing estimated solution algorithm:

Step 1: We find the minimum solution $x_{\min}$ (the lowest blue point in Fig. 6). We also find the value v_1, where (Π_{v_1}) is the oblique plane containing $x_{\min}$.

Step 2: We find the maximum solution $x_{\max}$ over the other oblique planes: (Π_{v_0}), with $v_0 \neq v_1$. We find v_2 knowing the oblique plane (Π_{v_2}) contains $x_{\max}$.

Step 3: We now have two distinct planes (Π_{v_1}) and (Π_{v_2}). The plane (Π_{v_1}) contains at most two solutions: $x_{\min}$ and $x^{v_1}_{\text{other}}$. The other plane (Π_{v_2}) contains $x_{\max}$ and $x^{v_2}_{\text{other}}$. The points $x_{\min}$ and $x_{\max}$ are different from each other. But there must be two solutions, one on each plane, which are equal or near to each other, because of the fact of the common solution. Thus we compute $H_1 = |x_{\max} - x^{v_1}_{\text{other}}|, H_2 = |x_{\min} - x^{v_2}_{\text{other}}|, H_3 = |x^{v_1}_{\text{other}} - x^{v_2}_{\text{other}}|$ and $H = \min\{H_1, H_2, H_3\}$.

Step 4:

- If $H = H_1$, then $x^{\text{est}} = \left(x_{\max} + x^{v_1}_{\text{other}}\right)/2$.
- If $H = H_2$, then $x^{\text{est}} = \left(x_{\min} + x^{v_2}_{\text{other}}\right)/2$.
- If $H = H_3$, then $x^{\text{est}} = \left(x^{v_1}_{\text{other}} + x^{v_2}_{\text{other}}\right)/2$.

After performing the above algorithm, we obtain the estimated solution $x^{\text{est}} \approx 0.3114$, comparing to the true solution $x^{\text{true}} = 0.312$, with the error $\approx 5.8 \times 10^{-4}$.

B. Experiment 2: Testing 100 random objects

We repeated this test for 100 random objects, where each of them is as the description in experiment 1. We have computed the error vector Err, such that $\text{Err}(i) = x^{\text{est}}(i) - x^{\text{true}}$, $i = 1, 2, \ldots, 100$; and mean $(\text{Err}) = 1/100 \sum_{i=1}^{100} \text{Err}(i) = -3.5 \times 10^{-5}$, is close to 0. We have: $1/100 \sum_{i=1}^{100} |\text{Err}(i)| = 3.8 \times 10^{-3}$, $\sqrt{1/100 \sum_{i=1}^{100} |\text{Err}(i)|^2} = 6.5 \times 10^{-3}$, $\max_i |\text{Err}(i)| = 2.6 \times 10^{-2}$.

VI. Discussions and Conclusions

We have presented a geometric self-calibration method for cone-beam projections with sources on a line. Assuming three source positions are given, we provide the source position for any new cone-beam projection from the projection data only. The method is based on 2D fan-beam data consistency conditions, more precisely on moments of order 0, 1 and 2. The idea is to consider cone-beam projections with sources on a line as a set of independent fan-beam projections in oblique planes (containing the source line). In this geometry, the moments of order 2 of the fan-beam projections, each centered on its center of mass, is a polynomial of order 2 in the source position variable. Thus the source position is a root of a polynomial of order 2. Thanks to many independent fan-beam projection sets (each contained in an oblique slice) of the cone-beam projections with sources on a line, we can easily identify the common root and thus the source position. We have presented preliminary numerical experiments, including statistics based on 100 phantoms of 15 random ball indicators. They show that this approach solves easily and precisely the source position self-calibration problem in cone-beam projections with sources on a line, provided three source positions are known.

References

[1] D. Ludwig, "The radon transform on euclidean space," *Commun. Pure Appl. Math.*, vol. 19, pp. 49–81, 1966.
[2] S. Helgason, *The Radon Transform*. Boston, MA, USA: Birkhaüser, 1980.
[3] F. Natterer, *The Mathematics of Computerized Tomography*. Wiley, 1986.
[4] L. Desbat and R. Clackdoyle, "Calibration and data consistency in parallel and fan-beam linogram geometries," in *IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings*, 2019. to appear.
[5] S. Basu and Y. Bresler, "Uniqueness of tomography with unknown view angles," *ieeeip*, vol. 9, pp. 1094–1106, 2000.
[6] S. Basu and Y. Bresler, "Feasibility of tomography with unknown view angles," *ieeeip*, vol. 9, pp. 1107–1122, 2000.
[7] A. Aichert, M. Berger, J. Wang, N. Maass, A. Doerfler, J. Hornegger, and A. Maier, "Epipolar consistency in transmission imaging," *IEEE Transactions on Medical Imaging*, vol. 34, pp. 2205–2219, Nov 2015.
[8] J. Lesaint, S. Rit, R. Clackdoyle, and L. Desbat, "Calibration for circular cone-beam ct based on consistency conditions," *IEEE Transaction on Radiation and Plasma Medical Sciences*, vol. 1, no. 6, pp. 517–526, 2017.
[9] R.Clackdoyle, "Data consistency for linograms and planograms," *IEEE Trans Rad Plasma Med*, vol. 4, pp. 288–299, 2018.
[10] R. Clackdoyle, "Necessary and sufficient consistency conditions for fan-beam projections along a line," *IEEE Trans.Nucl.Sci.*, vol. 60, pp. 1560–1569, 2013.

A Preliminary Study on the Inversion of DXT from Limited-Angular-Range Data

Zheng Zhang, Buxin Chen, Dan Xia, Emil Y. Sidky, and Xiaochuan Pan

***Abstract*—In this work, we investigate and develop an algorithm for image reconstruction by inverting a fan-beam, discrete-to-discrete X-ray-transform (DXT)-data model from its knowledge available over a limited-angular range considerably smaller than that required by conventional algorithms. Inspired by the recent works, we formulate the reconstruction problem as a convex optimization program with split-image-total-variation (TV) constraints, and subsequently develop an iterative algorithm, referred to as the limited-angle primal-dual (LAPD) algorithm, to reconstruct images through solving the optimization program. We carry out image reconstructions from DXT data generated over limited-angular ranges from numerical phantoms with a spectrum of practical anatomy characteristics to verify numerically that the LAPD algorithm theoretically derived can accurately solve the optimization program. Using the LAPD algorithm verified, we then explore empirically the potential of the LAPD algorithm for numerically accurately reconstructing images, or, equivalently, inverting the DXT-data model, as a function of the angular range only over which data are available. The LAPD algorithm can also be exploited for investigating how various key factors, such as angular range, subject characteristics, and data quality, may impact image reconstruction from data available over limited-angular ranges in CT.**

I. Introduction

In CT imaging, the problem of image reconstruction from data collected over an limited-angular range is of high practical significance and theoretical challenge, as it generally entails the inversion of a mathematically severely ill-conditioned data model, such as the fan-beam, discrete-to-discrete X-ray transform (DXT) data model, from its knowledge available over a limited-angular range. Throughout the summary, for simplicity without ambiguity, we refer to the fan-beam DXT as the DXT, and also use "image reconstruction" and "inversion of the DXT" interchangeably. Direct inversion of the DXT-data model from limited-angular-range data by use of conventional algorithms, e.g., filtered-backprojection (FBP) algorithm or simultaneous algebraic reconstruction technique (SART) [1] can result in images with significant artifacts.

Inspired by the recent works on anisotropic regularization for mitigating artifacts in reconstruction from data over limited-angular ranges [2]–[4], we develop in the work an algorithm to reconstruct an image, or, equivalently, to invert the DXT-data model, from data available over a limited-angular range. Specifically, we formulate the reconstruction problem as an optimization program with a data-ℓ_2-minimization term and two split-image-total-variation (TV) constraints that are the ℓ_1-norms of image gradients along two special orthogonal axes of a 2D image array, respectively. As the optimization program is convex, we develop an iterative algorithm, referred to as the limited-angle primal-dual (LAPD) algorithm, based upon the Chamboll-Pock instance of the general primal-dual algorithm, to reconstruct the image through solving the optimization program [5]–[7] from data available over a limited-angular range.

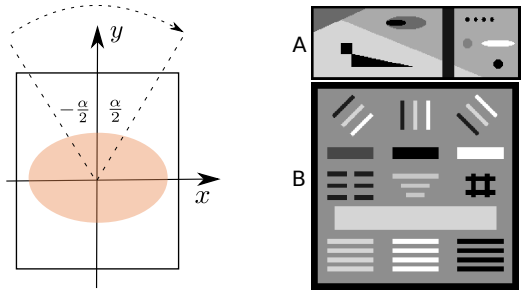


Figure 1. Illustration of a circular fan-beam scan with a limited-angular range (left), and phantoms A (top, right) & B (bottom, right)

With the LAPD algorithm theoretically developed, we carry out numerical studies from data generated from numerical phantoms with a spectrum of practical anatomy characteristics over limited-angular ranges to verify that the LAPD algorithm developed can indeed numerically accurately solve the optimization program. Using the LAPD algorithm verified numerically, we then explore empirically the potential of the LAPD algorithm for image reconstruction by varying the angular range over which data, or, equivalently, knowledge of the DXT, are available. The work yields an algorithm for inverting numerically accurately the DXT-data model from its knowledge over certain limited-angular ranges; and the LAPD algorithm can also be exploited for investigating how various key factors, such as angular range, subject characteristics, and data quality, may impact image reconstruction from limited angle data in CT. The optimization program and LAPD algorithm can readily be extended to invert a cone-beam-DXT-data model.

II. Method

We use vector $\mathbf{f}$ of size N to denote a 2D discrete image in a concatenate form, $N = N_1 \times N_2$, N_1 and N_2 are the total numbers of pixels along two orthogonal directions, and f_i to depict entry i of $\mathbf{f}$ with $i = 1, 2, ..., N$.

Z. Zhang, B. Chen, D. Xia, and E. Y. Sidky are with the Department of Radiology, the University of Chicago.

X. Pan is with Departments of Radiology and Radiation & Cellular Oncology, the University of Chicago.

A. The DXT-data model

The DXT-data model considered in the work is given by

$$\mathbf{g}(\mathbf{f}) = \mathcal{H}\,\mathbf{f}, \tag{1}$$

where vector $\mathbf{g}$ of size M denotes discrete model-data, with entry $g_j(\mathbf{f})$ indicating the model-data value corresponding to ray j, where $j = 1, 2., ..., M$; and $\mathcal{H}$ the system matrix of size $M \times N$, with element $h_{i,j}$ representing an intersection of ray j with pixel i. Image reconstruction considered in the work is equivalent to inverting the DXT-data model in Eq. (1) from limited knowledge of $\mathbf{g}^{(m)}$ available over a limited-angular range of interest.

B. Optimization program with split-TV constraints

For image reconstruction, or, equivalently, for inversion of the DXT-data model in Eq. (1), we first formulate a data-ℓ_2-minimization program with split-TV-constraints [3], [4] as

$$\begin{aligned} \mathbf{f}^{\star} = \underset{\mathbf{f}}{\text{argmin}} \left\{ \frac{1}{2} \parallel \mathbf{g}^{(m)} - \mathbf{g}(\mathbf{f}) \parallel^2 \right\} \\ \text{s.t.} \quad ||\nabla_1 \mathbf{f}||_{\text{TV}} \leq t_1, \; ||\nabla_2 \mathbf{f}||_{\text{TV}} \leq t_2, \text{ and } f_i \geq 0, \end{aligned} \tag{2}$$

where the transposes of matrices ∇_1 and ∇_2 of size $N \times N$ denote two-point differences along the two orthogonal axes, respectively; vectors $\nabla_1 \mathbf{f}$ and $\nabla_2 \mathbf{f}$ are of size N; scalars $||\nabla_1 \mathbf{f}||_{\text{TV}}$ and $||\nabla_2 \mathbf{f}||_{\text{TV}}$ indicate the ℓ_1-norms of gradient images along directions 1 and 2; and parameters t_1 and t_2 depict the upper bounds on the split-TV-constraints.

C. The LAPD algorithm

Because the optimization program in Eq. (2) is convex, iterative algorithms can be developed mathematically for solving the program. In the work, we derive an iterative LAPD algorithm and its convergence conditions. At the meeting, we will present in detail the derivation, pseudo codes, and convergence conditions of the LAPD algorithm.

In the verification or quantitative studies involving numerical phantoms presented in this summary, we consider consistent data, i.e., $\mathbf{g}^{(m)} = \mathbf{g}(\mathbf{f}_{\text{truth}})$, and use

$$\tilde{D}(\mathbf{f}_n) \to 0 \quad \text{as } n \to \infty \tag{3}$$

as a sufficient convergence condition on accurate image reconstruction or, equivalently, on accurate inversion of the DXT-data model, where n is the iteration number,

$$\tilde{D}(\mathbf{f}_n) = ||\mathbf{f}_n - \mathbf{f}_{\text{truth}}||_2 / ||\mathbf{f}_{\text{truth}}||_2, \tag{4}$$

and $\mathbf{f}_{\text{truth}}$ denotes the phantom image.

III. Numerical Verification Studies

We first perform numerical verification studies for three purposes: (1) to verify numerically that the LAPD algorithm theoretically developed and its computer implementation can solve accurately the optimization program in Eq. (2); (2) to numerically explore, with the LAPD algorithm verified, image reconstructions from data available over different angular ranges for phantoms with different levels of anatomic complexity; and (3) to gain empirically insights into the angular ranges for which the LAPD algorithm can numerically accurately yield image reconstruction, or, equivalently, invert the DXT-data model in Eq. (1).

A. Scanning geometry

Without loss of generality, we consider two special orthogonal directions along x- and y-axis of a 2D Cartesian coordinate system and a segment of a circular scanning trajectory covering angular range α symmetric relative to the y-axis, as shown in the left panel of Fig. 1. From each of the phantoms studied, we use the DXT-data model in Eq. (1) to generate consistent data, i.e., $\mathbf{g}^{(m)} = \mathbf{g}(\mathbf{f}_{\text{truth}})$, over angular range α. The fan-beam geometry considered in the work has source-to-rotation-axis and source-to-detector distances of 36 cm and 72 cm. For given α, data are generated at views with an angular interval of 1° by use of a linear detector array consisting of 512 detector bins of 0.71 mm that covers completely the phantom.

B. Data generation

While we have performed studies on an array of phantoms with different levels of anatomy and contrast complexity, we present in the summary the study results of two phantoms displayed in the right panel of Fig. 1, and will report results of additional phantom studies at the conference. In the study, each of both phantoms is placed in the scanning field of view such that their middle horizontal and vertical lines coincide with the x- and y-axis, respectively, of the coordinate system. Since both phantoms are known in the studies, we calculate their respective split TVs along x- and y-axis and use them as parameters t_1 and t_2. (When knowledge of t_1 and t_2 is unknown, one may need to sweep through ranges of t_1 and t_2 to select their values appropriate for a particular study.)

Phantom A in the right panel of Fig 1, containing numerous disks, ellipse, squares, and triangles [3], [4], is on a 256×128 array of square pixels of size 0.7 mm. Three data sets are generated from phantom A over angular ranges $\alpha = 30^\circ, 60^\circ$, and 90°, respectively. Conversely, phantom B in the right panel of Fig 1, including bar-shape anatomies of different orientations and contrast levels, is on a 256×256 array of square pixels of size 0.7 mm. Again, three data sets are generated from phantom B over angular ranges $\alpha = 90^\circ, 120^\circ$, and 160°, respectively.

IV. Study results

We first summarize the verification study that establishes numerically that the LAPD algorithm, and its computer implementation, can solve numerically accurately the optimization program with split-image-TV constraints in Eq. (2), and then discuss the results of studies, enabled by the verified LAPD algorithm, on exploring accurate image reconstruction from data over limited-angular ranges.

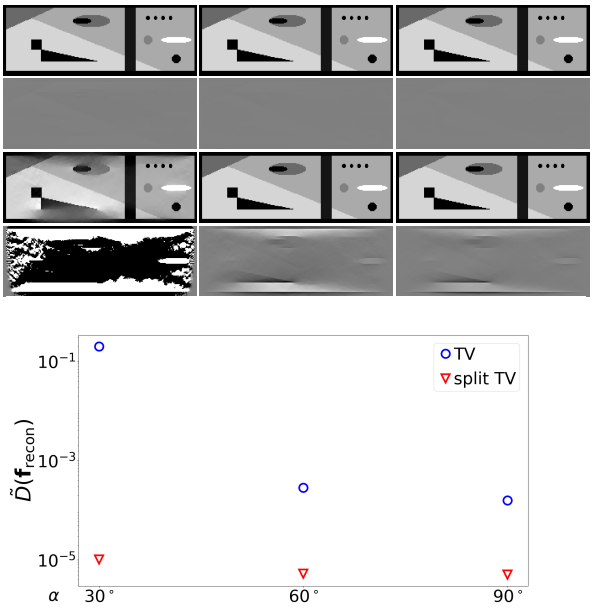


Figure 2. Images of phantom A obtained with split-TV (row 1) and conventional-TV (row 3) constraints, along with their respective differences relative to phantom A in rows 2 & 4, from data over angular ranges $\alpha = 30°$ (column 1), $60°$ (column 2), and $90°$ (column 3), respectively. The display window is [-0.001, 0.001] cm^{-1} for rows 3 & 4. The plot (bottom row) shows normalized image distance $\tilde{D}(\mathbf{f}_{\text{recon}})$ as functions of α for split-TV and conventional-TV reconstructions.

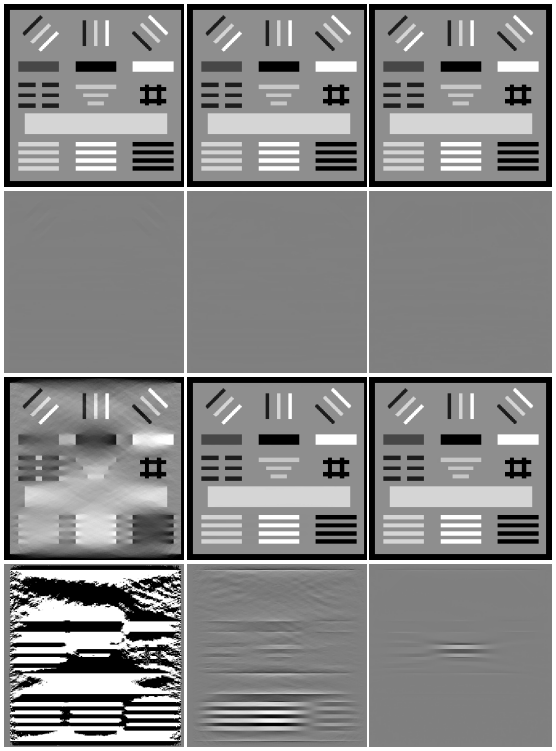


Figure 3. Images of phantom B obtained with split-TV (row 1) and conventional-TV (row 3) constraints, along with their respective differences relative to phantom B in rows 2 & 4, from data over angular ranges $\alpha = 90°$ (column 1), $120°$ (column 2), and $160°$ (column 3), respectively. The display window is [-0.001, 0.001] cm^{-1} for rows 2 & 4.

A. Verification of the LAPD algorithm derived

While the LAPD algorithm derived can theoretically solve the convex optimization program in Eq. (2), we must also establish in numerically studies that the LAPD algorithm can indeed accurately solve the optimization program in Eq. (2) in terms of satisfying the convergence conditions derived for the LAPD algorithm up to the computer precision. The numerical verification studies are absolutely necessary as they also assure that the algorithm is implemented correctly on a computer. We have carried out extensive studies using data sets generated from phantoms over different angular ranges, and the results confirm that for each of the cases studied, the convergence conditions of the LAPD algorithm are satisfied up to computer precision. We will present these extensive study results in detail at the meeting, which complete the evidence that the LAPD algorithm can solve the program in Eq. (2).

With the LAPD algorithm numerically verified, we present below some of the study results exploring how the performance of the LAPD algorithm, in terms of reconstruction visualization and accuracy, may be impacted by limited-angular ranges over which data are available for different phantoms. It should be reminded that each of the reconstructions below with the LAPD algorithm is obtained when the LAPD algorithm achieves its convergence conditions (up to the computer precision,) and that the impact of iteration numbers, if any, on the reconstruction is thus on the order of the computer precision. We use $\mathbf{f}_{\text{recon}}$ to denote the convergent reconstruction.

B. Image reconstruction of phantom A

We reconstruct images of phantom A by applying the LAPD algorithm to its three data sets generated over angular ranges $\alpha = 30°, 60°$, and $90°$, respectively, and display the respective convergent reconstructions $\mathbf{f}_{\text{recon}}$ in row 1 of Fig. 2, along with the differences $(\mathbf{f}_{\text{recon}} - \mathbf{f}_{\text{truth}})$ in row 2 of Fig. 2. From the results in rows 1 and 2 of Fig. 2, it can be observed that for the conditions considered, the LAPD algorithm can accurately reconstruct the images from data over an angular range of $\alpha \geq 30°$. As a reference, we also include reconstructions based upon a data-ℓ_2-norm-minimization with a conventional-TV constraint [5]–[7] from the three data sets and displayed the convergent reconstructions in row 3 of Fig. 2, along with their respective differences from phantom A in row 4 of Fig. 2. The reconstructions with the conventional-TV constraint contain visible artifacts that become more conspicuous as the angular range is reduced, especially for the artifacts that appear around contrast boundaries parallel to the x-axis. Furthermore, $\tilde{D}(\mathbf{f}_{\text{recon}})$, the normalized ℓ_2-norms of the difference images between convergent reconstructions and phantom A in the bottom panel of Fig. 2 corroborates the observation that under the conditions considered, the LAPD algorithm can yields numerically accurate reconstruction or, equivalently, numerically accurately invert the DXT-data model in Eq. (1), from data over an angular range as low as $30°$ for phantom A.

C. Image reconstruction of phantom B

We also reconstruct images of phantom B by applying the LAPD algorithm to its three data sets generated over angular ranges $\alpha = 90^\circ, 120^\circ$, and 160°, respectively, and display the respective convergent reconstructions $\mathbf{f}_{\rm recon}$ in row 1 of Fig. 3, along with the differences $(\mathbf{f}_{\rm recon} - \mathbf{f}_{\rm truth})$ in row 2 of Fig. 3. From the results in rows 1 and 2 of Fig. 3, it can be observed that for the conditions considered, the LAPD algorithm can accurately reconstruct the images from data over an angular range of $\alpha \geq 90^\circ$. As a reference, we also include reconstructions based upon a data-ℓ_2-norm-minimization with a conventional-TV constraint [5]–[7] from the three data sets and displayed the convergent reconstructions in row 3 of Fig. 3, along with their respective differences relative to phantom B in row 4 of Fig. 3. The reconstructions with the conventional-TV constraint contain visible artifacts that become more conspicuous as the angular range is reduced, especially for the artifacts appear around contrast boundaries parallel to the x-axis. For phantom B, we have also obtained results of $\tilde{D}(\mathbf{f}_{\rm recon})$ (not shown here for saving space,) which are similar to that of phantom A in Fig. 2. The reconstructions and $\tilde{D}(\mathbf{f}_{\rm recon})$ obtained demonstrate that under the conditions considered, the LAPD algorithm can yields numerically accurate reconstruction or, equivalently, numerically accurately invert the DXT-data model in Eq. (1), from data over an angular range as low as 90° for phantom B.

V. Discussions

We have investigated image reconstruction or, equivalently, inversion of the DXT-data model from data available only over a limited-angular range. Inspired by the recent work, we formulate the reconstruction problem into a convex optimization program with split-TV constraints and develop a theoretical LAPD algorithm to solve the program. Following extensive verification studies that establish numerically that the LAPD algorithm and its computer implementation can accurately solve the optimization program, we use the LAPD algorithm to explore numerically image reconstructions under a variety of conditions of interests, including limited-angular ranges over which data are generated from numerous phantoms with different anatomies of varying complexity levels. These studies enabled by the LAPD algorithm reveal that under the conditions considered, the LAPD algorithm can yield numerically accurate reconstructions or, equivalently, invert numerically accurately the DXT-data model, from data over angular ranges that are considerably smaller than what conventional algorithms require. We will report in detail the theoretical derivation, pseudo codes, and numerical verification of the LAPD algorithm, along with extensive numerical studies, at the conference.

We have also performed image reconstruction with the LAPD algorithm from limited-angular-range data containing components such as noise that are inconsistent with the DXT-data model and will report the results at the meeting. It should be reminded that the current work focuses only on investigation of inverting accurately the DXT-data model from limited-angular-range data. In the future, it is necessary and interesting to investigate the performance characteristics of the LAPD algorithm when applied to limited-angular-range data containing additional physical factors such as scatter, beam hardening, and focal spot of finite size, which are inconsistent to the DXT-data model. Though the study reported here may offer some insights into the upper bound on the algorithm performance, it remains to be established if the algorithm is any practical value in terms of devising practical procedures of high task utility when applied to data collected in realistic imaging applications. The work reported can readily be extended to investigate image reconstruction from cone-beam data over limited-angular ranges.

VI. Acknowledgment

This work was supported in part by NIH R01 Grant Nos. EB018102 and EB023968 and the Grayson-Jockey Club Research.

References

[1] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," *Ultrason. imaging*, vol. 6, pp. 81–94, 1984.

[2] E. T. Quinto, "Artifacts and visible singularities in limited data X-ray tomography," *Sens. & Imag.*, vol. 18, no. 1, pp. 9–23, 2017.

[3] J. Xu, S. Deng, H. Zhang, Y. Zhu, and P. Zhang, "A second derivative based regularization model for limited-angle computed tomography," *Proc. 5th Intern. Meet. Imag. Form. in X-Ray Comput. Tomogr.*, pp. 256–260, 2018.

[4] J. Xu, Y. Zhao, H. Li, and P. Zhang, "An Image Reconstruction Model Regularized by Edge-preserving Diffusion and Smoothing for Limited-angle Computed Tomography," *Inverse Prob.*, vol. 35, no. 8, pp. 5004–5038, 2019.

[5] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, pp. 1 – 26, 2011.

[6] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm," *Phys. Med. Biol.*, vol. 57, no. 10, pp. 3065–3091, 2012.

[7] Z. Zhang, X. Han, E. Pearson, C. Pelizzari, E. Y. Sidky, and X. Pan, "Artifact reduction in short-scan CBCT by use of optimization-based reconstruction," *Phys. Med. Biol.*, vol. 61, no. 9, pp. 3387–3406, 2016.

Simultaneous Image Reconstruction and Element Decomposition for Iodine Contrast Agent Visualization in Multi-energy Cone Beam CT

Hyunuk Jung, Chenyang Shen, Xun Jia

***Abstract*—Iodine contrast agent is widely used in liver cancer radiotherapy at CT simulation stage to enhance detectability of tumor. However, its application in cone beam CT (CBCT) for image guidance before treatment delivery is still limited because of poor image quality and excessive dose of contrast agent during multiple treatment fractions. In this study, we develop a multi-energy element-resolved (MEER) CBCT framework to enhance detectability of low-concentration iodine contrast agent. We acquired a kVp-switching scan on a conventional CBCT platform with 600 projections in a full gantry rotation, in which the kVp level sequentially changed among 80,100, and 120 kVps. We developed an MEER-CBCT reconstruction algorithm that simultaneously reconstructed x-ray attenuation images at each kVp, an electrong density image, and elemental composition images (H, O, Ca, and I). The composition of each voxel was subject to a constraint of a sparse representation of materials in a dictionary containing typical human tissues and iodine. To evaluate our method, we performed simulation studies and experimental validations using a phantom with inserts containing water and different concentrations of iodine solution. We analyzed iodine detectability as quantified by contrast-to-noise ratio (CNR) and compared results with those of mono-kVp CBCT images reconstructed by the standard filter back projection (FBP) method with 600 projections. MEER-CBCT achieved similar contrast enhancement as FBP method but about 6 times higher CNR. At 4% iodine solution concentration, FBP method achieved ~250 HU enhancement and CNR of ~2.5, considered as the standard CNR for successful tumor visualization. MEER-CBCT achieved the same CNR but at ~8 times lower iodine concentration of 0.5%.**

***Index Terms*—cone beam CT, multi-energy, iodine, contrast agent**

I. Introduction

Cone beam CT (CBCT) is widely used for imaging guidance in radiation therapy. Yet, due to relatively low image quality and soft-tissue contrast, tumor visualization is sometimes challenging, such as for liver cancer, leading to tumor targeting uncertainty. Intravenous contrast-enhancement (IVC) is routinely used in treatment planning CT of liver cancer radiotherapy because of improved tumor contrast. However, it cannot be reliably employed on current CBCT due to poor image quality: IVC-CBCT was found to be effective in only 1/4 of patients with tumor size larger than 120cc under breath-hold scans. Small tumors or free-breathing CBCT do not show contrast enhancement[1]. In-room mobile CT or CT-on-rail can realize IVC scans. The required hardware and cost limited their wide-spread use. Recent advance of MRI-LINAC enables MRI-based image guidance. Yet they are still in the early adoption stage and the cost and hardware complexity are likely to hinder clinical adoption.

Aiming at addressing this problem, we propose in this study a multi-energy element-resolved (MEER) CBCT framework achieved on a standard CBCT platform to enhance detectability of iodine contrast agent at a substantially reduced contrast agent concentration level. By innovatively employing multi-kVp data acquisition, a novel spatial-spectral dimension regularization approach, and a sparse-dictionary based elemental decomposition strategy, we will be able to improve image quality to achieve iodine contrast detection for tumor delineation.

All authors are with innovative Technology Of Radiotherapy Computations and Hardware (iTORCH) Laboratory, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX, 75587, USA. Email: chenyang.shen and xun.jia@utsouthwestern.edu

II. Method

A. Data acquisition

The multi-energy CBCT scan was acquired using a kVp-switching technique realized on a standard CBCT platform available in the on-board imaging system of a Varian medical linear accelerator (LINAC). We programmed the data acquisition protocol via its developer model to allow a CBCT scan with 600 x-ray projections in a full gantry rotation. The kVp levels were controlled to alternate among 80, 100, and 120 kVps in subsequent projections.

B. The proposed model

The general form of MEER-CBCT model can be formulated as follows:

$$\min_{\mathbf{F},\boldsymbol{\rho},\boldsymbol{\lambda}} \frac{1}{2}R(\mathbf{F}) + \frac{\beta}{2}G(\mathbf{F},\boldsymbol{\rho},\boldsymbol{\lambda}). \quad (1)$$

The first term $R(\mathbf{F})$ is established for ME-CBCT reconstruction, where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N] \in \mathbb{R}^{M\times N}$ denotes ME-CBCT images at N different energy channels. Each $\mathbf{f}_i$ represents an image of the x-ray attenuation coefficient at the i-th kVp level. For this specific application, $N = 3$ corresponding to the hree energy levels of 80, 100, 120 kVps.

The second term $G(\mathbf{F},\boldsymbol{\rho},\boldsymbol{\lambda})$ characterizes the relationship among $\mathbf{F}$, relative electron density (rED) to water $\boldsymbol{\rho} \in \mathbb{R}^{M\times M}$, a diagonal matrix with each diagonal element corresponding to rED of each voxel, and elemental compositions (EC) $\boldsymbol{\lambda} \in \mathbb{R}^{M\times D}$, where D indicates the number of elements of interest. In this study, we focused on four elements ($D = 4$),

i.e. Hydrogen (H), Oxygen (O), Calcium (Ca), and Iodine (I). The first three are major elements constructing majority of human tissues, while I is the effective element in the iodine contrast agent. β is the parameter balancing the contribution from the two terms. Note that $\boldsymbol{\rho} \geq 0$ and $\boldsymbol{\lambda} \geq 0$ are naturally requested. In addition, let $\mathbf{1}_D$ and $\mathbf{1}_M$ denote column vectors of length D and M with all entries being unity, $\boldsymbol{\lambda}\mathbf{1}_D = \mathbf{1}_M$ is required since the compositions of all elements in each voxel should sum up to 1.

1) Reconstruction: Among many possible choices of models, we considered a tight-frame based image reconstruction model [2], [3] in this study. As such, the detailed expression of $R(\mathbf{F})$ can be given as

$$R(\mathbf{F}) = \|P\mathbf{F} - B\|_F^2 + 2\alpha_1 \sum_{i=1}^{N} \|W\mathbf{f}_i\|_1, \tag{2}$$

where P indicates the system matrix characterizing the ME-CBCT data acquisition process, while B represents the acquired ME-CBCT projection data. α_1 is the regularization parameter, and W is the tight frame (TF) operator. $\|\cdot\|_F$ and $\|\cdot\|_1$ denote the Frobenius Norm and l_1 norm, respectively. The first term in Eq. (2) enforces fidelity between the reconstructed ME-CBCT images and acquired projections. The second term encourages sparsity in transformed images to help remove noise and artifacts while preserving edges in $\mathbf{F}$.

2) Decomposition: $G(\mathbf{F}, \boldsymbol{\rho}, \boldsymbol{\lambda})$ is established to relate the x-ray attenuation coefficients with rED and EC, which was derived from an empirical forward model in [4]:

$$\mathbf{f}_i = \boldsymbol{\rho}(k_i^{PE}\tilde{\mathbf{Z}}^{3.62} + k_i^{R}\hat{\mathbf{Z}}^{1.86} + k_i^{C}). \tag{3}$$

Here k_i^{PE}, k_i^{R}, and k_i^{C} are machine-specific parameters characterizing contributions of photoelectric effect, Rayleigh scattering, and Compton scattering, respectively, to the x-ray attenuation at i-th energy level,. For a given scanner, these parameters can be calibrated before-hand using calibration materials . In addition, $\boldsymbol{\lambda}$ is encoded in $\tilde{\mathbf{Z}}$ and $\hat{\mathbf{Z}}$ as $\tilde{\mathbf{Z}} = (\boldsymbol{\lambda} z^{3.62})^{\frac{1}{3.62}}$ and $\hat{\mathbf{Z}} = (\boldsymbol{\lambda} z^{1.86})^{\frac{1}{1.86}}$, where $z = [z_1, z_2, \ldots, z_D]^T \in \mathbb{R}^D$ gives the atomic numbers for D different elements. Further, let $K^C = (k_1^C, k_2^C, \ldots, k_N^C)$, and $K = (k_1, k_2, \ldots, k_N)$ where $k_i = k_i^{PE} z^{3.62} + k_i^{R} z^{1.86}$. Through a simple derivation, the forward model in (3) can be further rewritten into a general form for multiple energy levels, i.e.

$$\mathbf{F} = \boldsymbol{\rho}(\boldsymbol{\lambda} K + \mathbf{1}_M K^C), \tag{4}$$

According to [5], [6], EC of each voxel in a CBCT image can be sparsely represented over a dictionary consisting of EC of commonly encountered human tissues. In a similar way as [5], [6], we construct a dictionary $\Lambda \in \mathbb{R}^{E\times D}$ consisting compositions of D elements (including I) of in total E materials. Note that I was not contained in most of the commonly encountered human tissues. We added random fractions of I to 50% of materials in the dictionary, where I compositions for the rest of the materials remained to be 0. Afterwards, the dictionary was then normalized, such that $\Lambda\mathbf{1}_D = \mathbf{1}_E$ to enforce the unity summation on elemental fractions for each material. Based on results in [5], [6], we are able to accurately represent EC of any material via $\boldsymbol{\lambda} \approx \mathbf{v}\Lambda$, where $\mathbf{v} \in \mathbb{R}^{M\times E}$ gives the dictionary coefficients with each row being a sparse vector specifying the contribution of each dictionary material. With these, we can reformulate the decomposition term $G(\mathbf{F}, \boldsymbol{\rho}, \lambda)$ in Eq. (2) as:

$$G(\mathbf{F}, \boldsymbol{\rho}, \mathbf{v}) = \|\mathbf{F} - \boldsymbol{\rho}(\mathbf{v}\Lambda K + \mathbf{1}_M K^C)\|_F^2, \tag{5}$$

with $\mathbf{v}$ being sparse. One obvious limitation of this model (5) is its non-convex formdue to product of $\boldsymbol{\rho}$ and $\mathbf{v}$. Given $\mathbf{v}\Lambda$ as approximations of EC, it is also natural to request $\mathbf{v}\Lambda\mathbf{1}_D = \mathbf{1}_M$. As a consequence, $\mathbf{v}\mathbf{1}_E = \mathbf{v}\Lambda\mathbf{1}_D = \mathbf{1}_M$, where $\mathbf{1}_E$ is the column vector of length E with all elements being 1. Then if we combine $\boldsymbol{\rho}$ and $\mathbf{v}$ as $\mathbf{x} = \boldsymbol{\rho} \cdot \mathbf{v}$, we may further rewrite this decomposition term together with the sparsity constraint as

$$G(\mathbf{F}, \mathbf{x}) = \|\mathbf{F} - \mathbf{x}\Lambda\mathbf{K} + \mathrm{diag}(\mathbf{x}\mathbf{1}_E)\mathbf{1}_M K^C\|_F^2 + \frac{2\alpha_2}{\beta}\|\mathbf{x}\|_1. \tag{6}$$

α_2 is the sparsity regularization parameter. $\mathrm{diag}(\cdot)$ indicates the two-way conversion between a diagonal matrix and a column vector. In contrast to (5), (6) is a convex model which can be solved more effectively and efficiently. With $\mathbf{x}$ solved, one can easily compute $\boldsymbol{\rho}$ and $\mathbf{v}$ as

$$\boldsymbol{\rho} = \mathrm{diag}(\mathbf{x}\mathbf{1}_E), \text{ and } \mathbf{v} = \boldsymbol{\rho}^{-1}\mathbf{x}, \tag{7}$$

3) Combined model: The explicit form of the proposed model can be given by combining the reconstruction and decomposition models derived in previous sections:

$$\begin{aligned}\min_{\mathbf{F},\mathbf{x}\geq 0} \frac{1}{2}\|P\mathbf{F} - B\|_F^2 + \frac{\beta}{2}\|\mathbf{F} - \mathbf{x}\Lambda\mathbf{K} + \mathrm{diag}(\mathbf{x}\mathbf{1}_E)\mathbf{1}_M K^C\|_F^2 \\ +\alpha_1 \sum_{i=1}^{N} \|W\mathbf{f}_i\|_1 + \alpha_2\|\mathbf{x}\|_1.\end{aligned} \tag{8}$$

The proposed model in (8) is convex and can be solved iteratively by alternating direction method of multipliers (ADMM) [7]. Similar to the iterative scheme proposed in [6], we incorporated ADMM to split the optimization problem into several subproblems, and iteratively solve F and X in an alternating fashion. Note that the closed-form solution for each subproblem is available, and the convergence of the proposed algorithm is guaranteed[7], [6].

C. Evaluations

To demonstrate feasibility and effectiveness of our method, we performed both simulation studies and experimental validations using a phantom with inserts containing water and iodine solution (175 mgI/mL) of different concentrations, see Figure 1. Here we used 9 inserts with iodine solution concentration of 0, 0.1, 0.5, 1.0, 1.5, 2.0, 3.0, 5.0, and 10%.

We considered a kVp-switching scan with 600 projections acquired in a full gantry rotation under 80, 100, and 120 kVps. While the experimental study realized this data acquisition on the Varian LINAC using its developer mode, in the simulation study, x-ray projections of a digital phantom representing the actual phantom was calculated using our in-house developed Monte Carlo simulation tool [8].

To benchmark our method, we compared the reconstruction performance against the clinical standard Filtered Back Projection (FBP) method. For FBP reconstruction, projections at

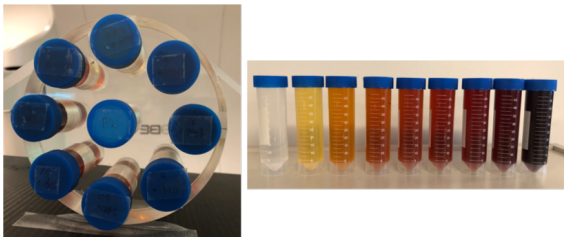


Fig. 1. Phantom used in this study with inserts of different concentrations of iodine.

80, 100, and 120 kVps were simulated/acquired individually with 600 projections covering a whole gantry rotation. Hence, the number of projections available at each kVp level for the FBP case is three times as that of for the proposed method.

To quantitatively evaluate the results, we considered two metrics. The first was contrast enhancement defined as $|\overline{x}_{roi} - \overline{x}_{bg}|$, where $\overline{x}_{roi}$ is the mean value of the image intensity of a region of interest (ROI) and $\overline{x}_{bg}$ is the mean value of the image intensity of the background non-iodine phantom region. The second metric was contrast-to-noise ratio (CNR), a key quantity characterizing the detectability of an object. CNR was defined as CNR $= \frac{2|\overline{x}_{\text{roi}} - \overline{x}_{\text{bg}}|}{\sigma_{\text{roi}} + \sigma_{\text{bg}}}$, where σ_{roi} and σ_{bg} are the the standard deviation of image intensity at the ROI and the background, respectively. In general, higher CNR indicates better detectability.

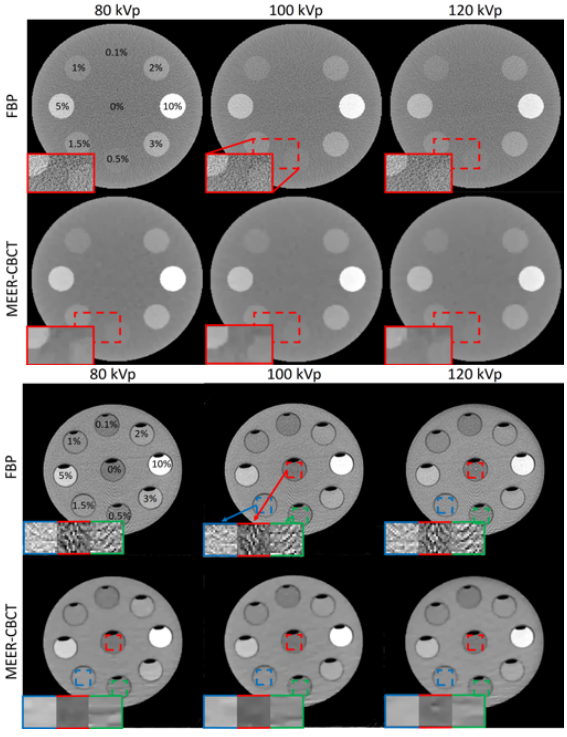


Fig. 2. Reconstructed images in simulation study (top) and experimental study (bottom).

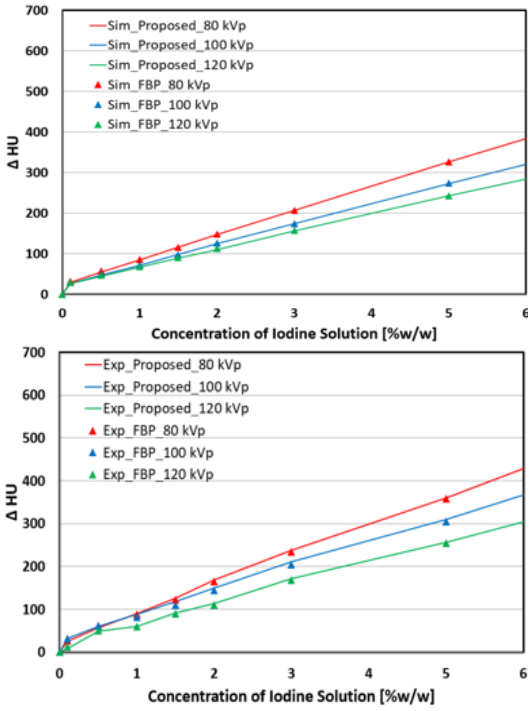


Fig. 3. Comparison of contrast enhancements between our method and FBP in simulation study (top) and experimental study (bottom). Result of 10% concentration not displayed to focus on relevant data range.

III. Results

Figure 2 presents the reconstructed x-ray attenuation images in the simulation and experimental studies. First and second rows are images of the phantom images reconstructed by FBP and MEER-CBCT. The images were displayed at a widow of [-500, 500] HU, while the zoomed-in patches were desplayed at [-200, 200] HU. It is easy to observe that the results using FBP method contain severe noise and streak artifacts. In contrast, the image quality produced by the MEER-CBCT method is much better, with noise and streak artifacts effectively suppressed.

For quantitative comparison, we first present the results of contrast enhancement in Figure 3. As expected, the contrast enhancement increases with iodine solution concentration. The enhancement levels in the images reconstructed by the FBP algorithm and by the proposed algorithm generally agree with each other, indicating that the use of regularization terms in our method did not suppress the image contrast.

Moreover, the resulting CNRs for iodine contrast agent of different concentrations are depicted in Figure 4. In both simulation and experimental studies, for a given concentration of iodine solution , the corresponding CNR obtained by our method is on average 6 times higher than that of the FBP method. In particular, in the experiment (Figure 4 (bottom)), with 4% of iodine solution concentration, the FBP method achieved contrast enhancement of ~250 HU, and CNR of ~2.5. This was considered as the minimal CNR required for successful tumor visualization in liver cancer. MEER-CBCT may achieve the same CNR at 0.5% of iodine solution con-

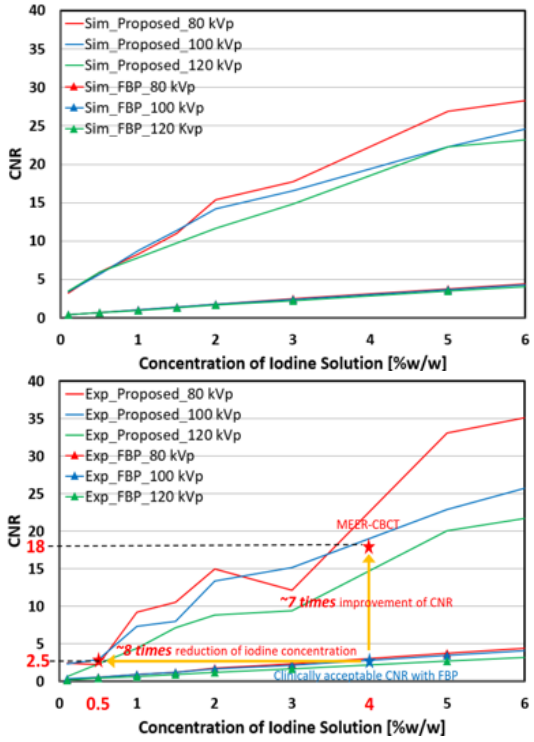


Fig. 4. Comparison of CNR between our method and FBP in simulation (top) and experimental (bottom) study. Result of 10% concentration not displayed to focus on relevant data range.

TABLE I
CONTRAST ENHANCEMENTS AND CNRS WITH DIFFERENT CONCENTRATION OF IODINE SOLUTION IN SIMULATION STUDY.

Iodine Solution Concentration (% w/w)	FBP		MEER-CBCT	
	Enhancement (HU)	CNR	Enhancement (HU)	CNR
0.1%	30.4	0.4	28.0	3.4
0.5%	51.6	0.7	48.9	5.9
1%	75.9	1.0	74.4	8.3
1.5%	102.0	1.4	101.1	10.7
2%	129.3	1.7	127.3	13.8
3%	179.7	2.4	179.0	16.4
5%	281.2	3.7	280.9	23.8

centration, about 1/8 of concentration needed by the standard FBP method.

Finally, we summarized the quantitative results averaged over the three kVps in Tables I and II for simulation and experimental studies, respectively.

IV. CONCLUSION AND DISCUSSIONS

To improve the detectability of iodine contrast agent in CBCT for image guidance of liver cancer radiotherapy, we developed a MEER CBCT framework that acquired a kVp-switching scan on a conventional CBCT platform and simultaneously reconstructed x-ray attenuation images at each kVp, the image of rED and images of EC. The composition of each voxel was subject to a constraint of a sparse representation of materials in a dictionary containing typical human tissues and iodine. In both simulation and experimental studies,

TABLE II
CONTRAST ENHANCEMENTS AND CNRS WITH DIFFERENT CONCENTRATION OF IODINE SOLUTION IN EXPERIMENTAL STUDY.

Iodine Solution Concentration (% w/w)	FBP		MEER-CBCT	
	Enhancement (HU)	CNR	Enhancement (HU)	CNR
0.1%	23.4	0.3	21.9	1.8
0.5%	57.9	0.5	55.0	2.5
1%	76.8	0.9	79.8	7.0
1.5%	107.9	1.1	111.9	8.6
2%	140.3	1.6	144.1	12.4
3%	202.6	2.1	206.4	12.3
5%	305.6	3.3	308.9	25.4

MEER-CBCT achieved similar contrast enhancement as FBP method, but about 6 times higher CNR on average. At 4% iodine solution concentration, FBP method achieved ~250 HU enhancement and CNR of ~2.5, considered as the standard CNR for successful tumor visualization. MEER-CBCT yielded the same CNR but at ~8 times lower iodine concentration of 0.5%.

Despite the exciting results, the current study has a few limitations. The liver tumor contrast enhancement has a complex kinetic behavior occurring during a time interval with a length comparable to the CBCT data acquisition time. Organ motion during the CBCT scan, if unaddressed, would further blur the resulting images and diminish the gain of contrast enhancement. These issues will likely hinder clinical translation of the proposed method. We are in the process of developing novel techniques to address these challenges.

ACKNOWLEDGMENT

This study is supported in part by a grant from the National Institutes of Health (NIH/NCI 1R01CA227289).

REFERENCES

[1] C. L. Eccles, R. V. Tse, M. A. Hawkins, M. T. Lee, D. J. Moseley, and L. A. Dawson, "Intravenous contrast-enhanced cone beam computed tomography (ivcbct) of intrahepatic tumors and vessels," *Advances in Radiation Oncology*, vol. 1, no. 1, pp. 43 – 50, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2452109416000117

[2] X. Jia, B. Dong, Y. Lou, and S. B. Jiang, "Gpu-based iterative cone-beam ct reconstruction using tight frame regularization," *Physics in Medicine & Biology*, vol. 56, no. 13, p. 3787, 2011.

[3] J.-F. Cai, S. Osher, and Z. Shen, "Split bregman methods and frame based image restoration," *Multiscale modeling & simulation*, vol. 8, no. 2, pp. 337–369, 2009.

[4] R. Rutherford, B. Pullan, and I. Isherwood, "Measurement of effective atomic number and electron density using an emi scanner," *Neuroradiology*, vol. 11, no. 1, pp. 15–21, 1976.

[5] C. Shen, B. Li, L. Chen, M. Yang, Y. Lou, and X. Jia, "Material elemental decomposition in dual and multi-energy ct via a sparsity-dictionary approach for proton stopping power ratio calculation," *Medical physics*, vol. 45, no. 4, pp. 1491–1503, 2018.

[6] C. Shen, B. Li, Y. Lou, M. Yang, L. Zhou, and X. Jia, "Multienergy element-resolved cone beam ct (meer-cbct) realized on a conventional cbct platform," *Medical physics*, vol. 45, no. 10, pp. 4461–4470, 2018.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends ® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[8] X. Jia, H. Yan, L. Cerviño, M. Folkerts, and S. B. Jiang, "A gpu tool for efficient, accurate, and realistic simulation of cone beam ct projections," *Medical physics*, vol. 39, no. 12, pp. 7368–7378, 2012.

A Preconditioning Scheme for Accelerating A Non-Convex Primal-Dual Reconstruction Algorithm in Multispectral CT

Buxin Chen, Emil Y. Sidky, Zheng Zhang, Dan Xia, and Xiaochuan Pan

***Abstract*—In this work, we aim to accelerate the convergence speed of the non-convex primal-dual (NCPD) algorithm for multispectral CT reconstruction by devising non-diagonal preconditioning (NDPC) step matrices in the NCPD algorithm. The NDPC strategy is designed to offset the spread of singular values in the discrete X-ray transform and improve the overall conditioning of the system matrix in the multispectral CT data model. An instance of the NCPD algorithm was derived and applied to data with both scalar steps and NDPC step matrices in the dual and primal updates for comparison. Quantitative studies have been carried out with data generated from digital phantom images for demonstrating and characterizing the convergence trends of both algorithms in terms of convergence metrics such as data and image RMSEs. The results of the study reveal clearly that the NCPD algorithm with NDPC step matrices achieves a numerical convergence (limited by computer precision) faster than its counterpart without NDPC step matrices.**

I. Introduction

We have previously proposed a non-convex optimization program for multispectral CT reconstruction and developed a non-convex primal-dual (NCPD) algorithm for numerically solving the optimization program [1]. The NCPD algorithm is based on a first-order, primal-dual algorithm [2], initially developed by Chambolle and Pock [3] (CPPD), with an addtional data correction step to account for the non-linearity in the multispectral data model. In the original CPPD algorithm, the dual and primal updates employ scalar steps, which was inherited to our NCPD algorithm. More recently, a step-preconditioner in the form of block-diagonal matrices has been proposed to accelerate the convergence speed of the MOCCA algorithm [4]. The preconditioner is motivated by preserving the rotation invariance of the material expansion set. We have also proposed a different form of non-diagonal preconditioning, or NCPD, step matrices designed to offset the spread in singular values of the discrete X-ray transform (DXT) matrix, as poor conditioning in the X-ray transform matrix often leads to inefficient first-order primal-dual algorithm. In this work, we aim to incorporate such NDPC step matrices into the NCPD algorithm for multispectral CT reconstruction. The system matrix in the data model for multispectral CT combines the DXT matrix and a μ matrix, and is often poorly conditioned as well, further impacting the overall conditioning of the system matrix and thus welcoming the introduction of preconditioner. We derive a specific instance of the NCPD algorithm for a total-variation-constrained data-ℓ_2 minimization problem using the non-linear data model in multispectral CT, and replace the scalar steps in the dual and primal updates with NDPC step matrices. We compare the NDPC step matrices with the scalar steps and characterize their convergence trends in terms of image and data RMSEs and also verify numerically that adding NDPC does not alter the eventual convergence of the NCPD algorithm.

In this work, we show that the NCPD algorithms with scalar steps and NDPC step matrices converge numerically to the solutions specified by the program considered. Moreover, we are also interested in the convergence speeds and trends of the algorithms at iterations prior to reaching eventually their numerical convergences that are limited by the computer precision in the study. In particular, we seek to improve the NCPD algorithm's convergence speed to achieve its numerical convergence with reduced iterations. This investigation may yield insights into the design of practical procedures, based on or motivated by the iterative updates of the algorithm with improved convergence speed, for yielding reconstructions of practical, task-specific utility with fewer iterations. Furthermore, an algorithm that achieves its numerical convergence (limited by the computer precision) with reduced iterations would also be of direct value to research on iterative algorithms in which one often seeks numerical convergences of the iterative algorithms.

B. Chen, E. Y. Sidky. Z. Zhang, amd D. Xia are with the Department of Radiology, The University of Chicago, Chicago, IL 60637, USA

X. Pan is with the Departments of Radiology & Radiation and Cellular Oncology, The University of Chicago, Chicago, IL 60637, USA

II. Materials and methods

A. Discrete-to-discrete data model

The discrete-to-discrete data model in multispectral CT can be written as [5],

$$g_j^{[s]}(\mathbf{b}) = -\ln \sum_m q_{jm}^{[s]} \exp\left(-\sum_k \mu_{km} \sum_i a_{ji}^{[s]} b_{ki}\right), \quad (1)$$

where $q_{jm}^{[s]}$ denotes the normalized X-ray spectrum at energy m at ray j in spectral set s, μ_{km} the mass attenuation coefficient at energy m for basis material k, $a_{ji}^{[s]}$ the intersection length of ray j in spectral set s with voxel i, and b_{ki} the basis image of material k at voxel i. Vector $\mathbf{b}$ denotes an aggregate basis-image vector formed by concatenating individual basis-image vectors $\mathbf{b}_k$ with element b_{ki}, and $\mathbf{g}(\mathbf{b})$ an aggregate model data vector by concatenating model data vector from each spectral set $\mathbf{g}^{[s]}(\mathbf{b})$ with element $g_j^{[s]}(\mathbf{b})$. Similarly, we use $\mathbf{g}_{\mathcal{M}}$, $\mathbf{g}_{\mathcal{M}}^{[s]}$,

$g^{[s]}_{\mathcal{M}j}$ to denote the measured data for the aggregate vector, each spectral set, and each ray, respectively.

B. Non-convex optimization program and NCPD algorithm

Based on the non-linear data model in Eq. (1), we form a non-convex optimization program involving the ℓ_2-norm between model and measured data subject to a total-variation (TV) constraint on a monochromatic image at energy E_0, as

$$\begin{aligned} &\min_{\mathbf{b}} \frac{1}{2}||\mathbf{g}_{\mathcal{M}} - \mathbf{g}(\mathbf{b})||_2^2 \\ &\text{s.t. } ||(|\nabla(\mu_1(E_0)\mathbf{b}_1 + \mu_2(E_0)\mathbf{b}_2)|)||_1 \leq \gamma, \end{aligned} \tag{2}$$

where, without loss of generality, a two-basis decomposition model is used with $\mathbf{b} = (\mathbf{b}_1^\top, \mathbf{b}_2^\top)^\top$ and $\mu_1(E_0)$, $\mu_2(E_0)$, simply referred to as μ_1 and μ_2 hereafter, mass attenuation coefficients of the two basis materials at energy E_0, $|\nabla u|$ the gradient-magnitude image of u, and ∇ a linear transform of finite differencing [2].

A key observation in developing the NCPD algorithm is that the non-linear data model in Eq. (1) can be separated into a linear term and a non-linear term using Taylor expansion, as

$$g_j^{[s]}(\mathbf{b}) = \bar{g}_j^{[s]}(\mathbf{b}) + \Delta g_j^{[s]}(\mathbf{b}), \tag{3}$$

where

$$\bar{g}_j^{[s]}(\mathbf{b}) = \sum_k \bar{\mu}_{jk}^{[s]} \sum_i a_{ji}^{[s]} b_{ki}, \tag{4}$$

$$\Delta g_j^{[s]}(\mathbf{b}) = -\ln \sum_m q_{jm}^{[s]} \exp\left(-\sum_k \Delta\mu_{jkm}^{[s]} \sum_i a_{ji}^{[s]} b_{ki}\right), \tag{5}$$

$\bar{\mu}_{jk}^{[s]} = \sum_m q_{jm}^{[s]} \mu_{km}$, and $\Delta\mu_{jkm}^{[s]} = \mu_{km} - \bar{\mu}_{jk}^{[s]}$. As a result, in the same concatenating fashion as aforementioned, two aggregate vectors $\bar{\mathbf{g}}(\mathbf{b})$ and $\Delta\mathbf{g}(\mathbf{b})$ can be formed from $\bar{g}_j^{[s]}(\mathbf{b})$ and $\Delta g_j^{[s]}(\mathbf{b})$, respectively, and are connected to $\mathbf{g}(\mathbf{b})$ as

$$\mathbf{g}(\mathbf{b}) - \Delta\mathbf{g}(\mathbf{b}) = \bar{\mathbf{g}}(\mathbf{b}) = \mathcal{H}\mathbf{b}, \tag{6}$$

where system matrix $\mathcal{H}$ has elements $h_{jki}^{[s]} = \bar{\mu}_{jk}^{[s]} a_{ji}^{[s]}$, with spectral set s and ray j indexing the rows of $\mathcal{H}$, and basis material k and pixel i indexing the columns of $\mathcal{H}$. The resulting system matrix, with the addition of $\bar{\mu}_{jk}^{[s]}$, is different from that of a conventional CT reconstruction. Using Eq. (6), we can re-write the optimization program in Eq. (2) as

$$\begin{aligned} &\min_{\mathbf{b}} \frac{1}{2}||\mathbf{g}_{\mathcal{M}} - \Delta\mathbf{g}(\mathbf{b}) - \mathcal{H}\mathbf{b}||_2^2 \\ &\text{s.t. } ||(|\nabla(\mu_1\mathbf{b}_1 + \mu_2\mathbf{b}_2)|)||_1 \leq \gamma. \end{aligned} \tag{7}$$

We then derive the NCPD algorithm by first deriving an instance of the CPPD algorithm from Eq. (7) assuming $(\mathbf{g}_{\mathcal{M}} - \Delta\mathbf{g}(\mathbf{b}))$ is known and then use the current iteration of the basis images to update and estimate the term as $(\mathbf{g}_{\mathcal{M}} - \Delta\mathbf{g}(\mathbf{b}^{(n)}))$. In the derivation of the CPPD algorithm instance, a stacked matrix is used and defined as

$$\mathcal{K} = \begin{pmatrix} \mathcal{H} \\ \alpha\mu_1\nabla, \alpha\mu_2\nabla \end{pmatrix}, \tag{8}$$

where $\alpha = ||\mathcal{H}||_2 / ||(\mu_1\nabla, \mu_2\nabla)||_2$. This stacked matrix $\mathcal{K}$ is key to the computation of the proposed NDPC step matrix.

C. Dual and primal updates

1) Scalar steps: The pseudo-code for the NCPD algorithm tailored to the non-convex optimization program in Eq. (2) can be found in Ref. [1]. We extract two essential steps, the dual and primal updates, and summarize them below as

$$\mathbf{y}^{(n+1)} = \text{prox}_\sigma[F^*](\mathbf{y}^{(n)} + \sigma\mathcal{K}\bar{\mathbf{b}}^{(n)}) \tag{9}$$

as the dual update, and

$$\mathbf{b}^{(n+1)} = \mathbf{b}^{(n)} - \tau\mathcal{K}^\top\mathbf{y}^{(n+1)} \tag{10}$$

as the primal update. $\mathbf{y}$ is the dual variable, $\text{prox}[\cdot]$ the proximal mapping operation, F^* the convex conjugate of function $F(\mathbf{y}) = F(\mathbf{u}, \mathbf{v}) = \frac{1}{2}||\mathbf{u} - \mathbf{g}_{\mathcal{M}}||_2^2 + \delta_{\text{diamond}(\alpha\gamma)}(|\mathbf{v}|)$, $\sigma = \frac{\rho}{||\mathcal{K}||_2}$, and $\tau = \frac{1}{\rho||\mathcal{K}||_2}$. ρ is the step size ratio parameter balancing out the dual and primal updates and can be tuned.

2) NDPC step matrices: Instead of σ and τ, we use two matrices, Σ and T, as the step sizes for the dual and primal updates, which, with $\sigma = 1/||T\mathcal{K}^\top\mathcal{K}||_2$, are defined as

$$\Sigma = \rho\sigma I, \quad T \approx (\mathcal{K}^\top\mathcal{K})^{-1}/\rho, \tag{11}$$

where σ is calculated following T and ρ is again the step size ratio parameter that needs to be tuned. Σ is a diagonal matrix, while T, a non-diagonal preconditioing step matrix, is the inverse of the symmetric positive semi-definite matrix $(\mathcal{K}^\top\mathcal{K})$. The inverse can be approximated using a truncated series of eigenvalues and eigenvectors of $(\mathcal{K}^\top\mathcal{K})$, as

$$T = \frac{I}{e_L} + \sum_{i=1}^{L-1} u_i \left(\frac{1}{e_i} - \frac{1}{e_L}\right) u_i^\top, \tag{12}$$

where e_i is the i-th eigenvalue (sorted from largest to smallest), u_i its corresponding eigenvector, and L the number of eigenvectors used in the expansion. As a result, the dual and primal updates with step matrices are now

$$\mathbf{y}^{(n+1)} = \text{prox}_\sigma[F^*](\mathbf{y}^{(n)} + \Sigma\mathcal{K}\bar{\mathbf{b}}^{(n)}) \tag{13}$$

as the dual update, and

$$\mathbf{b}^{(n+1)} = \mathbf{b}^{(n)} - T\mathcal{K}^\top\mathbf{y}^{(n+1)} \tag{14}$$

as the primal update.

D. Study design

Multispectral CT data are generated using Eq. (1), assuming a circular fan-beam geometry with 160 views uniformly distributed over 2π. The source-to-detector and source-to-center-of-rotation distances are 1,500 and 1,000 mm, respectively. The linear detector is 400 mm in length and consists of 256 1.56-mm bins. Two X-ray spectra of 80 and 140 kVp are generated using the TASMICS worksheet [6], assuming a tungsten anode and a 5-mm Al filter. A digital phantom is used mimicking GAMMEX DE-472 dual-energy phantom with iodine and calcium inserts of various contrast levels. It is sampled on a 128×128 image array with 1.95-mm square pixels. The digital phantom is then decomposed into two basis materials, water and bone, as shown in Fig. 1, which are used as the truth basis images in the data generation.

(a) water basis (b) water basis (c) mono@100keV

Fig. 1: Water (a) and bone (b) basis images used to generate the ideal data, and a monochromatic images at 100 keV (c) combined by the basis images. The display windows are [0.0, 1.2] for the basis images, and [-100, 100] HU for the monochromatic image. Two red arrows point to two inserts with the smallest contrast levels in the water basis image.

The NCPD algorithm for the non-convex optimization program in Eq. (2) is derived in Ref. [1] and applied to the consistent data in this study with both scalar steps and NDPC step matrices for comparison. Two convergence metrics, namely, data and image RMSEs, are computed to characterize the convergence trend of both algorithms in terms of minimization of the data-ℓ_2 objective function and the accurate recovery of the basis images. They are defined as

$$\begin{aligned} \text{data RMSE} &= \frac{||\mathbf{g}_{\mathcal{M}} - \mathbf{g}(\mathbf{b}^{(n)})||_2}{||\mathbf{g}_{\mathcal{M}}||_2}, \\ \text{image RMSE} &= \frac{||\mathbf{b}^{\text{truth}} - \mathbf{b}^{(n)}||_2}{||\mathbf{b}^{\text{truth}}||_2}. \end{aligned} \tag{15}$$

III. Results

We show in Fig. 2 convergence metrics, the data and image RMSEs, as functions of the iteration number for the NCPD algorithms with scalar steps and NDPC step matrices. They are plotted on a log-log scale to reveal in detail convergence trends for the algorithms. Each algorithm was tuned with the ρ parameter to find the fastest convergence, i.e., a number of ρ values were sampled and the one with the fastest descent was selected for the respective algorithm ($\rho = 0.1$ and 0.01 for the scalar and NDPC strategies, respectively, in this study). Both algorithms were run for 2,000 iteration here. The NDPC strategy shows faster convergence in terms of data and image RMSEs than the scalar-step strategy.

We further display in Fig. 3 water basis images and monochromatic images at 100 keV from both algorithms at various intermediate iterations at, from top to bottom, 200, 300, 500, and 1,000. In the water basis images, the two inserts with the smallest contrast level (pointed to by arrows) would be most impacted by lack of convergence. At 200 iterations, the NDPC water basis image seems to have significantly less cross-talk from the bone basis than the scalar one. The two inserts pointed to by arrows in the scalar water basis image show negative contrast (darker than the background), and the polarity of these two contrast levels did not correctly flip until 500 iterations for the scalar water basis images, and gradually recovered to its accurate contrast levels around 1,000 iterations. For the monochromatic images at 100 keV with a

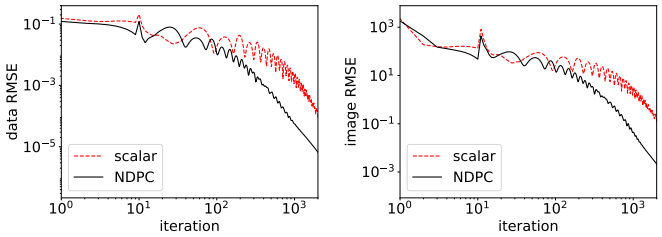


Fig. 2: Convergence metrics, data (left) and image (right) RMSEs, as functions of iteration number for the NDPC algorithm with scalar steps and NDPC step matrices. The metrics are defined in Eq. (15). Both algorithm were run for 2,000 iterations to show their convergence trends.

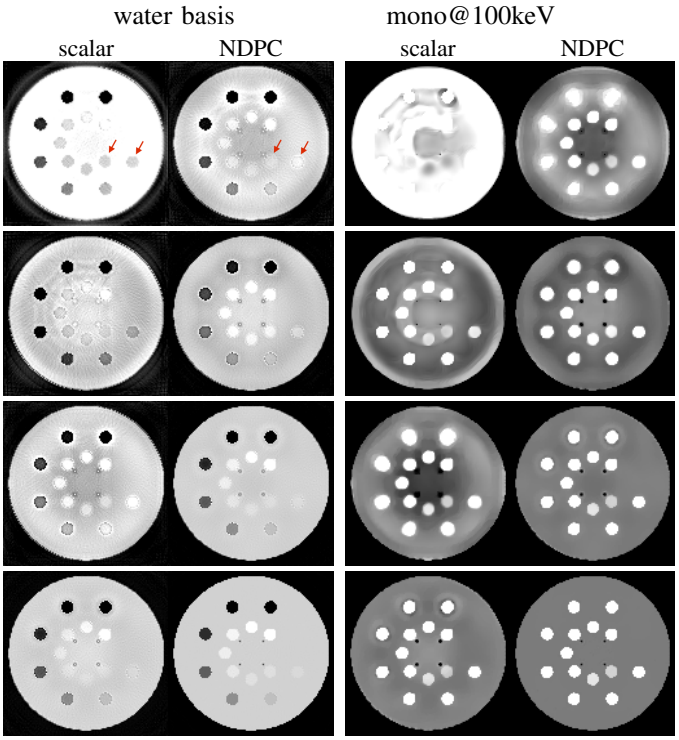


Fig. 3: Water basis images (left two columns) and monochromatic images at 100 keV (right two columns) at intermediate iterations n obtained by using the NCPD algorithm with scalar steps and NDPC step matrices. From top to bottom iteration numbers are: 200, 300, 500, 1000. The display windows are [0.0, 1.2] for the water basis images, and [-100, 100] HU for the monochromatic images.

narrow display window of [-100, 100] HU, the scalar ones show significant artifacts at 200, 300, and 500 iterations, and rings artifacts and background non-uniformity at 1,000 iteration, where the NDPC ones look visually the same to the truth in Fig. 1b at 1,000 iteration and very similar at 500 iteration. Overall, it can be noted that the NCPD algorithm with NDPC step matrices accurately recovers the basis images much faster than the scalar one.

Lastly, we note that, given enough iterations, both algorithms converged in this study. In Fig. 4, we show both convergence metrics from the NCPD algorithm with scalar

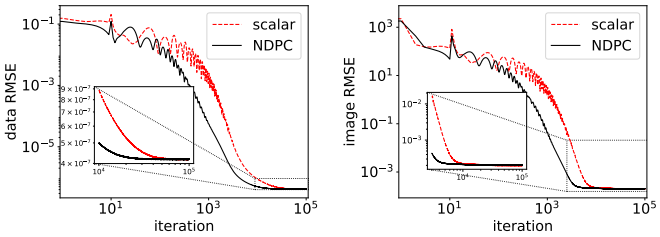


Fig. 4: Convergence metrics, data (left) and image (right) RMSEs, as functions of iteration number for the NDPC algorithm with scalar steps and NDPC step matrices. Both algorithm were run for 100,000 iterations to show that they both eventually converged. The tails of the convergence curves are zoomed in for revealing that the NDPC strategy was faster to computer precision, although both algorithms converged to the same level of computer precision eventually.

and NDPC step matrices as functions of the iteration nubmer, plotted on a log-log scale. Both algorithms were run for 100,000 iterations. The algorithm with scalar steps, although slower than the NDPC one in the initial convergence rate, eventually converged to the same level of computer precision of the latter for both metrics. While the eventual convergence of the algorithm with scalar steps is consistent with our previous results [1] (where the algorithm was simply referred to as ncCP), it is reassuring to observe the numerical results of the eventual convergence of the NCPD algorithm with NDPC step matrices, considering the lack of theoretical proof for the algorithm's convergence. In addition, zoomed-in views of the tails of both convergence curves in Fig. 4 show that the NDPC strategy reached to numerical convergence at the computer precision faster than the scalar strategy.

IV. Conclusions

In this work, we have applied non-diagonal preconditioning step matrices to an instance of the NCPD algorithm tailored for the TV-constrained data-ℓ_2 minimization problem in multispectral CT reconstruction. Inverse crime studies were carried out from consistent data generated from a digital phantom mimicking a standardized physical dual-energy phantom, to characterize the convergence trend of the algorithm in terms of minimization of the objective function and accurate recovery of the basis images. The results suggest that, compared the NCPD algorithm with scalar steps, the NDPC step matrices could achieve faster convergence speed under the conditions in this study. Further, it was shown that adding the NDPC step matrices to the dual and primal updates did not change the numerical convergence of the NDPC algorithm.

Future work will focus on further characterizing the NDPC strategy by applying it to inconsistent data with, e.g., noise and decomposition error. In addition, it is of our interest to study and implement different designs of the NDPC step matrices to speed up the algorithm convergence. The block-diagonal preconditioning matrices used in Ref. [4] are an option. The exact inversion of the block-diagonal matrix would be feasible in the multispectral CT reconstruction, since the size of the block is equal to the number of spectral set (or energy window in the case of photon-counting detector). The computation overhead is also only one-time for the NCPD algorithm, given that the Taylor expansion in Eq. (3) is around a fixed point. Another preconditioner to consider is the μ-preconditioner used in Ref. [7], which rotates the material basis expansion into an orthogonal one. Such rotation, which can lead to senstive convergence behavior of the algorithm, can be incorporated into the NCPD algorithm with both scalar steps and preconditioning step matrices. Additional results will be presented at the conference.

Acknowledgment

This work was supported in part by NIH R01 Grants EB018102 and EB023968, and the Grayson-Jockey Club Research Grant.

References

[1] B. Chen, Z. Zhang, D. Xia, E. Sidky, and X. Pan, "Non-convex chambolle-pock algorithm for multispectral CT," in *The Fifth International Conference on Image Formation in X-Ray Computed Tomography*, 2018.

[2] E. Y. Sidky, J. H. Jorgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm," *Physics in Medicine and Biology*, vol. 57, no. 10, p. 3065, 2012. [Online]. Available: http://stacks.iop.org/0031-9155/57/i=10/a=3065

[3] A. Chambolle and T. Pock, "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, May 2011. [Online]. Available: http://link.springer.com/article/10.1007/s10851-010-0251-1

[4] E. Y. Sidky, R. F. Barber, T. Gilat-Schmidt, and X. Pan, "Three material decomposition for spectral computed tomography enabled by block-diagonal step-preconditioning," *arXiv preprint arXiv:1801.06263*, 2018.

[5] B. Chen, Z. Zhang, E. Y. Sidky, D. Xia, and X. Pan, "Image reconstruction and scan configurations enabled by optimization-based algorithms in multispectral CT," *Physics in Medicine & Biology*, vol. 62, no. 22, p. 8763, 2017.

[6] A. M. Hernandez and J. M. Boone, "Tungsten anode spectral model using interpolating cubic splines: Unfiltered xray spectra from 20 kV to 640 kV," *Medical Physics*, vol. 41, no. 4, pp. 042 101–1–042 101–15, Apr. 2014. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1118/1.4866216/abstract

[7] R. F. Barber, E. Y. Sidky, T. G. Schmidt, and X. Pan, "An algorithm for constrained one-step inversion of spectral CT data," *Physics in Medicine and Biology*, vol. 61, no. 10, p. 3784, 2016.

Radiopaque Temporary Embolic Agent for Direct Visualization During Surgical Procedure and Follow up Treatments

Priyanka Ray, Parikshit Moitra, Yicheng Zhang, Tor Jensen, Scott Santeler and Dipanjan Pan*

***Abstract*— Embolization is an effective, minimally invasive technique to control the flow of blood in the body. This procedure is used to treat many conditions including: tumors, aneurysms and arteriovenous malformations. Currently, no inherent radiopaque temporary embolic agents are available in the clinic that allows direct imaging of the material during the procedure and during follow up treatments. An ideal agent would combine the visibility of contrast agents, have the shape-ability and versatility of Gelfoam®, and still maintain its efficacy to stop blood flow in the target tissue without causing irreversible ischemia in the process. It would allow the physician direct visualization of the actual plug without the need for accompanying contrast. More precise placement and faster procedures could result from improved visualization of a Gelfoam® like agent. We have developed a translatable, radiopaque temporary embolic agent and characterized its' physico-chemical and biological properties followed by a proof-concept in vivo pilot study. A radiopaque embolization agent could reduce repeated radiation exposure to the patient, the physician and staff and may also allow subsequent imaging of treated areas of concern with less imaging artifact from permanent metallic plugs.**

I. Introduction

Embolization is an effective, minimally invasive technique to control the flow of blood in the body. [1-2] This procedure is used to treat many conditions including: tumors, aneurysms and arteriovenous malformations. Typically, an artificial embolus is administered with a catheter. The embolus serves to clot the blood vessel by mechanically preventing the flow of blood. Embolization can effectively shrink tumors by preventing nutrient rich blood from reaching them and reduce the risk of an aneurysm rupturing by arresting blood flow through the damaged vessel. Multiple different embolization products exist including materials such as plastic beads, metal coils, mesh plugs and glue. [3] Embolics are commonly classified as either temporary or permanent depending on their lifespan. Permanent embolics include polyvinyl alcohol (PVA) particles, liquid glue, detachable balloons, and coils. These embolics are intended to remain in the patient, whereas temporary embolics are not. Only one temporary, shapeable agent is currently available for clinical use, i.e. gelatin sponge. When choosing an embolic agent to employ, the interventional radiologist must balance the risk of depriving all blood supply to an area (ischemia), versus temporarily stopping a source of active bleeding or abnormal flow. [1-7] Micron-sized particulates are radio-transparent (not visible by standard imaging) but can permanently and irreversibly block the blood supply to a tissue or organ. The location of the embolized particulate can only be implied by the absence of blood flow when a contrast agent is administered, and flow is not visible within the plugged vessels. The absence of visualization can lead to longer and greater amounts of x-ray exposure to perform angiography to assess the location of the newly placed embolic material. Larger metallic plugs and coils are visible and can be seen easily with imaging techniques but remain in the body for the rest of the patient's life and can create artifacts for future imaging when MRIs or CTs are later used for a patient. Gelatin sponge functions like a particulate but is shapeable and can used as a plug or as a slurry. Gelfoam®, a commercially available (Pfizer) gelatin sponge, is often used in trauma as it is thought to be temporary. Gelfoam® is a water insoluble compressed sponge made from porcine skin that is intended to be used as a surface hemostatic. Although not originally developed for use as an embolic agent, the off-label use of Gelfoam® as a temporary embolic fall within the medical standard of care. Gelfoam is chosen as preferred embolics in multiple applications. Arterial embolization of the splenic artery in blunt trauma is good clinical example from which to draw comparisons. Use of Gelfoam for embolization is faster (32 min on average) than coil embolization (53 min) and has a 6% failure rate as opposed to 12% failure rate in coil embolization. [4] However, other studies of efficacy or embolization, suggest failures from splenic artery embolization were procedural and technical in origin. [8] One complication faced by radiologists when working with Gelfoam as a temporary embolic is the lack of fidelity associated with relocating it after its insertion. Gelfoam® is white, and once exposed to blood, it will blend in with its surroundings. Several studies suggest Gelfoam® will eventually liquefy and be absorbed in the body within six weeks of insertion. However, when a procedure related complication arises before the complete disintegration of the Gelfoam®, it is challenging for a radiologist to pinpoint the location of the Gelfoam® once the patient is back on the operating table. It can stop the blood flow to an area of bleeding and will allow the tissue to heal without leaving

Rsearch supported by University of Illinois, National Institute of Health (NIBIB).

P.R. and P.M.. are with Department of Chemical Biochemical and Environmental Engineering and Pediatrics at University of Maryland Baltimore County and University of Maryland Baltimore, MD, USA.

T.J. and S. S. .are with Carle Foundation Hospital and University of Illinois College of Medicine, Urbana, IL 61801 USA.

*D. P. was with Department of Bioengineering, Beckman Institute of Advanced Science and Technology and Department of Materials Science and Engineering University of Illinois at Urbana-Champaign, Urbana IL 61801 USA and now at Department of Chemical Biochemical and Environmental Engineering and Pediatrics at University of Maryland Baltimore County and University of Maryland Baltimore, MD, USA.

(e-mail: dipanjan@som.umaryland.edu).

behind metal or plastic in the body. Unfortunately, like other particulates, the distribution of its coverage can only be implied by diminished or absent flow during contrast angiography. Patients who have sustained multiple wounds or have multiple sources of bleeding can require repeated imaging and embolization procedures. As blood reroutes through different blood vessel networks following an embolization, new areas of bleeding can present themselves. Being able to identify previously treated areas with a radio-opaque embolic agent that has the flexible application of a gelatin sponge would be helpful to the operator. With current materials only permanent devices allow this post treatment identification of embolized blood supplies. Failures of successful embolization for stasis were theorized to be secondary to rebleeding events in the absorbable embolic group. It is our hope that radiopaque absorbable embolics could reduce technical failure by better identifying areas of successful embolization or allow repeat embolization with greater efficacy due to successful identification or previously embolized segments of the target organ. We therefore conclude that an ideal temporary embolic agent should exhibit mechanical stability allowing resistance of the blood flow and reducing the possibility of recanalization before the bleeding vessel heals. Moreover, the agent should be relatively benign and exhibit no toxicity compared with existing materials, which require more biocompatible composition to avoid cardiovascular toxicity or strong immune reaction. Finally, the agent should be surgery-amenable to achieve effective embolization results.

II. Materials and Methods

A. Integrating polyiohexol chain into Gelfoam using a Crosslinker

A synthetic protocol was followed to ensure that that iohexol is chemically bound to Gelfoam matrix to integrate a polyiohexol chain through the porous openings of Gelfoam (Figure 1). By creating this connected matrix, iohexol should not diffuse from the Gelfoam into the bloodstream, and thus, should be traceable under a Computed Tomography (CT) scanner. The protocol that was followed involved using the crosslinking agent, Hexamethylene diisocyanate along with a catalyst, dibutylin dilaurate to drive the formation of polyiohexol. 50 mg of 0.6 mmol iohexol is dissolved in 2 mL of N, N-Dimethylmethanamine. Afterwards, 30.7 mg of 16 µmol Hexamethylene diisocyanate and 1 µL of 16 µmol dibutylin dilaurate is introduced into the solution. This entire mixture is stirred at 80 °C for 24 hours with a Gelfoam volume of 2 mm x 2 mm x 7 mm. Following that timeframe, the Gelfoam is removed from the beaker, encapsulated in an Eppendorf tube, and placed in an evaporator to remove excess liquid. Drying the Gelfoam can take 3-5 days. Further tests need to be done to optimize drying time. Experimental groups prepared were polyiohexolinated Gelfoam, Gelfoam dipped in excess polyiohexol solution, Gelfoam dipped in control iohexol, and Gelfoam with no iohexol.

B. Modeling Vasculature using Polyurethane and Donkey Serum

Once the Gelfoam has completely dried, it was moved into a 50 mm strip of polyurethane (PU) tubing with 2 mm inner diameter. In a 15 mL conical, a 1:1 ratio of 5 mL donkey serum and 5 mL red food coloring was prepared as a blood analogue. From there, 200 uL of donkey serum would be injected to each experimental PU tube. To mimic vasculature flow, the blood analog would be extracted and reintroduced into each tube 10 times per day for three days.

In addition to the flush system described above, a water pump was used to provide an improved representation circulatory flow of fluid. With continuous flow, the

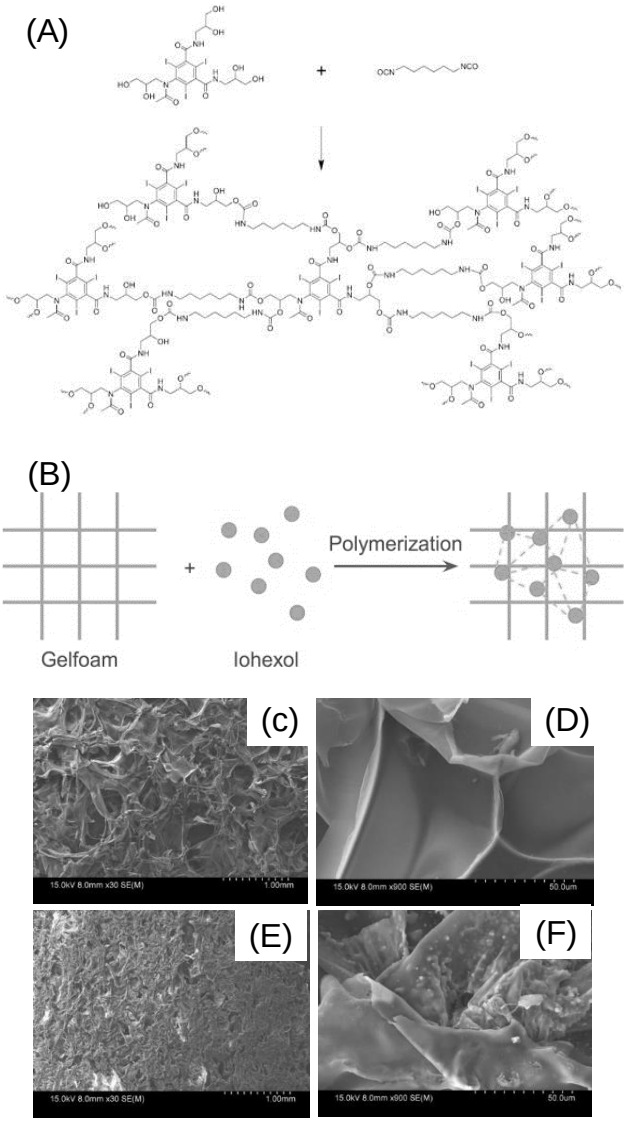


Figure 1. A) One-pot synthesis of radiopaque temporary embolic agent (RTEA); B) Schematics of polyiohexol interwoven in Gelfoam®. SEM images of Gelfoam Samples. C) & D) Commercial Gelfoam. E) & F) Radiopaque Gelfoam (RTEA).

polyiohexolinated Gelfoam showed embolic properties by preventing the blood analogue from passing through the polyurethane tube. was gently washed with PBS. Finally, a mixture of 15µL of ethanolamine (1%), glycine (1%)

solution in PBS (v/v 9:1) was prepared and used for 20 min to block the unreacted succinimidyl group of DTSP with antibody. Finally, the sensor was dried overnight under a

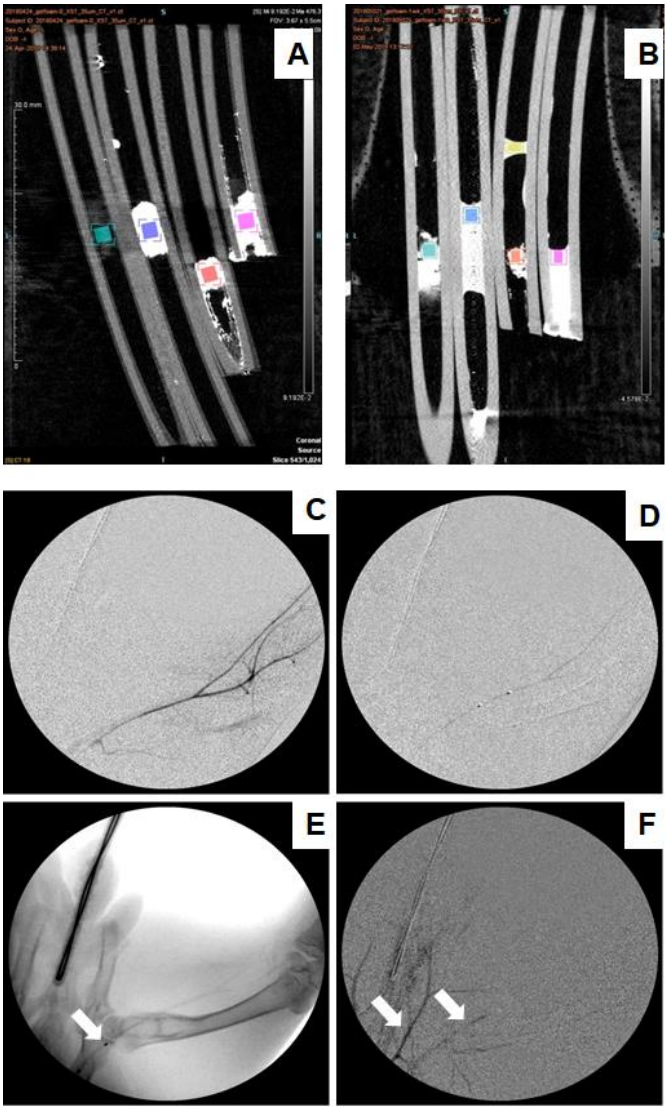


Figure 2. (A) Before washing without fluid in the tube. From left to right, commercial Gelfoam (Control); commercial Gelfoam mixed with iohexol; RTEA; Gelfoam in polyiohexol Solution. (B) After one-week wash test with simulated blood (same order). (C) A pre-occlusion iodine shot was applied to visualize the vasculature. (D) Using a subtraction angiography, the occlusion of the vessel by use of the radiopaque polymer was monitored. (E) The catheter was retreated prior to post-occlusion iodine shot; (F) that showed the retrograding agent circumventing the occluded vessel.

biosafety hood and stored in an airtight container at 4 °C until further used.

C. Imaging Gelfoam using a Computed Tomography (CT) Scanner

All the experimental groups of Gelfoam being tested were imaged using a Siemens Invion CT Scanner at Beckman Institute. Scans were performed prior to introducing blood analog and after introducing blood analog. Scans after introducing blood analog were performed for two sets: the first set was an experimental group that was imaged 5 days after testing and the second set was an experimental group that was imaged 7 days after testing.

III. Results and Discussion

To demonstrate proof of concept for the proposed iohexol [9-12] modification method, an RTEA was synthesized using hexamethylene diisocyanate as an iohexol crosslinking agent in the presence of Gelfoam and a dibutylin dilaurate catalyst (Figure 1) [13]. The resulting product showed similar gelation structure by SEM as compared to unmodified Gelfoam (Figure 1C-F). To characterizing the radiopacity of RTEA, microCT scans were performed with the setting of 80 kV, 500 μA in 650 ms exposure time. The RTEA produced initial Hounsfield Unit (HU) of 3400±217, approximately 50-fold greater than commercial Gelfoam (HU 69±33). After the samples were subjected to 1 week of media washing, the samples were scanned again. The RTEA retained a radiodensity of 640±150 and the control Gelfoam measured 400±42 with the presence of washing media (HU 395±18). Thought the gelatin sponge structure gently disintegrated under the flow, resulting part of the iohexel network losing, the RTEA sample still presented a high enough radiopacity for visualization under fluoroscopy with the presence of blood analogue.

To model vascular gelation of the RTEA, RTEA was introduced into a 50 mm length of polyurethane tubing with a 2 mm inner diameter. Simulated blood was made with a 1:1 ratio of donkey serum and red food coloring. Samples were exposed to the blood analogue for a period of 7 days with a media exchanged performed 10 times daily (Figure 2A-B). For imaging RTEA and controls, a Siemens Inveon microCT Scanner was used. Scans were performed before introducing blood analogue initially, and 7 days after the washing test (Figure 2A-B). Both images from the scans were analyzed in Siemens Inveon Research Workplace 4.0. Cylindroid Regions of Interests (ROIs) of all 4 samples were chosen for defining the average HU value with the same size in the middle of the tubes. ROIs for the samples after washing test were chosen proximal to the flow. The RTEA demonstrated temporary embolic characteristics in these preliminary tests. It is expected that specific degradation characteristics can be incorporated into the final material by optimizing the cross-linking conditions during the material modification. For in vivo application, a pre-occlusion iodine shot was applied to visualize the vasculature. Using a subtraction angiography, the occlusion of the vessel by use of the radiopaque polymer was monitored. The catheter was retreated prior to post-occlusion iodine shot and it that showed the retrograding agent circumventing the occluded vessel (Figure 2 C-F).

IV. Conclusion

An ideal embolic agent would combine the visibility of contrast agents, have the shape-ability and versatility of Gelfoam® and still maintain its efficacy to stop blood flow in the target tissue without causing irreversible ischemia in the tissue in the process. It would allow the physician direct visualization of the actual plug without the need for

accompanying contrast. More precise placement and faster procedures could result from improved visualization of a Gelfoam® like agent. We developed and optimized a translatable, radiopaque temporary embolic agent (RTEA) and characterized its' physico-chemical and biological properties followed by a proof-of-concept pilot in vivo study. Further development of these agents for temporary embolization procedures will also require understanding of their degradation behavior and hemostatic ability. This work will address the challenge associated with identifying the location of Gelfoam® when it is used as a temporary embolic. A radiopaque embolization agent could reduce repeated radiation exposure to the patient, the physician and staff and may also allow subsequent imaging of treated areas of concern with less imaging artifact from permanent metallic plugs. The major innovation of this proposal is the development of a radiopaque temporary embolic. Existing injectable embolic agents (e.g. Gelfoam®) integrated with X-ray absorbing contrast agents (e.g., iohexol) can easily meet all the clinical requirements of efficient delivery to target arteries, rapid setting process without adhesion to catheters, good visibility by x-ray imaging and efficient embolization results.

REFERENCES

[1] Vaidya, S., Tozer, K. R., & Chen, J. (2008, September). An overview of embolic agents. In Seminars in interventional radiology (Vol. 25, No. 3, p. 204). Thieme Medical Publishers.

[2] Pfitzer Inc., "Description of Gelfoam® Absorbable Gelatin Compressed Sponge" February 2012.

[3] Patel A A, Soulen M C. In: Baum S, Pentecost MJ, editor. Abrams' Angiography Interventional Radiology. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2006. Agents for small vessel/tissue embolization and transcatheter tissue ablation. pp. 169–175.

[4] Correll JT, Prentice HR, Wise EC: Biologic investigations of a new absorbable sponge. Surg Gynecol Obstet. 1945; 181:585-589.

[5] Council on Pharmacy and Chemistry: Absorbable Gelatin sponge - new and nonofficial remedies. JAMA. 1947; 135:921.

[6] Treves N: Prophylaxis of postmammectomy lymphedema by the use of GELFOAM laminated rolls. Cancer 1952; 5:73-83.

[7] Barnes AC: The use of gelatin foam sponges in obstetrics and gynecology. Am J Obstet Gynecol 1963; 86:105-107.

[8] Rarig HR: Successful use of gelatin foam sponge in surgical restoration of fertility. Am J Obstet Gynecol. 1963; 86:136.

[9] Higgins, C. B. (1985). Cardiotolerance of iohexol. Survey of experimental evidence. Investigative radiology, 20(1 Suppl), S65-9.

[10] Bettmann, M. A., & Higgins, C. B. (1985). Comparison of an ionic with a nonionic contrast agent for cardiac angiography. Results of a multicenter trial. Investigative radiology, 20(1 Suppl), S70-4.

[11] Dahlstrom, K., Shaw, D. D., Clauss, W., Andrew, E., & Sveen, K. (1985). Summary of US and European intravascular experience with iohexol based on the clinical trial program. Investigative radiology, 20(1 Suppl), S117-21.

[12] General Electric Company, "Omnipaque (iohexol) Injection, Section II- Clinical Pharmacology- Intravascular" Datasheet Reference ID: 4080358, [2017]

[13] Yin, Q., Yap, F. Y., Yin, L., Ma, L., Zhou, Q., Cheng, J. (2013). Poly(iohexol) Nanoparticles As Contrast Agents for in Vivo X-ray Computed Tomography Imaging. Journal of the American Chemical Society, 135(37), 13620-13623. doi:10.1021/ja405196f

U-net based Automatic CBCT-based Liver Tumor Localization using Biomechanical modeling

Xiaokun Huang, You Zhang and Jing Wang

Abstract*— Liver tumor targeting remains a challenge due to its low inherent contrast in the liver region in cone beam computed tomography (CBCT) images. Instead of directly localizing tumor in CBCT images, deformation vector fields (DVFs) can be used to propagate the CT tumor contour to CBCT images for automatic localization. Two dimensional-three dimensional (2D-3D) registration has been used to solve the intra-liver DVFs by matching acquired projection with forward projections of deformed images, for automatic liver localization. To improve the deformation accuracy of low-contrast liver regions, biomechanical modeling has also been proposed to improve the intra-liver DVFs by finite-element-analysis based on boundary conditions defined as the 2D-3D registration-solved DVFs at liver boundary (2D-3D-Bio). However, the 2D-3D DVFs at lower liver boundary are limited in accuracy due to worse contrast as compared to liver upper boundary, which will negatively affect the liver tumor localization results. To improve the intra-liver DVFs accuracy, we should improve the accuracy of lower liver boundary DVFs further. In this work, we applied a U-net based method to improve the lower liver boundary DVFs accuracy. The improved DVFs then served as boundary condition to generate intra-liver DVFs by biomechanical modeling (U-net-Bio). The method is evaluated with 32 liver SBRT patients. DICE coefficients calculated between the deformed liver contour and reference liver contour are used to evaluate the boundary DVFs accuracy. Tumor center-of-mass error (CoME) between deformed tumor contour and reference tumor contour from physician contouring is used to evaluate the accuracy of intra-liver DVFs. The CoMEs of 2D-3D, 2D-3D-Bio and U-net-Bio are 5.8 mm, 2.9 mm and 2.0 mm respectively. The proposed method can improve the accuracy of liver boundary DVFs accuracy and further improve intra-liver tumor localization.

***Index Terms*—4D-CBCT, Tumor localization, Liver, Biomechanical modeling**

This work was supported in part by US National Institutes of Health (R01 EB020366 and R01 362 EB027898) and by the Cancer Prevention and Research Institute of Texas (RP160661).

X. Huang. BS, Y. Zhang. Ph.D, and J. Wang .Ph.D are with the Department of Radiation Oncology, University of Texas Southwestern, Medical Center, Dallas, TX 75390 USA (e-mail: Xiaokun.Huang@UTSouthwestern.edu; You.Zhang@UTSouthwestern.edu; Jing.Wang@UTSouthwestern.edu).

X. Huang and Y. Zhang are co-first authors.

I. Introduction

Liver cancer has increasing incidences and mortalities globally [1]. Radiation therapy has been applied to treat un-resectable liver tumor and has advantages over other treatments such as surgery or radiofrequency ablation in tumor control rate [2]. Stereotactic body radiotherapy (SBRT), characterized by delivering high doses in a single or few fractions, has been investigated for liver tumor treatment with encouraging and promising tumor local control [3-5]. Since liver normal tissues have low tolerance to radiation, it is important to narrow the margins added to the tumor to spare more normal tissues to allow dose escalation while maintaining low toxicity rate. However, liver tumor localization remains a challenge due to a lack of inherent contrast between tumors and normal liver tissues in cone beam computed tomography (CBCT) images. Respiration involved motion brings artifacts to CBCT images and further affects the tumor targeting accuracy, which necessitates large margins.

Four-dimensional CBCT (4D-CBCT) plays an increasingly important role in motion management by offering on-board time-resolved volumetric images. Instead of delineating tumor location in 4D-CBCT images of each phase through the whole respiration cycle, a motion model can be applied in the form of deformation vector fields (DVFs) which contains the displacement of voxels to propagate liver tumor contour from planning CT to CBCT images for automatic localization. People have investigated finite-element-analysis-based biomechanical modeling to improve the intra-organ motion model accuracy with boundary conditions defined as the DVFs along the organ boundary. The DVFs in the organ boundary are solved using two dimensional-three dimensional (2D-3D) deformable registrations by matching the acquired projection with digitally reconstructed radiographs (DRRs) of the deformed image [6, 7]. This method can solve DVFs accurately in the projection around high-contrast regions while the performance would be degraded where the contrast is low. Therefore, DVFs of liver lower boundary is hard to be accurately solved by 2D-3D deformable registration due to low contrast between liver and surrounding lower-abdominal organs, compared with DVFs in liver upper boundary region. Errors in the lower boundary DVFs would be propagated to intra-organ motion model by biomechanical modeling.

To improve the accuracy of liver motion modeling, previous

research studies applied principle component analysis (PCA) based method to find the correlation between lower and upper liver boundary DVFs to correct DVFs of liver lower boundary using the correlation [8]. However, PCA-based method can only generate patient-specific correlation from patient-specific 4D-CT images, which cannot be generalizable. The 4D-CT may also not be available for all liver cancer patients. Since the convolutional neural network (CNN) has advantages in solving complex non-linear functions, we proposed to use CNN to find a generalized solution for lower-liver DVF correction. The updated DVFs can serve as the boundary condition in the biomechanical modeling to derive and improve the intra-liver DVFs.

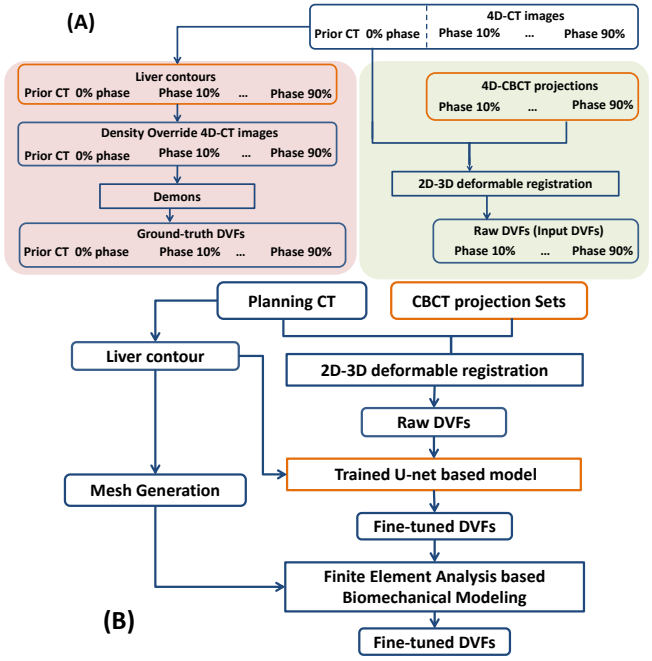


Fig. 1. (A) The Input and ground-truth DVFs generation. Green part shows the process of generating input raw DVFs. Red part shows the process of obtaining ground-truth DVFs. (B) Workflow of Automatic liver tumor localization through CNN based method and Biomechanical modeling.

II. METHOD

Figure 1 showed the workflow of the proposed method. We simulated 4D-CBCT projections from 4D-CT. We performed 2D-3D deformable registration to obtain the DVFs between reference phase and other respiration phases. We selected the end-of-exhale (EE) phase as the reference phase. The DVFs generally have low accuracy around the lower liver boundary. These DVFs obtained through 2D-3D registration served as the input for CNN-based DVFs optimization to obtain fine-tuned DVFs. The fine-tuned DVFs with corrected DVFs in the liver lower boundary can then serve as the boundary condition to derive intra-liver motion model through biomechanical modeling.

A. 2D-3D Deformation registration

The 2D-3D deformation registration method solves the DVFs by matching the measured projections from CBCT with DRRs simulated from deformed volumetric images. DVFs are solved by minimizing the objective function (f) which is defined below:

$$f(D^{1\to t}) = \langle \|\boldsymbol{P}^t - A\mu^t(x+\boldsymbol{D}^{1\to t})\|_{l_2}^2 + \|\boldsymbol{P}^1 - A\mu^1(x+\boldsymbol{D}^{t\to 1})\|_{l_2}^2 + \beta[\phi(\boldsymbol{D}^{1\to t}) + \phi(\boldsymbol{D}^{t\to 1})]\rangle \quad (1)$$

In the Eq. 1, the $\boldsymbol{D}^{1\to t}$ denotes the forward DVF matrix and $\boldsymbol{D}^{t\to 1}$ denotes the inverse DVF matrix. The $\boldsymbol{P}^t$ denotes the CBCT acquired projection at phase t while the $\boldsymbol{P}^1$ denotes the projection at the selected reference phase (phase 1). μ^t denotes the voxel-wise attenuation coefficient value of CBCT image at phase t. A denotes the projection system matrix. DVFs are used to deform the image from reference coordinate x to other phase coordinate $(x+\boldsymbol{D}^{1\to t})$. The first term and second term in the right of equal sign is the data fidelity term, calculating the Euclidian norm of the difference between measure projection ($\boldsymbol{P}^t$) and DRR $(A\mu^t(x+\boldsymbol{D}^{1\to t})$), considering the inverse-consistency.

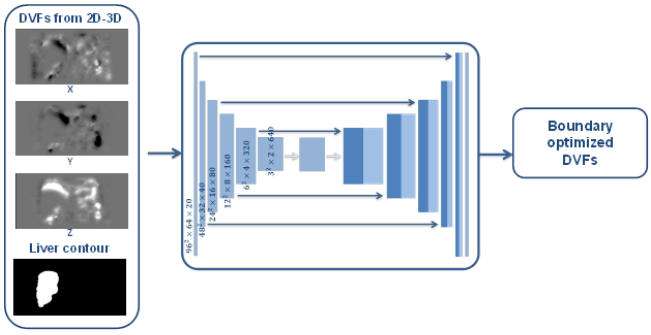


Fig. 2. U-net structure

The objective function introduced a trade-off between data-fidelity and DVFs smoothness by introducing a smoothness regularization term. We can solve the DVFs with the regularization term to achieve fast and stable convergence.

B. U-net based DVFs optimization

We deployed U-net based method to improve the accuracy of the liver lower boundary DVFs solved by 2D-3D registration. The structure of U-net is shown in Figure 2.

The input contains 4 channels and each channel is a 3D volume. Three of them are raw DVFs of each direction (anterior-posterior, left-right and superior-inferior) obtained through 2D-3D deformable registration. The fourth channel is the liver contour of the reference phase image. We chose the fourth channel as liver contour over liver CBCT image considering that the intensity variation of CBCT images within liver is small and liver contour can bring emphasis in the liver boundary regions. We contoured liver of each phase in the 4D-CT images through the Eclipse™ Treatment Planning System (Varian Medical Systems, Palo Alto, CA, USA). We overrode the liver volumes with voxel values 10 times of the original values to generate density-overridden 4D-CT to enhance the contrast along the liver boundary. Then we applied Demons [9] registration method between reference phase and other respiration phases to generate the inter-phase motion model which works as the "ground-truth" to train CNN based DVFs optimization in a supervised fashion. We

included the liver contour enable DVF optimization along the lower liver boundary. The U-net structure contains 5 blocks on both contracting and expansive paths. In the contacting path, each block contains 2 convolutional layers, each followed by a rectified linear unit (ReLU) layer, and a max-pooling layer for

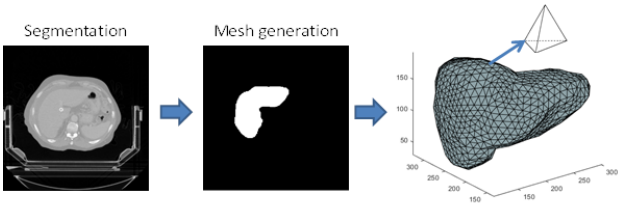


Fig. 3. Mesh generation for biomechanical modeling

down-sampling. In the expansive path, each block contains an up-sampling layer and 2 "up-convolutional layer". The initial filter number is 16. The batch size is 4. The output of the U-net method contains 3 channels, denoting optimized DVFs in each direction. Input and output channels had same size. We choose mean-absolute-error defined in the liver boundary region as the loss function for the model, since we focused on improving the lower boundary DVFs. The initial learning rate is 1e-4. We applied Adam [10] to minimize the loss function.

C. *Biomechanical modeling*

Biomechanical modeling is a physical-properties driven method that has been widely investigated to derive intra-organ motion from displacement-based boundary condition by applying finite element analysis.

We generate tetrahedral mesh based on the reference phase liver contour using open-source Iso2Mesh package [11]. The process for generating the mesh file is shown in Figure 3. We applied the hyper-elastic Mooney-Rivlin material model to describe the liver. The surface DVFs extracted from the output DVFs of the U-net based method works as the boundary condition. We applied the FEBio package to solve the finite element analysis to obtain the in intra-liver DVFs [12].

III. Materials

We used 32 liver SBRT patients with 4D-CT images to train the CNN model and evaluate the proposed method. Twenty-four among 32 patients have 10 phases (from 0% to 90%) 4D-CT, and another seven patients have 3 phases (0%, 50% and 90%). The last case has 2 phases (0% and 50%). 4D-CTs are acquired in different CT scanners with different resolution and size. Therefore, we resized all 4D-CTs into 256×256×128 with resolution $2\times2\times2mm^3$. For each phase, 20 cone-beam projections are simulated using ray-tracing method [4]. Projections are evenly distributed among 360 degrees for the acquisition angle. The simulation source-to-axis distance was 1000 mm and the source-to-imager distance was 1500 mm. A projection contains 512×384 pixels with resolution $0.776\times0.776mm^2$. We took 18 patients for training, 7 for validation and another 7 for testing.

The model is trained on a NVIDIA Tesla v100-sxm2 GPU. We trained the model with fixed 200 epochs first and saved the model in the 200th epoch. The saved model served as the initialization for following training. The training would early-stop when validation loss stops decreasing for 50 epochs. The training process took 40-50 hours. Once finished training the model, the trained model takes 10 seconds to generate improved boundary DVFs for subsequent biomechanical modeling.

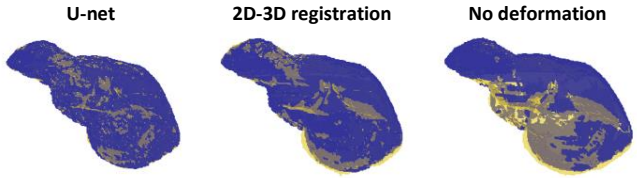


Fig. 4. Visual comparison between deformed liver contour with different methods and "gold-standard" liver contour. The blue mask denotes the gold-standard liver contour in EI phase. The yellow masks, from left to right, denote deformed liver contour from reference phase to EI phase through DVFs generated by U-net based method, 2D-3D registration and liver contour with no DVFs applied.

IV. Results

We hypothesized that the proposed method can improve the liver boundary DVFs accuracy. We deformed the reference phase liver contour to other respiration phases using DVFs obtained from 2D-3D registration and the proposed U-net based method and made visual comparison between the deformed liver contours and the "gold-standard" liver contour from physician. We showed the visual comparison results of 1 test patient in Figure 4 by overlaying the deformed contour with "gold-standard" contour in end-of-inhale phase where the respiration-induced motion was largest.

TABLE I
DICE COEFFICIENTS CALCULATED BETWEEN DEFORMED LIVER CONTOURS WITH DIFFERENT METHODS (2D-3D DEFORMATION REGISTRATION AND U-NET BASED METHOD) AND "GOLD-STANDARD" LIVER CONTOUR FOR EACH PHASE. NO DEFORMATION COLUMN DENOTES NO DVFS APPLIED TO REFERENCE PHASE LIVER CONTOUR. THE RESULTS ARE AVERAGED OVER 7 TEST PATIENTS.

Phase	2D-3D registration	U-net	No deformation
2	0.972	0.979	0.964
3	0.971	0.974	0.960
4	0.964	0.967	0.952
5	0.958	0.963	0.944
6 (EI phase)	**0.955**	**0.960**	**0.935**
7	0.954	0.959	0.931
8	0.960	0.963	0.949
9	0.966	0.968	0.954
10	**0.971**	**0.975**	**0.961**

We can see in the Figure 4 that deformed liver contour using DVFs from U-net based method matched better in the lower liver boundary with "gold-standard" liver contour, compared with deformed liver contour using DVFs from 2D-3D registration. The overlay comparison demonstrated that the proposed U-net based method can improve the DVFs accuracy especially in lower liver boundary.

We quantified above comparison by calculating the DICE coefficient [13] between deformed liver contour and "gold-standard" liver contour for each respiration phase. The quantitative comparison between 2D-3D deformation

TABLE II
TUMOR CENTER-OF-MASS ERRORS OF DIFFERENT METHODS (UNIT: MM). THE RESULTS ARE AVERAGED OVER 7 TEST PATIENTS.

Phase	2	3	4	5	6	7	8	9	10
2D-3D	1.0 ± 0.6	2.4 ± 2.0	4.3 ± 2.5	5.7 ± 2.3	6.3 ± 2.3	6.5 ± 2.3	5.9 ± 2.4	4.3 ± 2.2	2.0 ± 1.2
2D-3D-Bio	0.8 ± 0.3	2.0 ± 1.1	2.7 ± 1.9	3.2 ± 1.8	3.5 ± 1.7	3.6 ± 1.6	3.3 ± 1.7	2.4 ± 1.6	1.4 ± 1.3
U-net-Bio	1.0 ± 0.5	1.3 ± 0.8	2.0 ± 1.0	2.2 ± 1.4	2.3 ± 1.0	2.4 ± 1.0	2.4 ± 1.2	1.7 ± 1.0	1.3 ± 0.9

registration method and U-net based method are shown in the Table 1.

As shown in Table 1, the averaged DICE coefficients for no deformation, 2D-3D registration method and U-net based method in EI phase are 0.935, 0.955 and 0.960, respectively. U-net based method can provide high accuracy DVFs compared with 2D-3D registration method.

We derived the intra-liver DVFs through biomechanical modeling with boundary condition extracted from DVFs generated by U-net based method (U-net-Bio) and 2D-3D registration method (2D-3D-Bio) separately. Since the goal of our project is to automatically localize liver tumor in CBCT, we evaluated the performance of tumor localization accuracy by calculating the tumor center-of-mass-errors (CoMEs) which is defined as the magnitude of the difference vector between center-of-mass of "gold-standard" tumor contour from physician and center-of-mass of deformed tumor contour. The results are shown in Table 2.

The averaged (± SD) tumor CoMEs in EI phase (phase 6) for 2D-3D, 2D-3D-Bio and U-net-Bio are 6.3 (± 2.3) mm, 3.5 (± 1.7) mm and 2.3 (± 1.0) mm, respectively. U-net-Bio can improve the intra-liver DVFs accuracy and achieve better liver tumor localization with smaller tumor CoMEs compared with 2D-3D and 2D-3D-Bio.

V. CONCLUSION

The proposed method applied CNN based method to learn the correlation between liver upper and lower boundary to improve the lower liver boundary DVFs accuracy. Combined deep learning and biomechanical modeling can improve liver tumor localization in low-contrast CBCT images close to 2 mm accuracy.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, pp. 394-424, 2018.

[2] A. Mahadevan, O. Blanck, R. Lanciano, A. Peddada, S. Sundararaman, D. D'Ambrosio, *et al.*, "Stereotactic body radiotherapy (SBRT) for liver metastasis–clinical outcomes from the international multi-institutional RSSearch® Patient Registry," *Radiation Oncology*, vol. 13, p. 26, 2018.

[3] T. E. Schefter, B. D. Kavanagh, R. D. Timmerman, H. R. Cardenes, A. Baron, and L. E. Gaspar, "A phase I trial of stereotactic body radiation therapy (SBRT) for liver metastases," *International Journal of Radiation Oncology* Biology* Physics*, vol. 62, pp. 1371-1378, 2005.

[4] A. E. van der Pool, A. Mendez Romero, W. Wunderink, B. J. Heijmen, P. C. Levendag, C. Verhoef, *et al.*, "Stereotactic body radiation therapy for colorectal liver metastases," *British Journal of Surgery: Incorporating European Journal of Surgery and Swiss Surgery*, vol. 97, pp. 377-382, 2010.

[5] C. Vautravers-Dewas, S. Dewas, F. Bonodeau, A. Adenis, T. Lacornerie, N. Penel, *et al.*, "Image-guided robotic stereotactic body radiation therapy for liver metastases: is there a dose response relationship?," *International Journal of Radiation Oncology* Biology* Physics*, vol. 81, pp. e39-e47, 2011.

[6] D. Yang, H. Li, D. A. Low, J. O. Deasy, and I. El Naqa, "A fast inverse consistent deformable image registration method based on symmetric optical flow computation," *Physics in medicine and biology*, vol. 53, p. 6143, 2008.

[7] J. Wang and X. Gu, "Simultaneous motion estimation and image reconstruction (SMEIR) for 4D cone-beam CT," *Medical physics*, vol. 40, 2013.

[8] Y. Zhang, M. R. Folkert, X. Huang, L. Ren, J. Meyer, J. N. Tehrani, *et al.*, "Enhancing liver tumor localization accuracy by prior-knowledge-guided motion modeling and a biomechanical model," *Quantitative imaging in medicine and surgery*, vol. 9, p. 1337, 2019.

[9] X. Gu, H. Pan, Y. Liang, R. Castillo, D. Yang, D. Choi, *et al.*, "Implementation and evaluation of various demons deformable image registration algorithms on a GPU," *Physics in medicine and biology*, vol. 55, p. 207, 2009.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[11] Q. Fang and D. A. Boas, "Tetrahedral mesh generation from volumetric binary and grayscale images," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, 2009, pp. 1142-1145.

[12] S. A. Maas, B. J. Ellis, G. A. Ateshian, and J. A. Weiss, "FEBio: finite elements for biomechanics," *Journal of biomechanical engineering*, vol. 134, p. 011005, 2012.

[13] T. J. Sørensen, *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons*. København: I kommission hos E. Munksgaard, 1948.

Gram Scale Synthesis of Hafnia Nanodots for Detection of Bone Microcracks using Photon Counting CT Imaging

Fatemeh Ostadhossein, Indu Tripathi, Lily Benig, Denae LoBato, Mahdieh Moghiseh, Chiara Lowe, Aamir Raja, Anthony Butler, Raj Panta, Marzieh Anjomrouz, Alex Chernoglazov and Dipanjan Pan*

***Abstract*— Herein, the bone microcrack imaging has been demonstrated using ligand-directed nanoparticles in combination with photon counting spectral CT. We report here a gram-scale synthesis of hafnia (HfO_2) nanoparticle with surface modification by a chelator moiety, i.e. nitrilotriacetic acid (NTA). The feasibility of delineating these nanoparticles from bone and soft tissue of muscle was demonstrated with photon counting spectral CT equipped with advanced detector technology. The ex vivo and in vivo studies pointed to the accumulation of hafnia nanoparticles at microdamage site featuring distinct spectral signal. Due to their small <5 nm size, hafnia nanoparticles were excreted through reticuloendothelial system organs without noticeable aggregation while not triggering any adverse side effects based on histological and liver enzyme function assessments. These preclinical studies highlight the potential of HfO_2 based nanoparticle contrast agents for skeletal system diseases due to their well-placed K-edge binding energy.**

I. Introduction

The assessment of the bone microcracks relies on the invasive histomorphometry in which the 2D cross-sections of bone are imaged with fluorochromes specific to the bone minerals [5]. Several clinically used imaging modalities are currently being investigated for the non-invasive three-dimensional assessment of bone microdamage [6]. Among them, computed tomography (CT) imaging is an excellent candidate due to its capability to rapidly provide information on the anatomical features [7]. However, the confounding radiographic attenuation from calcium makes the damage detection difficult in bone [8]. Furthermore, the spatial resolution of conventional CT (approximately 0.5-0.625 mm due to the large detector element size) does not allow the detection of damage on the micrometer scale [9]. A non-invasive imaging technique is thus needed that may enable microdamage to be measured and quantified and provide useful data.

Emerging spectral photon counting CT enables the detection of contrast agents based on their characteristic X-ray attenuation profile and respective energies with minimal background from hard tissues [10]. This powerful technique can resolve the strong X-ray attenuation from calcium, seen in the conventional CT imaging instruments [11]. Moreover, the high spatial and spectral resolution, an enhanced contrast to noise ratio (CNR) with reduced imaging artifacts make this technique suitable for the preclinical and clinical applications offering molecular level information [12]. Hafnium (Hf; Z=72) as an emerging radiopaque material with well-positioned K-edge energy (65.3 keV) has been rarely investigated as a contrast agent for the spectral photon counting CT imaging [13-17].

Herein, we disclose for the first time the synthesis of the ligand-modified sub 5 nm hafnia nanoparticles for the sensitive detection of bone microdamage using spectral photon counting CT (Figure 1). These nanoparticles were directed to bone minerals by the introduction of an multidentate chelating agent i.e. nitrilotriacetic acid. At the site of the microdamage: a fracture in hydroxyapatite, the major bone constituent, resulted in the exposure of freshly charged calcium ions due to the breakage of atomic bonds in the matrix [18]. Hence, the nanoparticles targeted to the more concentrated minerals 'lit up' the microdamage site. It was demonstrated that hafnia nanoparticles can show in 'true color' the location and anatomy of the microdamage using spectral photon counting CT in ex vivo human sample. We indicated in the in vivo studies that these nanoparticles could effectively locate the microcrack anatomy in the rat's tibia while the non-targeted control nanoparticles accumulated in the muscle non-specifically. Furthermore, we demonstrated that intramuscular injection of the nanoparticles did not induce any adverse effects histopathologically, neither did they negatively alter liver enzyme functions. Bio-distributive properties of these nanoparticles revealed their preferential distribution through nutrient transport canals (e.g. Haversian canals, canaliculi, Volkmann's canals, etc.) in the muscle near the bone and microcrack itself without noticeable aggregation. Therefore, these ligand-directed hafnia nanoparticles would open new avenues for the safe, non-destructive, sensitive bone microdamage detection in vivo.

Rsearch supported by University of Illinois, National Institute of Health (NIBIB).

F.O., I.T., L.B., D.L. are with Department of Bioengineering, Beckman Institute of Advanced Science and Technology and Department of Materials Science and Engineering University of Illinois at Urbana-Champaign, Urbana IL 61801 USA.

M.M., C.L., A.R., A.B., R.P., M.A., A.C. are with MARS Bioimaging Limited, University of Otago Christchurch, Christchurch, New Zealand.

*D. P. was with Department of Bioengineering, Beckman Institute of Advanced Science and Technology and Department of Materials Science and Engineering University of Illinois at Urbana-Champaign, Urbana IL 61801 USA and now at Department of Chemical Biochemical and Environmental Engineering and Pediatrics at University of Maryland Baltimore County and University of Maryland Baltimore, MD, USA.

(e-mail: dipanjan@som.umaryland.edu).

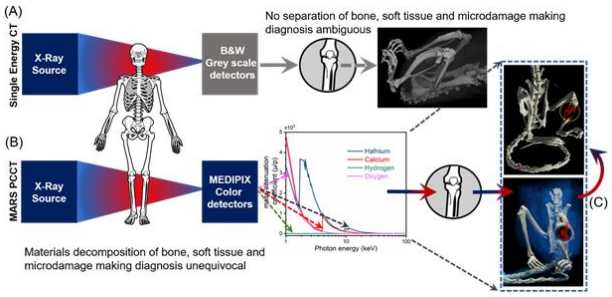


Figure 1. Schematic illustration of the workflow. Bone as a functional material would experience various stresses. For the early detection of microdamages, here, targeted hafnia nanoparticles were designed and used in the photon counting spectral CT. This advanced technology would determine signal from hafnia nanoparticles bound to the hard tissue of bone both in the ex vivo model and in vivo. The highlighted area in C is magic lens of Hf in MARS spectral CT. The diagram of the mass attenuation coefficient vs. photon energy is extracted from NIST database.

MATERIALS AND METHODS

A. *Synthesis of hafnia nanodots*

Briefly, 300 ml cyclohexane, 34.5 g IGEPAL® CO-520, 11.25 ml ethyl alcohol, 3.75 ml NaOH (75 mM) were stirred vigorously (500 rpm) under reflux condition at 65 °C in 500 ml round bottom flask for 30 min to achieve homogeneity. 1.61 g hafnium ethoxide powder was added to the above pot followed immediately with 750 µl water and the reaction progressed overnight during which HfO_2 nanoparticles nucleated in the water-in-oil microemulsion. 7.5 g trimethoxysilylpropyl-modified polyethylenimine was then added dropwise at a rate of 80 µl.min-1 under the mentioned condition and the stirring continued for another 2 h until the turbid suspension became transparent, indicating the reaction completion. The solvent was evaporated in rotary evaporator at 55 °C and the polyethylene imine silane modified nanoparticles were washed with 100 ml of hexane: ethyl ether (50:50 vol/vol) once. The nanoparticles were washed with ethyl ether three times (50 ml each) and were dried in vacuum oven overnight. For nitrilotriacetic acid (NTA) conjugation, (7.18 mg) NTA was activated with 2.63 mg EDC and 1.58 mg NHS for 1 h in 1 ml of dimethylformamide (DMF) and was then added to 200 mg of non-targeted nanoparticles dispersed in 9 ml DMF under stirring condition at room temperature and was stirred overnight to ensure complete conjugation. The targeted nanoparticles were washed with methyl alcohol repeatedly at 10,000 rpm (at least 4x) and then were dialyzed after dispersing in water by sonication.

B. *In vivo studies with MARS Photon Counting CT.*

All the animal experiments were carried out according to the ethical guidelines outlined by the Illinois Institutional Animal Care and Use Committee at University of Illinois Urbana Champaign and were approved by the board (protocol #17125). For the imaging experiment, 7 five-month-old female Sprague Dawley rates were divided into three groups; 1) control sham operated (n=1); 2) non-targeted nanoparticles (n=3); 3) targeted nanoparticles (n=3). The number per group was chosen to minimize the overall number of the animals. The fur was shaved around the hind limbs and then an incision was made to expose the tibia. A bone damage (1-2 cm long x 0.2 cm deep) was created longitudinally in the tibia with a surgical scalpel. The surgery site was sutured and then nanoparticles (100 µl, 259 mM Hf) were injected intramuscularly in the muscle surrounding the tibia. Their wounds were taken care of by applying sterile gas, iodophor and injecting carprofen (1.5 mg.kg-1, SC) at 12 h intervals for the first 24 h post-surgery to minimize pain. During the surgery the rats were anesthetized using isoflurane/oxygen.

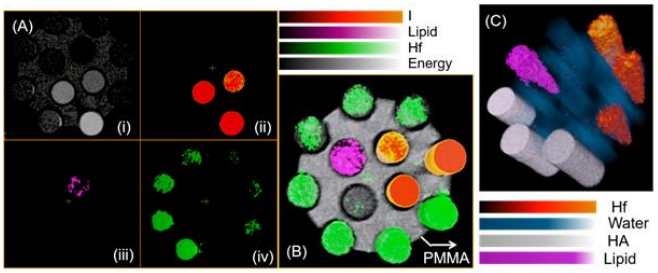


Figure 2. Targeted hafnia nanoparticles with iohexol scan-material images (i) The scanned energy channel is 30-45 keV; (ii) iodine channel; (iii) lipid channel; (iv) Hf channel; (B) fused and 3D reconstructed image of the phantom. The corresponding axial slices of the phantom containing iohexol, hafnium, water, and lipid at different energy bins have been indicated in. (C) 3D reconstruction of the second phantom.

The animals were imaged with General Electric Light speed 16-slice CT scanner (GE, Milwaukee, WI, USA) using the following parameters: slice thickness: 0.625 mm; tube voltage: 80 kV; tube current: 100 mA; gantry rotation time: 0.8 s; and pitch: 9.38 mm, 30 min and 5 h post injection. The images were processed on 3DSlicer software.

For photon counting spectral CT, adjustments were made to MARS scanner to accommodate the larger rat specimen and the scanning FOV was increased to 68 mm. Furthermore, SDD and SOD were increased to 281.6 mm and 215 mm, respectively. To achieve an equivalent photon count, the tube current was increased to 25 µA and the exposure time decreased to 150 ms. Energy filtration and energy thresholds/bins remained unchanged compared to the ex vivo sample.

II. RESULTS AND DISCUSSION

We have utilized the ligand-directed ultra-small HfO_2 (hafnia) nanoparticles for the sensitive detection of bone microdamage ex vivo and in vivo. Nanoparticle core was synthesized by a sol-gel chemistry process and NTA was introduced by a facile EDC-NHS chemistry. Physicochemical properties and the conjugation process were investigated using various analytical techniques. These nanoparticles showed the response in the spectral CT imaging in the corresponding K-edge of Hf which was properly located within the bone microdamage site. The quantification was also achieved in the spectral CT and the calculated values were commensurate with the analytically measured concentrations. The decomposition algorithms

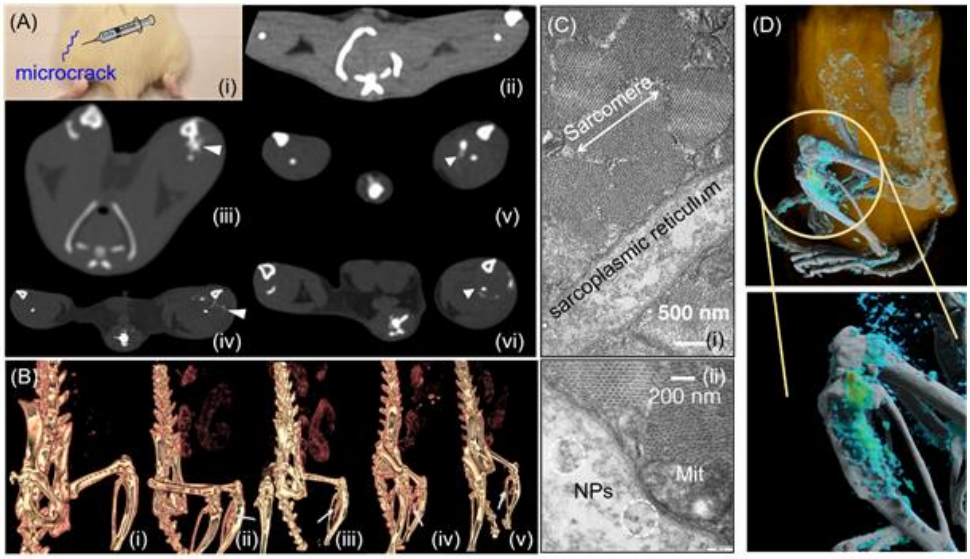


Figure 3. In vivo investigation of the targeted nanoparticles for bone microdamage imaging with conventional CT. (A) (i) the microdamage was induced in the hind limb of Sprague Dawley rat by scalpel and the nanoparticles were injected intramuscularly after suturing and subsequently were scanned, The conventional CT images for (ii) control sham operated animal, (iii) non-targeted nanoparticles 30 min, (iv) non-targeted nanoparticles 5 h, (v) targeted nanoparticles 30 min, (vi) targeted nanoparticles 5 h, the arrows point to the accumulation of hafnia nanoparticles and its contrast against bone; (B) The 3D reconstructed image of (A) (i) control sham operated animal, (ii) non-targeted nanoparticles 30 min, (iii) non-targeted nanoparticles 5 h, (iv) targeted nanoparticles 30 min, (v) targeted nanoparticles 5 h; (C) The TEM image of the harvested muscle at the injection site, No significant aggregation of the nanoparticles is perceived even in the biological tissues. Mitochondria, sarcomere, sarcoplasmic reticulum and nanoparticles are marked on the image, (i) and (ii) are the same with different magnification. The in vivo evaluation of the nanoparticles for bone microdamage imaging using spectral CT; (D) The nanoparticles based on Hf can be distinguished from bone and soft tissue and the signal from Hf can correctly point to the site of the surgery. Cyan-green representing high concentrations of Hf, blue representing low concentrations and the calcium channel in white.

could accurately differentiate between Hf, lipid, air and HA in the phantom studies. (Figure 2) Ex vivo human bone microdamage model was subsequently used to confirm the binding of the nanoparticles to the crack site while Hf could be distinguished from bone itself. Finally, the imaging of the animals with the induced microdamage and injected with the targeted nanoparticles revealed the accumulation in the region of interest and the signal from soft tissue, bone and Hf in microcrack could be separated. (Figure 3) Due to their high safety as verified histopathologically, their small size, stability in the biological media, and their favorable imaging properties, these HfO_2 based nanoparticles could potentially be used for the anatomical evaluations in the future.

III. Conclusion

Our results indicated that the quantification by MARS photon counting technique can reveal the ratio of the nanoparticles accumulated in the microdamage limb over the healthy limb was higher in the targeted nanoparticles compared with the non-targeted group (258% increase in the total amount of Hf and 27% increase in the max concentration of Hf) therefore pointing to the targeting capability of these nanoparticles. Materials decomposition image reconstruction algorithm presented the targeted hafnia nanoparticle with a high signal-to-noise ratio successfully separated from background noise. Bone as a functional material would experience various stresses which can lead to microdamages. Targeted hafnia nanoparticles were designed and used in the photon counting spectral CT for the detection of the microcracks. This advanced technology would determine signal from hafnia nanoparticles bound to the hard tissue of bone both in the ex vivo model and in vivo.

References

[1] [1] a) S. Nagaraja, T. L. Couse, R. E. Guldberg, Journal of biomechanics 2005, 38, 707; b) M. R. Allen, D. B. Burr, Bone 2011, 49, 56.

[2] a) D. Taylor, J. G. Hazenberg, T. C. Lee, Nature materials 2007, 6, 263; b) S. Mohsin, F. O'Brien, T. Lee, Journal of anatomy 2006, 209, 119; c) F. Barthelat, Z. Yin, M. J. Buehler, Nature Reviews Materials 2016, 1, 16007.

[3] a) C. Acevedo, V. A. Stadelmann, D. P. Pioletti, T. Alliston, R. O. Ritchie, Nature Biomedical Engineering 2018, 1; b) S. Colopy, J. Benz-Dean, J. Barrett, S. Sample, Y. Lu, N. Danova, V. Kalscheur, R. Vanderby Jr, M. Markel, P. Muir, Bone 2004, 35, 881.

[4] F. M. Lambers, G. Kuhn, R. Müller, BoneKEy reports 2012, 1.

[5] S. Haupert, S. Guerard, F. Peyrin, D. Mitton, P. Laugier, PLOS one 2014, 9, e83599.

[6] a) F. Ostadhossein, L. Benig, I. Tripathi, S. K. Misra, D. Pan, ACS applied materials & interfaces 2018; b) M. Muller, D. Mitton, M. Talmant, P. Johnson, P. Laugier, Journal of Biomechanics 2008, 41, 1062; c) A.-I. Henry, B. Sharma, M. F. Cardinal, D. Kurouski, R. P. Van Duyne, Analytical chemistry 2016, 88, 6638; d) C. Jiang, Y. Wang, J. Wang, W. Song, L. Lu, Biomaterials 2017, 114, 54.

[7] a) B. M. Yeh, P. F. FitzGerald, P. M. Edic, J. W. Lambert, R. E. Colborn, M. E. Marino, P. M. Evans, J. C. Roberts, Z. J. Wang, M. J. Wong, Advanced drug delivery reviews 2017, 113, 201; b) F. Ostadhossein, S. K. Misra, I. Tripathi, V. Kravchuk, G. Vulugundam, D. LoBato, L. E. Selmic, D. Pan, Biomaterials 2018, 181, 252.

[8] H. W. Goo, J. M. Goo, Korean journal of radiology 2017, 18, 555.

[9] E. Lin, A. Alessio, Journal of cardiovascular computed tomography 2009, 3, 403.

[10] a) D. Pan, C. O. Schirra, S. A. Wickline, G. M. Lanza, Contrast media

& molecular imaging 2014, 9, 13; b) C. O. Schirra, E. Roessl, T. Koehler, B. Brendel, A. Thran, D. Pan, M. A. Anastasio, R. Proksa, IEEE transactions on medical imaging 2013, 32, 1249; c) C. O. Schirra, A. Senpan, E. Roessl, A. Thran, A. J. Stacy, L. Wu, R. Proksa, D. Pan, Journal of materials chemistry 2012, 22, 23071; d) R. Panta, S. T. Bell, J. Healy, R. Aamir, C. Bateman, M. Moghiseh, A. Butler, N. Anderson, Journal of Instrumentation 2018, 13, T02001; e) M. Moghiseh, R. Aamir, R. K. Panta, N. de Ruiter, A. Chernoglazov, J. Healy, A. Butler, N. Anderson, JSM Biomed Imaging Data Pap 2016, 61, 1007.

[11] D. Pan, T. A. Williams, A. Senpan, J. S. Allen, M. J. Scott, P. J. Gaffney, S. A. Wickline, G. M. Lanza, Journal of the American Chemical Society 2009, 131, 15522.

[12] a) N. G. Anderson, A. P. Butler, Contrast media & molecular imaging 2014, 9, 3; b) C. H. McCollough, S. Leng, L. Yu, J. G. Fletcher, Radiology 2015, 276, 637; c) D. Pan, C. O. Schirra, A. Senpan, A. H. Schmieder, A. J. Stacy, E. Roessl, A. Thran, S. A. Wickline, R. Proska, G. M. Lanza, ACS nano 2012, 6, 3364; d) D. Pan, A. H. Schmieder, A. SenPan, X. Yang, S. A. Wickline, E. Roessl, R. Proksa, C. O. Schirra, G. M. Lanza, in Design and Applications of Nanoparticles in Biomedical Imaging, Springer 2017, p. 385; e) D. Pan, E. Roessl, J. P. Schlomka, S. D. Caruthers, A. Senpan, M. J. Scott, J. S. Allen, H. Zhang, G. Hu, P. J. Gaffney, Angewandte Chemie International Edition 2010, 49, 9635.

[13] P. F. FitzGerald, R. E. Colborn, P. M. Edic, J. W. Lambert, A. S. Torres, P. J. Bonitatibus Jr, B. M. Yeh, Radiology 2015, 278, 723.

[14] J. Kim, D. Bar-Ness, S. Si-Mohamed, P. Coulon, I. Blevis, P. Douek, D. P. Cormode, Scientific reports 2018, 8, 12119.

[15] T. Nowak, M. Hupfer, R. Brauweiler, F. Eisa, W. A. Kalender, Medical physics 2011, 38, 6469.

[16] T. L. McGinnity, O. Dominguez, T. E. Curtis, P. D. Nallathamby, A. J. Hoffman, R. K. Roeder, Nanoscale 2016, 8, 13627.

[17] a) S. Bonvalot, P. Rutkowski, J. Thariat, S. Carrere, M.-P. Sunyach, E. Saada, P. Agoston, A. Hong, A. Mervoyer, M. Rastrelli, Annals of Oncology 2018, 29, mdy424. 081; b) P. Zhang, S. Vivet, S. Paris, A. Pottier, AACR, 2017; c) C. Le Tourneau, J. Thariat, E. C. Rodriguez, P. Sargos, C. Le Pechoux, G. Kantor, C. Hoffmann, V. M. Garcia, T. De Baere, V. Vendrely, International Journal of Radiation Oncology• Biology• Physics 2018, 102, e185.

[18] T. C. Lee, S. Mohsin, D. Taylor, R. Parkesh, T. Gunnlaugsson, F. J. O'Brien, M. Giehl, W. Gowin, Journal of Anatomy 2003, 203, 161

Three-Material Dual Energy Decomposition Using A Constrained Model-Based Algorithm

Stephen Z. Liu, Qian Cao, Jeffrey H. Siewerdsen, J. Webster Stayman, and Wojciech Zbijewski*

***Abstract*—Conventionally, Dual Energy CT (DE CT) provides a decomposition of the object into two base materials. However, in many clinical application, more base materials are needed to adequately represent the object. A volume constraint is often included in the decomposition – i.e. the sum of volume fractions of the base materials is assumed to be one in each voxel. In this work, we propose an algorithm that considers the volume constraint in model-based material decomposition (constrained MBMD, C-MBMD) to enable estimation of concentrations of three base material directly from DE projection data (without the need for coinciding projection rays or intermediate reconstructions of energy channels). The algorithm uses a penalized-weighted least-squares (PWLS) objective function with bound constraints and is optimized using a gradient projection approach. This algorithm is applied to quantitative DE imaging of bone marrow edema (BME), where we seek the concentrations of soft tissue, bone mineral and fat within the trabecular bone. We tested the proposed C-MBMD in a kVp-switching scenario using two digital phantoms containing water-bone-fat mixtures. The algorithm generated ~10% quantitative errors for all material estimates, comparing favorably to >20% errors using the classic image-domain decomposition (IDD). Compared to IDD, C-MBMD also demonstrated reduced noise, beam hardening-induced inaccuracy, and crosstalk between soft tissue and fat.**

***Index Terms*—Dual energy computed tomography, multi-material decomposition, model-based reconstruction, constrained optimization, extremity imaging, bone marrow edema.**

I. Introduction

We have previously reported a Model-Based Material Decomposition (MBMD) algorithm [1] that generates dual energy (DE) material concentration images directly from projection data. MBMD does not rely on the coinciding rays in the high and low energy data (required by the conventional projection-domain decomposition) or intermediate reconstructions of the energy channels (required by image-domain decomposition, IDD). It has been shown to outperform IDD in both image quality and quantitative accuracy for two-material DE decompositions [1]. However, the object model in the original formulation of MBMD allows only two base materials, consistent with the number of available energy channels. In this work, we consider the extension of MBMD to three base materials. To this end, we utilize the approach conventionally deployed in three-material IDD: a volume constraint [2,3], requiring that the sum of volume fractions of the base materials (each bounded between zero and one) in each voxel adds to one.

Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205. *Email: wzbijewski@jhu.edu

DE CT has substantial applications in evaluating bone compositions. For example, bone marrow edema (BME) is an important biomarker for rheumatoid arthritis and osteoarthritis [4]. BME presents as increased water-like signals in marrow compartment of trabecular bone. It is typically detected by MRI, but there are also great efforts in translating this biomarker into DE CT [5,6]. DE imaging of BME has been demonstrated by IDD with soft-tissue (water), bone mineral (calcium) and fat as base materials [6]. In this work, we investigate whether the proposed constrained MBMD (C-MBMD) method can achieve improved quantitative DE imaging of BME compared to IDD.

In the following sections, we present the mathematical formulation of a constrained objective function and optimization approaches used in C-MBMD for three-material DE decompositions. The algorithm was validated using simulated DE data of digital phantoms containing mixtures of water, bone and fat that mimicked clinical DE imaging of BME.

II. Method

A. Objective Function and Forward Model for C-MBMD

The constrained three-material decomposition from DE data is performed as the optimization of a penalized-weighted least-square (PWLS) objective function:

$$\hat{\pi} = \arg\min_{f_*}(y-\bar{y})^T\mathbf{K}^{-1}(y-\bar{y}) + R(\pi)$$
$$\text{s.t.} \sum_{\kappa=1}^{3}\pi_{j,\kappa} = c, b_l \le \pi_{j,\kappa} \le b_u \quad (1)$$

where $\pi_{j,\kappa}$ is the κth material volume fraction in the jth voxel, y is the vector of DE measurements, $\mathbf{K}$ is the covariance matrix, and R is the penalty. $\bar{y}$ denotes the forward model written as:

$$\bar{y}_i = \sum_{\epsilon=1}^{\mathrm{E}} S_{i,\epsilon}\Delta\epsilon\exp\left(\sum_{\kappa=1}^{3}-\mathbf{A}_i\mu_{\epsilon,\kappa}\tilde{\rho}_\kappa\pi_\kappa\right) \quad (2)$$

where $\bar{y}_i$ is the ith x-ray passing through the object, $S_{i,\epsilon}$ is the spectral response at the energy index ϵ, and $\mathbf{A}_i$ is the ith row of the projection matrix. $\mu_{\epsilon,\kappa}$ and $\tilde{\rho}_\kappa$ represent material mass attenuations and nominal densities of the κth basis, respectively.

The optimization problem in (1) is constrained by i) material fractions summing up to a constant c, and ii) each material fraction staying between a lower (b_l) and an upper (b_u) bounds. We modify the problem by merging the equality constraint into the

forward model in (2):

$$\bar{y}_i = \sum_{\epsilon=1}^{\mathrm{E}} S_{i,\epsilon}\Delta\epsilon\exp\left(-\mathbf{A}_i\begin{pmatrix}\mu_{\epsilon,1}\tilde{\rho}_1\\ \mu_{\epsilon,2}\tilde{\rho}_2\\ \mu_{\epsilon,3}\tilde{\rho}_3\end{pmatrix}^T\begin{pmatrix}\pi_1\\ \pi_2\\ c-\pi_1-\pi_2\end{pmatrix}\right) \quad (3)$$

After rearrangement, we obtain:

$$\bar{y}_i = \sum_{\epsilon=1}^{\mathrm{E}} S_{i,\epsilon}\Delta\epsilon\exp(-\mathbf{A}_i z_3 c)\exp\left(-\mathbf{A}_i\begin{pmatrix}z_1\\ z_2\end{pmatrix}^T\begin{pmatrix}\pi_1\\ \pi_2\end{pmatrix}\right) \quad (4)$$

$$\begin{cases} z_1 := \mu_{\epsilon,1}\tilde{\rho}_1 - \mu_{\epsilon,3}\tilde{\rho}_3, \quad z_2 := \mu_{\epsilon,2}\tilde{\rho}_2 - \mu_{\epsilon,3}\tilde{\rho}_3 \\ z_3 := \mu_{\epsilon,3}\tilde{\rho}_3 \end{cases} \quad (5)$$

Note that since c is pre-defined as a constant, (4) can be formulated using matrix representations analogous to the KC-MBMD algorithm [7] reported in Liu et al.

Based on the modification in (4), we can re-write the original constrained optimization problem into:

$$\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2)^T = \arg\min_{\pi}(y-\bar{y})^T\mathbf{K}^{-1}(y-\bar{y}) + R(\pi)$$
$$\text{s.t.} \quad b_l \le \pi_{j,1} + \pi_{j,2}, \pi_{j,1}, \pi_{j,2} \le b_u \quad (6)$$

where we only need to estimate the volume fractions of the two materials, subject to the three bounding constraints. The volume fraction of the third material can then be obtained from $(\hat{\pi}_1, \hat{\pi}_2)$ using the volume constraint.

B. Optimal Solution

We first apply optimization transfer onto (6) using separable parabolic surrogate [8,9] to simplify the task. By doing this, we obtain an objective function in a quadratic form:

$$\hat{\pi} = \arg\min_{\pi}\sum_i\sum_j \zeta + \zeta'\pi_j + \frac{1}{2}\zeta''\pi_j^2$$
$$\text{s.t.} \quad b_l \le \pi_{j,1} + \pi_{j,2}, \pi_{j,1}, \pi_{j,2} \le b_u \quad (7)$$

where the quadratic coefficients are:

$$\begin{cases} \zeta := \dfrac{M_{ij}}{\gamma_i}\left(\phi_i^{(n)}\left(\ell_i^{(n)}\right) - \nabla\phi_i^{(n)}\left(\ell_i^{(n)}\right)\gamma_i\pi_j^{(n)} + \dfrac{1}{2}\nabla^2\phi_i^{(n)}\left(\ell_i^{(n)}\right)\gamma_i^2\left(\pi_j^{(n)}\right)^2\right) + r \\ \zeta' := \dfrac{M_{ij}}{\gamma_i}\left(\nabla\phi_i^{(n)}\left(\ell_i^{(n)}\right)\gamma_i - \nabla^2\phi_i^{(n)}\left(\ell_i^{(n)}\right)\gamma_i^2\pi_j^{(n)}\right) + r' \\ \zeta'' := \dfrac{M_{ij}}{\gamma_i}\left(\nabla^2\phi_{1,i}^{(n)}\left(\ell_i^{(n)}\right)\gamma_i^2\right) + r'' \end{cases} \quad (8)$$

$\pi_j^{(n)}$ denotes the previous estimates of material fractions. r, r' and r'' are coefficients of the parabolic surrogate for the penalty R. Remaining terms are all constants as defined in [9].

To derive the optimal solution, we incorporate the constraints into (7) using Karush-Kuhn-Tucker (KKT) multipliers:

$$\mathcal{L}^{(n)} = \sum_i\sum_j \zeta + \zeta'\pi_j + \frac{1}{2}\zeta''\pi_j^2 + \lambda_0\left(\pi_{j,1}+\pi_{j,2}-b_u\right) - \lambda_1\left(\pi_{j,1}-b_l\right) - \lambda_2\left(\pi_{j,2}-b_l\right) \quad (9)$$

where λ_0, λ_1 and λ_2 are constants. The optimal solution is thus given by the following KKT conditions:

$$\begin{cases} \nabla_{\pi_\kappa}\mathcal{L}_\kappa^{(n)} = \sum_i\sum_j \zeta' + \zeta''\pi_{j,\kappa} - \lambda_\kappa + \lambda_0 = 0 \\ \lambda_\kappa\left(\pi_{j,\kappa} - b_l\right) = 0, \quad \lambda_0\left(\pi_{j,1}+\pi_{j,2}-b_u\right) = 0 \\ \lambda_\kappa \ge 0, \quad \lambda_0 \ge 0 \end{cases} \quad (10)$$

for κ = 1 or 2, where the first equation defines the primal feasibility, the second and third give the complementary slackness, and the last two ensure the dual feasibilities.

C. Projected Gradient and Cauchy Point

Numerous strategies have been reported to solve (7). We employ the gradient projection approach [10] (Fig. 1), which can be derived from the gradient descent for each voxel:

$$\pi_{j,\kappa}^{(n+1)}\left(\alpha_j^{(n)}\right) = \pi_{j,\kappa}^{(n)} - \alpha_j^{(n)}g_{j,\kappa}^{(n)} \quad (11)$$

$$g_{j,\kappa}^{(n)} = \zeta' + \zeta''\pi_{j,\kappa}^{(n)} \quad (12)$$

where $g_j^{(n)}$ is the descent direction at the nth iteration, and $\alpha_j^{(n)}$ is the step length that could be obtained by:

$$\hat{\alpha}_j^{(n)} = \arg\min_{\alpha_j^{(n)}}\sum_i\sum_\kappa \zeta + \zeta'\pi_{j,\kappa}^{(n+1)} + \frac{1}{2}\zeta''\left(\pi_{j,\kappa}^{(n+1)}\right)^2 \quad (13)$$

To add the inequality constraints, we determine a breakpoint step length (BSL) for each material. The BSL is defined as the step that brings the objective exactly onto the closest bound. For the first constraint in (7), the BSL can be computed as:

$$\bar{\alpha}_{j,0} = \begin{cases} \dfrac{\pi_{j,1}^{(n)} + \pi_{j,2}^{(n)} - b_u}{g_{j,1}^{(n)} + g_{j,2}^{(n)}} & \text{if } g_{j,1}^{(n)} + g_{j,2}^{(n)} < 0 \\ \infty & \text{if } g_{j,1}^{(n)} + g_{j,2}^{(n)} = 0 \\ \dfrac{\pi_{j,1}^{(n)} + \pi_{j,2}^{(n)} - b_l}{g_{j,1}^{(n)} + g_{j,2}^{(n)}} & \text{if } g_{j,1}^{(n)} + g_{j,2}^{(n)} > 0 \end{cases} \quad (14)$$

Similarly, the BSL for the last two constraints is:

$$\bar{\alpha}_{j,\kappa} = \begin{cases} \left(\pi_{j,\kappa}^{(n)} - b_u\right)/g_{j,\kappa}^{(n)} & \text{if } g_{j,\kappa}^{(n)} < 0 \\ \infty & \text{if } g_{j,\kappa}^{(n)} = 0 \\ \left(\pi_{j,\kappa}^{(n)} - b_l\right)/g_{j,\kappa}^{(n)} & \text{if } g_{j,\kappa}^{(n)} > 0 \end{cases} \quad (15)$$

for κ = 1 or 2. The BSL satisfying all three constraints $(\tilde{\alpha}_{j,\kappa}^{(n)})$ is thus the smaller value between $\bar{\alpha}_{j,0}$ (14) and $\bar{\alpha}_{j,\kappa}$ (15).

Once the two BSLs are obtained (Fig. 1), we sort them (including 0) increasingly and generate a set of intervals [10]:

$$\mathcal{D}_j^{(n)} = \left\{\left[0, \tilde{\alpha}_{j,\kappa_1}^{(n)}\right], \left[\tilde{\alpha}_{j,\kappa_1}^{(n)}, \tilde{\alpha}_{j,\kappa_2}^{(n)}\right], 0 < \tilde{\alpha}_{j,\kappa_1}^{(n)} < \tilde{\alpha}_{j,\kappa_2}^{(n)}\right\} \quad (16)$$

The optimal step length can be thus obtained by evaluating (13) within each interval. To derive an analytical solution, we define an advancement parameter $\Delta\alpha_j$. Assuming we are at the interval $\left[\tilde{\alpha}_{j,\kappa_1}^{(n)}, \tilde{\alpha}_{j,\kappa_2}^{(n)}\right]$, this parameter can be expressed as:

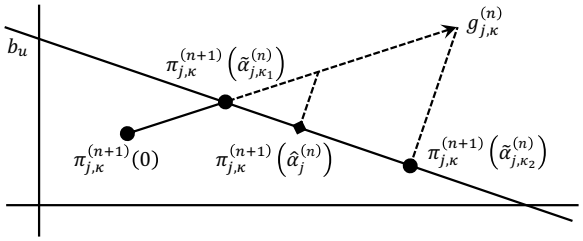


Figure 1. Projected gradient search path for the κth material basis. Arrow represents the descent direction. Connections between circles denote the intervals of sorted BSFs. Square is the material fraction at the optimized step length.

$$\Delta\alpha_j := \alpha_j^{(n)} - \tilde{\alpha}_{j,\kappa_1}^{(n)} \in \left[0, \tilde{\alpha}_{j,\kappa_2}^{(n)} - \tilde{\alpha}_{j,\kappa_1}^{(n)}\right] \quad (17)$$

Then, $\pi_j^{(n+1)}\left(\alpha_j^{(n)}\right)$ in (11) becomes:

$$\pi_j^{(n+1)} = \begin{cases} \pi_j^{(n+1)}\left(\tilde{\alpha}_{j,\kappa_1}^{(n)}\right) - \Delta\alpha_j g_j^{(n)} & \text{if } \tilde{\alpha}_{j,\kappa_1}^{(n)} < \tilde{\alpha}_{j,\kappa}^{(n)} \\ \pi_j^{(n+1)}\left(\tilde{\alpha}_{j,\kappa_1}^{(n)}\right) & \text{if } \tilde{\alpha}_{j,\kappa_1}^{(n)} \geq \tilde{\alpha}_{j,\kappa}^{(n)} \end{cases} \quad (18)$$

By doing this, we convert (13) into an unconstrained minimization with the advancement as the only variable:

$$\Delta\hat{\alpha}_j = \arg\min_{\Delta\alpha_j} \sum_i \sum_\kappa \zeta + \zeta'\left(\pi_{j,\kappa}^{(n+1)}\left(\tilde{\alpha}_{j,\kappa_1}^{(n)}\right) - \Delta\alpha_j g_{j,\kappa}^{(n)}\right) + \frac{1}{2}\zeta''\left(\pi_{j,\kappa}^{(n+1)}\left(\tilde{\alpha}_{j,\kappa_1}^{(n)}\right) - \Delta\alpha_j g_{j,\kappa}^{(n)}\right)^2 \quad (19)$$

After arrangement, we obtain another quadratic objective:

$$\Delta\hat{\alpha}_j = \arg\min_{\Delta\alpha_j} \sum_i \sum_\kappa \sigma + \Delta\alpha_j \sum_i \sum_\kappa \sigma' + \frac{1}{2}{\Delta\alpha_j}^2 \sum_i \sum_\kappa \sigma'' \quad (20)$$

where the coefficients are:

$$\begin{cases} \sigma := \zeta + \zeta'\pi_{j,\kappa}^{(n+1)}\left(\tilde{\alpha}_{j,\kappa_1}^{(n)}\right) + \frac{1}{2}\zeta''\left(\pi_{j,\kappa}^{(n+1)}\left(\tilde{\alpha}_{j,\kappa_1}^{(n)}\right)\right)^2 \\ \sigma' := -\zeta' g_{j,\kappa}^{(n)} - \zeta''\pi_{j,\kappa}^{(n+1)}\left(\tilde{\alpha}_{j,\kappa_1}^{(n)}\right) g_{j,\kappa}^{(n)} \\ \sigma'' := \zeta''\left(g_{j,\kappa}^{(n)}\right)^2 \end{cases} \quad (21)$$

If the result of (20) satisfies (17), then it is a Cauchy point. In this work, we directly apply the Cauchy point as the optimized advancement for computing the step length of the current iteration. Such point can be analytically calculated as:

$$\Delta\hat{\alpha}_j = -\left(\sum_i \sum_\kappa \sigma'\right) \Big/ \left(\sum_i \sum_\kappa \sigma''\right) \quad (22)$$

If (17) is not satisfied, we move on and evaluate (19) within the next interval until KKT condition in (10) is satisfied.

D. Simulation Studies

The proposed C-MBMD algorithm was first tested in a polyenergetic simulation study using a phantom containing water, bone and fat (Fig. 2A). The diameter of the outer water cylinder was 160 mm. Nine inserts with the diameter of 20 mm were placed inside, each filled with a unique mixture of water, bone and fat (Fig. 2A), where material volume fractions add to one.

The phantom was created using 0.125 mm voxels. The simulated system had a SAD of 400 mm, a SDD of 540 mm, and a flat-panel detector with 0.139 mm pixels. We implemented two kVp-switching schemes [7]: 60/120 and 60/140 (+0.225 mm Ag). Poisson noise was added onto the data (assuming a flux of 2×10^5 photons/pixel). Before the reconstruction, we binned the data by a factor of 4 (i.e. 2 × 2) to simulate partial volumes.

We compared decomposition performances of conventional IDD with a volume constraint [6] and the proposed C-MBMD. Prior to IDD, low energy (LE) and high energy (HE) volumes were individually reconstructed by PWLS algorithm with quadratic penalties. C-MBMD used water and bone as independent base materials, and each fraction was bounded between 0 and 1. C-MBMD was initialized with a scaled IDD result (scale = 0.7x) to test the accuracy when initialized away from the solution. We applied quadratic penalties with strengths of 5 and 10^3 for water and bone, respectively. Both for IDD and C-MBMD, the data was reconstructed onto 0.250 mm voxels.

We also validated the algorithm on a simulated knee phantom (Fig. 3). Three BME ROIs were added by increasing water fraction inside trabecular bone (mean water fraction ~0.881). We generated data using the 60/140 (+0.225 mm Ag) protocol. The same system geometry, projection post-processing and reconstructions were employed as for the previous study.

III. Results

Fig. 2A shows IDD and C-MBMD three-material decompositions of the cylinder phantom. IDD exhibited significant noise in the water and fat image, likely due to the matrix inversion in the IDD algorithm. Edge artifacts in water and fat fractions for C-MBMD can be resolved using a more sophisticated regularization scheme [11]. As anticipated, the 60/140 acquisition outperforms 60/120 because of the better spectral separation.

Mean estimated fractions and standard deviations for ROI-4, 5 and 6 are shown in Fig. 2B. Note that all three ROIs have the same fat fraction. Values estimated from C-MBMD are closer to the ground truths, and present reduced fluctuations than IDD.

Fig. 3 compares IDD and C-MBMD of the digital knee phantom. We include IDD results with both 180 views (kVp-switching) and 360 views. Like the previous study, C-MBMD exhibits improved noise in water and fat images compared to IDD. For C-MBMD, the estimated water fractions in ROI-1, 2 and 3 are 0.628 ± 0.029, 0.657 ± 0.064 and 0.581 ± 0.102, respectively. For IDD with 180 views, these values are 0.661 ± 0.216, 0.686 ± 0.213 and 0.690 ± 0.231. For IDD with 360 views, these values are 0.811 ± 0.107, 0.778 ± 0.145 and 0.677 ± 0.253. The relatively accurate estimation for IDD with 360 views is likely due to higher imaging dose and angular sampling. For comparison, Fig. 3B shows a decomposition using unconstrained two-material MBMD [1]. We note the difficulty of visually discriminating edema and fat in the water image, motivating the need for dedicated three-material decomposition algorithms.

IV. Discussion & Conclusion

The proposed bound-constrained MBMD enables simultaneous estimation of three-basis concentrations from DE data. The

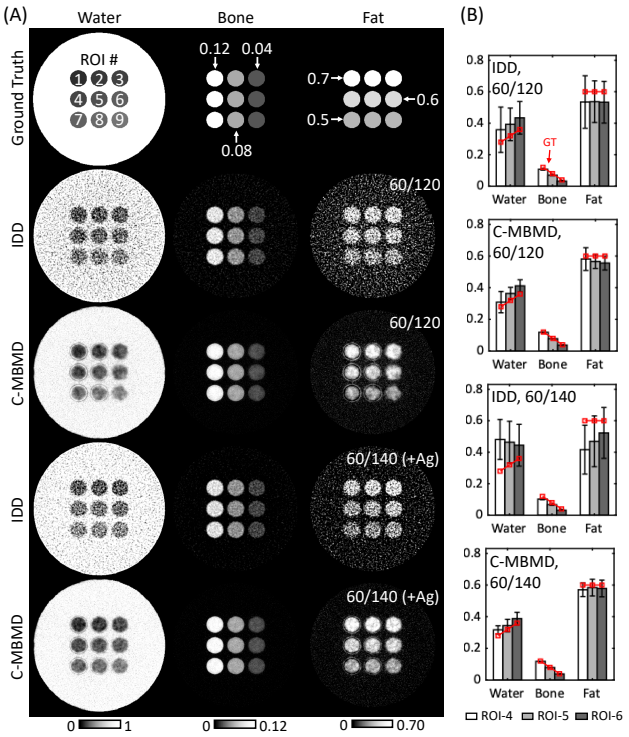


Figure 2. (A) IDD and C-MBMD estimates of water, bone and fat fractions for the cylinder phantom. Two kVp-switching protocols are compared: 60/120 and 60/140 (+Ag). (B) Mean estimated material fractions and their standard deviations in ROIs 4, 5 and 6. Red lines represent the ground truths.

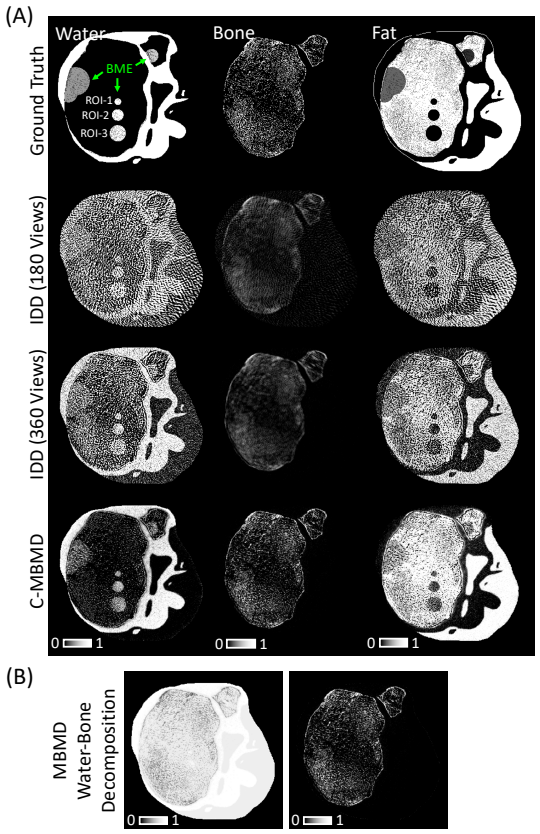


Figure 3. (A) IDD and C-MBMD estimates of water, bone and fat fractions for the digital knee phantom using the 60/140 (+Ag) protocol. (B) Unconstrained two-material decomposition of the knee phantom using MBMD.

algorithm does not require coinciding x-rays, thus can be applied to acquisition scenarios such as the presented kVp-switching. We validated C-MBMD in simulated DE imaging of BME, showing accurate estimation of base material fractions in mixtures of water, bone and fat. Ongoing work includes the experimental validation of C-MBMD on the weight-bearing extremity CBCT scanner [12,13]. When combined with KC-MBMD [7] and fracture biomechanical assessment algorithms [14,15], C-MBMD could enable simultaneous monitoring of tissue compositions and mechanical stability of injured extremities.

References

[1] S. Tilley, W. Zbijewski, and J. W. Stayman, "Model-based material decomposition with a penalized nonlinear least-squares CT reconstruction algorithm," *Phys. Med. Bio.*, vol. 64, no. 3, p. 035005, 2019.

[2] X. Liu, L. Yu, A. N. Primak, and C. H. Mccollough, "Quantitative imaging of element composition and mass fraction using dual-energy CT: Three-material decomposition," *Med. Phys.*, vol. 36, no. 5, pp. 1602–1609, 2009.

[3] Y. Long and J. A. Fessler, "Multi-material decomposition using statistical image reconstruction for spectral CT," *IEEE Trans. Med. Imaging*, vol. 33, no. 8, pp. 1614–1626, 2014.

[4] D. T. Felson, S. Mclaughlin, J. Goggins, M. P. Lavalley, M. E. Gale, S. Totterman, W. Li, C. Hill, and D. Gale, "Bone marrow edema and its relation to progression of knee osteoarthritis," *Ann. Intern. Med.*, vol. 139, no. 5 (Part 1), pp. 330-336, 2003.

[5] J. Garcia, S. Z. Liu, and A. Y. Louie, "Biological effects of MRI contrast agents: gadolinium retention, potential mechanisms and a role for phosphorus," *Phil. Trans. Royal Soc. A*, vol. 375, no. 2107, p. 20170180, 2017.

[6] W. Zbijewski, A. Sisniega, J. W. Stayman, G. Thawait, N. Packard, J. Yorkston, S. Demehri, J. Fritz, and J. H. Siewerdsen, "Dual-energy imaging of bone marrow edema on a dedicated multi-source cone-beam CT system for the extremities," *Medical Imaging 2015: Physics of Medical Imaging*, vol. 9412, p. 94120V, 2015.

[7] S. Z. Liu, S. Tilley, Q. Cao, J. H. Siewerdsen, J. W. Stayman, and W. Zbijewski, "Known-component model-based material decomposition for dual energy imaging of bone compositions in the presence of metal implant," *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 11072, p. 1107213, 2019.

[8] H. Erdogan and J. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imaging*, vol. 18, no. 9, pp. 801–814, 1999.

[9] S. Tilley, M. Jacobson, Q. Cao, M. Brehler, A. Sisniega, W. Zbijewski, and J. W. Stayman, "Penalized-likelihood reconstruction with high-fidelity measurement models for high-resolution cone-beam imaging," *IEEE Trans. Med. Imaging*, vol. 37, no. 4, pp. 988–999, 2018.

[10] J. Nocedal and S. J. Wright, *Numerical optimization*. New York: Springer, 2006.

[11] W. Wang, M. Tivnan, G. J. Gang, S. Tilley, and J. W. Stayman. "Generalized local impulse response prediction in model-based material decomposition of spectral CT," *Med. Phys.*, vol. 46, no. 6, pp. e257-e257, 2019.

[12] W. Zbijewski, P. D. Jean, P. Prakash, Y. Ding, J. W. Stayman, N. Packard, R. Senn, D. Yang, J. Yorkston, A. Machado, J. A. Carrino, and J. H. Siewerdsen, "A dedicated cone-beam CT system for musculoskeletal extremities imaging: Design, optimization, and initial performance characterization," *Med. Phys.*, vol. 38, no. 8, pp. 4700–4713, 2011.

[13] G. J. Gang, W. Zbijewski, M. Mahesh, G. Thawait, N. Packard, J. Yorkston, S. Demehri, and J. H. Siewerdsen, "Image quality and dose for a multisource cone-beam CT extremity scanner," *Med. Phys.*, vol. 45, no. 1, pp. 144–155, 2017.

[14] Q. Cao, S. Liu, G. Osgood, S. Demehri, J. Siewerdsen, J. Stayman, and W. Zbijewski, "Cone-beam CT of load-bearing surgical hardware using a mechanical model of implant deformation," *Med. Phys.*, vol. 46, no. 6, pp. e399-e399, 2019.

[15] S. Z. Liu, Q. Cao, G. M. Osgood, J. H. Siewerdsen, J. W. Stayman, and W. Zbijewski, "Quantitative assessment of weight-bearing fracture biomechanics using extremity cone-beam CT," *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 11317, p. 113170I, 2020.

Cone Beam CT Image Reconstruction by Origin Ensemble Algorithm

Andriy Andreyev, Arkadiusz Sitek and Christoph Graf vom Hagen

***Abstract*—Origin Ensemble algorithm is an iterative non-projecting image reconstruction algorithm, previously proposed for emission tomography. In this work we apply Origin Ensemble (OE) algorithm to reconstruction of Computed Tomography (CT) data. Absorption projection data is partitioned into sets of individual absorption events, after which OE algorithm is used similarly as in the reconstruction of emission tomography data. Using computer simulated and real data, we demonstrate basic feasibility of this approach applied to reconstruction of cone beam CT data.**

***Index Terms*— Computed Tomography, Image Reconstruction, Origin Ensemble, Markov chain, Metropolis Hastings, convergence.**

I. Introduction

THE Origin Ensemble (OE) tomographic image reconstruction algorithm has been successfully applied in Positron Emission Tomography [1][2], including Time-of-Flight [3], Single Photon Emission Tomography [4] and Compton Camera [5][6][7]. OE algorithm is an alternative non-projecting method for tomographic iterative reconstruction methods, meaning that it does not rely on the repeated forward- and backprojection to recover the tomographic image. Instead, it starts from initial randomized point cloud representation of the unknown distribution of locations of photon interactions, and progressively refines the individual coordinates of each point in using Markov Chain Monte Carlo (MCMC). In emission tomography the distribution represents locations of emissions of all detected counts. In transmission tomography, as used here, it represents locations of absorption amounts (AAs) within the volume.

OE algorithm is originally formulated for the emission tomography data, due to its direct representation of detected photon origins in SPECT or detected annihilation coincidences origins in PET. However, it similarly can be formulated to reconstruct CT projection data, in post-log domain as demonstrated here.

In this work we apply OE algorithm to simulated Cone-Beam CT data in post-log domain, describe the approach and review the preliminary results.

A. Andreyev and C. Graf vom Hagen are with Carl Zeiss X-Ray Microscopy, Pleasanton, CA 94588 USA (e-mail: andriy.andreyev@zeiss.com, christoph.graf-vom-hagen@zeiss.com).

A. Sitek is with IBM Watson Health, Cambridge, MA 02142 USA (e-mail: sarkadiu@gmail.com).

II. Methods

We define post-log transformed absorption data x along individual detector integrating element i as: $x_i = -ln\frac{N}{N_0}$, where N is measured photon flux across the studied object and N_0 is the unobstructed photon flux when no object is present in the FOV.

We treat the total value of absorption events detected in a certain projection bin x_i, as a sum of k individual absorption amounts $\sum_k x_{i,k}$ that are positioned at yet unknown locations along the line of response l_i that connects the X-ray source focal spot and the corresponding detector element i. OE algorithm execution starts from initialization state, where all absorption amounts (AAs) are spread randomly along respective LORs (AAs are not allowed to leave corresponding LORs in OE algorithm). At this stage, by summing the AAs in each image volume voxel an initial density matrix $D_{x_{t=0}}$ is created, representing the total AA in each image voxel j.

Similarly, as it was done in previous works, the OE algorithm for CBCT case executes the following loop which corresponds to Metropolis-Hastings iteration:

1. Randomly select AA $x_{i,k}$.
2. Randomly choose a new location along the LOR l_i.
3. The new location x_{t+1} is accepted with the probability of

$$p = \min\left(1, \frac{D_{x_{t+1}}+1}{D_{x_t}}\right),$$

where D_{x_t} is the absorption density in the current location x_{t+1} and $D_{x_{t+1}} + 1$ is the absorption density in the new location x_{t+1} including the event that would be transferred to this location, hence "+1". Per this probability definition the AAs will always change the location if higher density location is randomly selected in step 2. However, moves to lower density locations are still possible. Steps 1-3 are repeated across all absorption events until algorithm reaches equilibration state (or steady state) and the final updated density matrix D_{x_t} then becomes a final reconstructed image volume if the same voxel sampling is used. While in steady state, several samples of the density matrix can also be acquired and averaged providing another estimate of D_{x_t}. It can be shown that the loop of the algorithm defines reversible Markov chain which guarantees convergence

to equilibrium [11].

There can be multiple variations to the step 3 of the algorithm, such as it can be modified to include the effects of beam hardening or scatter, or for example an extreme case of edge enhancement can be realized with:

$$p = \begin{cases} 1, & D_{x_{t+1}} \geq D_{x_t} \\ 0, & D_{x_{t+1}} < D_{x_t} \end{cases}$$

During the iterations of OE algorithm, we stored current 3D location for each individual absorption amount. Certainly, as the number of detector elements and/or projection views increases this may increase the requirement on the computer memory. As an example, it may be challenging to use OE algorithm beyond 512x512 projection view sampling as in this case it can occupy the memory volume of about 512x512x800x3x4x10/1024/1024/1024 ~ 23GB assuming average 10 absorption events per detector bin and 800 CB projections.

III. Experiments

To provide some examples of the reconstruction, we first use computer generated data with single point object. In another experiment, we simulated the absorption data for simple water cylinder phantom with two steel rods inserted. The projections consisted of 180 projections of 256x256.

Finally, we took the real phantom data acquisition, courtesy Dr. Tom Case (Carl Zeiss Microscopy) acquired in Xradia Versa X-Ray Microscope (available from Carl Zeiss X-Ray Microscopy, Pleasanton CA). The data in cone beam geometry were downsampled and clipped to 262x110 projection pixels with 200 projection views as shown in Figure 4. Polynomial based beamhardening pre-correction was optionally applied to the data in order to reduce the cupping artifact.

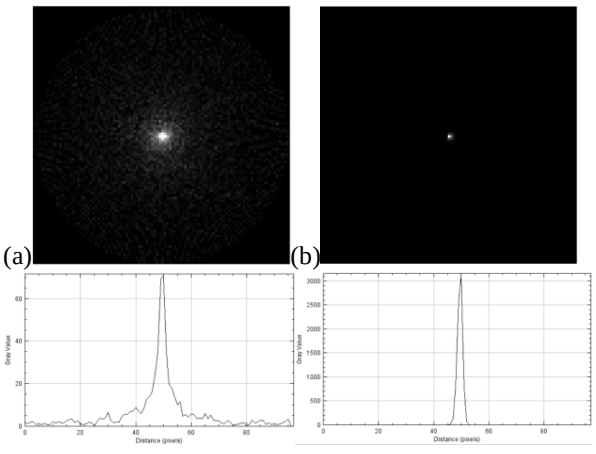


Figure 1. a) Initial density distribution. b) Image reconstructed after 100 iterations. Corresponding profiles are shown in the bottom.

IV. Results and Discussions

Figure 1a shows the result of an initial distribution where AAs are randomly distributed along corresponding LORs (roughly equivalent to simple unfiltered backprojection). Figure 1b is the result of 100 iterations of the OE algorithm where iteration is defined as the algorithm loop executed number of times equal to number of AAs.

The results with the simulated phantom are shown in Figure 2 with initial distribution, 10 iteration, and 100 iteration. The images were reconstructed from 256x256 pixel data with 180 projection views in cone beam CT geometry into 256x256x256 image volumes. The corresponding profiles over the two rods and background as well as convergence curves over the background and one of the rod ROIs are shown in Figure 3.

Reconstruction speed depends on the k number of partitioned absorption values. The larger number of k, the better accuracy can OE algorithm achieve, however, at the increased computational expense and higher memory requirement. The measured absorption values in the cylinder data in Figure 4 ranged from 0 to about 3. To achieve sufficient AA sampling density, they were sub-partitioned into 10 AAs per unit absorption value. For this data the reconstruction took about 20 mins with unoptimized C++ code, single tread CPU and 100 iterations. The reconstructed real data phantom is shown in Figure 5 a) without and b) with beamhardening correction with latter significantly reducing the cupping artifact.

Due to simplification of geometry, significant downsampling of acquired projections, the obtained images did not achieve the image quality that can be obtained with proper implementation of OE algorithm, e.g. [3]. In the future we consider using OE algorithm using parallel implementation on modern CPUs or GPUs, which can also be done analogous to emission tomography implementations.

The advantage of using OE algorithm is insensitivity to the reconstructed volume dimension since the reconstructed absorption coefficients are represented as point cloud. Similarly, as in implementations for PET and Compton Camera, models of complex processes of the data acquisition physics are much simpler than that of conventional iterative methods. For example, the finite focal spot size and/or detector PSF can be modeled by introducing an additional degree of freedom along each line of response as in [8]. Other vital corrections such as beamhardening correction can be applied similarly as in conventional CT reconstruction by pre-scaling the projection data by a certain polynomial. This ability of OE algorithm to model complex physics of the CBCT acquisition is the main motivation of using OE for CBCT reconstruction. We also plan to test OE approach to investigate how OE algorithm would perform in photon counting CT or dark field tomography applications.

V. Conclusion

We demonstrated the first application of OE image reconstruction algorithm to CT data reconstruction in cone-beam geometry. By treating the absorption data in the similar way to emission data, it is possible to use the OE image reconstruction algorithm developed for emission data in

reconstruction of transmission modality. Certainly, a further studies are needed find out whether this method is only an algorithmic novelty without real application benefits, or if it may have some real advantages over the conventional image reconstruction methods especially in cases with complex modeling of physics.

REFERENCES

[1] A. Sitek, "Representation of photon limited data in emission tomography using origin ensembles," Phys. Med. Biol. 53:3201, 2008.

[2] A. Sitek and D.J. Kadrmas, "Compton scatter and randoms correction for origin ensemble 3D PET reconstructions," Proc. Fully 3D, 2011.

[3] C. Wülker, A. Sitek, S. Prevrhal, "Time-of-flight PET image reconstruction using origin ensembles," Phys. Med. Biol., 60(5), 1919-44, 2015.

[4] M.C. Lyon, A. Sitek, S.D. Metzler, S.C. Moore, "Reconstruction of multiple-pinhole micro-SPECT data using origin ensembles," Med. Phys., 43(10), 2016.

[5] A. Andreyev, A. Sitek, and A. Celler, "Fast image reconstruction of Compton camera using stochastic origin ensemble approach," Med. Phys., 429–438, 2011.

[6] D. Mackin, S. Peterson, S. Beddar, and J. Polf, "Evaluation of a stochastic reconstruction algorithm for use in Compton camera imaging and beam range verification from secondary gamma emission during proton therapy," J. Physics in Medicine & Biology 57(11), 3537, 2012.

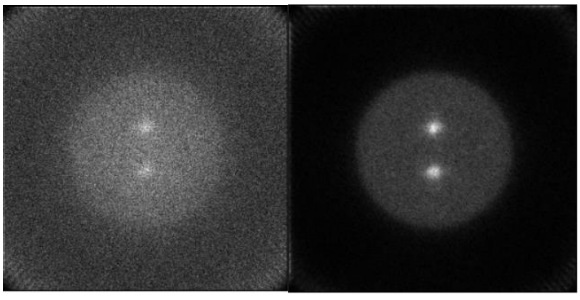
[7] Z. Yao, Y. Xiao, Z. Chen, B. Wang, Q. Hou, "Compton-based prompt gamma imaging using ordered origin ensemble algorithm with resolution recovery in proton therapy," Scientific Reports 9, Article number: 1133, 2019.

[8] A. Sitek, A. Andreyev, "Detector Response Correction for 3D PET Using Bayesian Modeling of the Location of Interaction," IEEE Nucl. Sci. Symp. MIC Conf. Rec. 2012.

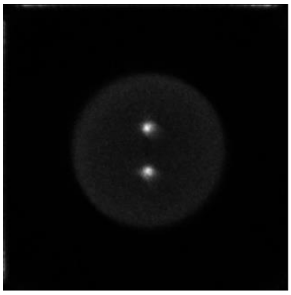
[9] A. Andreyev, A. Celler, I. Ozsahin, A. Sitek, "Resolution recovery for Compton camera using origin ensemble algorithm," J. Medical physics 43(8 Part 1), 4866–4876, 2016.

[10] A. Andreyev, A. Sitek, A. Celler, "Resolution recovery for Compton camera using origin ensemble algorithm," IEEE Nucl. Sci. Symp. MIC Conf. Rec., 2011.

[11] A. Sitek, "Statistical Computing in Nuclear Imaging," CRC Press 2014.

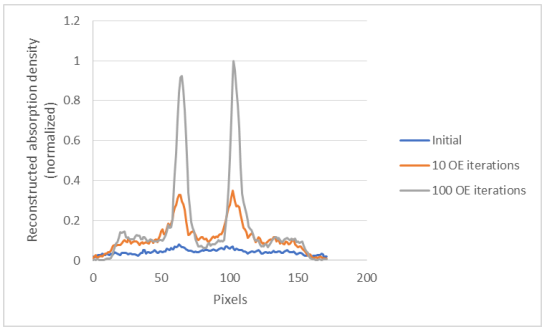


a) b)

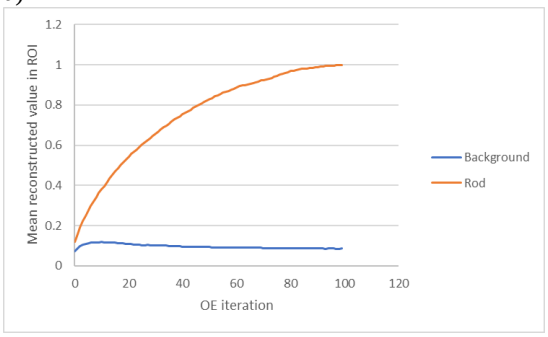


c)

Figure 2. Simulated water phantom with two metal inserts: a) initial random distribution; b) 10 OE iterations; c) 100 OE iterations.



a)



b)

Figure 3. (a) Intensity profiles and (b) convergence curves for the background and one of the rods (mean reconstructed value scaled to the absolute maximum).

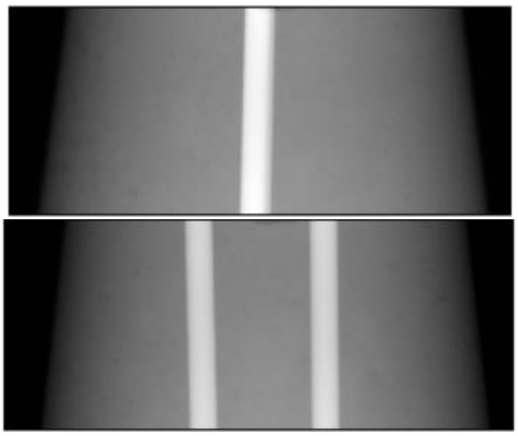
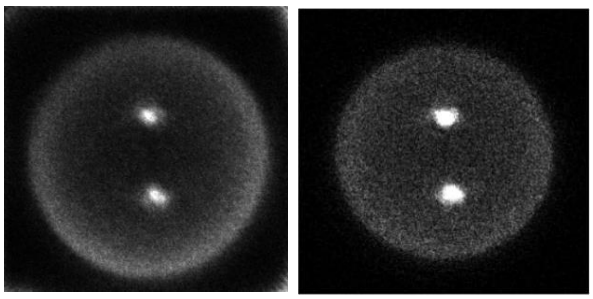


Figure 4. Two brass rods inserted in plaster phantom: original downsampled projection data. 90 degree projection view difference.



a) b)

Figure 5. Real data reconstruction after 100 iterations: a) no beamhardening correction, b) with polynomial beam hardening correction.

A Constrained Curve-Fitting Method for Scatter Estimation for Helical Cone-Beam CT

Zhicong Yu, Amit Jain, Chuanyong Bai and Daniel Gagnon
Imaging Technologies, Accuray Incorporated

***Abstract*—X-ray scatter estimation is critical for cone-beam CT imaging. The state-of-the-art methods either require intensive computation or extensive data training. The curve-fitting method using scatter measured in the shadow region (Siewerdsen J. et al. 2006) is simple and efficient. However, its performance degrades when the beam width becomes larger and the object becomes complex, due to lack of scatter information within the illuminated beam region. This work extends this direct curve-fitting method to an constrained version for helical scans of pitch less than 1. The idea is to use shadow scatter measured from neighbor rotations to estimate scatter within the illuminated beam region of the current view. The additional information enhances the quality of curve fitting and results improved scatter estimation. We refer to this method as constrained curve-fitting method and we demonstrate it using phantom data acquired from a benchtop system.**

I. Introduction

Cone-beam CT (CBCT) is an important imaging tool for image-guided radiation therapy and interventional radiology. However, the CT number accuracy and uniformity of CBCT are suboptimal, mainly because of x-ray scatter contamination. Significant progress has been made for x-ray scatter estimation and correction. The state-of-the-art methods either require intensive on-the-fly computation (e.g.,[1], [2]) or substantial prior information and extensive training (e.g.,[3]). In 2006, Siewerdsen J et al. proposed a method [4] that does not need either of the two requirements. It estimates scatter in the illuminated beam by curve fitting using scatter information located in the shadow region. For convenience, we will refer this method as the direct curve fitting (DCF) method. In general this method is less accurate compared to the kernel or deep-learning based methods, but it is simple, efficient and direct, and thus practical and attractive. The performance of the method is dependent on the width of the illuminated beam (size of aperture opening): The narrower the beam is, the better scatter estimation is. When the illuminated beam is large and the object is complex, its performance degrades due to lack of any scatter information within the illuminated beam region.

This work extends the DCF method to a constrained curve-fitting method (CCF) for helical CBCT. It first estimates scatter at a couple locations within the illuminated beam region of the current view by using shadow information from neighbor rotations. This estimated sparse scatter information is then combined with the shadow information of the current view for the final scatter estimation via curve fitting. Because of the extra scatter information approximated within the illuminated beam region, the curve fitting is expected to be more constrained and accurate, resulting in improved scatter estimation.

We demonstrate the CCF method using phantom data acquired from a benchtop system. We assess the quality of scatter estimation via comparison against scatter estimated using the DCF method and the beam-stopper-array (BSA) method [5]. Our results showed improved accuracy of scatter estimation of the proposed CCF method when compared to the DCF method. The rest of the paper is dedicated to the description of the CCF method as well as the detailed evaluation results.

II. Method and Materials

A. Assumptions and Convention

We consider a helical scan with a large flat-panel detector; see Figure 1. We make two assumptions: a) The helical scans are of pitch less than 1.0; b) The active detector area is noticeably larger than the illuminated beam along the axial direction. Both assumptions are necessary to ensure that the proposed method has sufficient amount of data to work with.

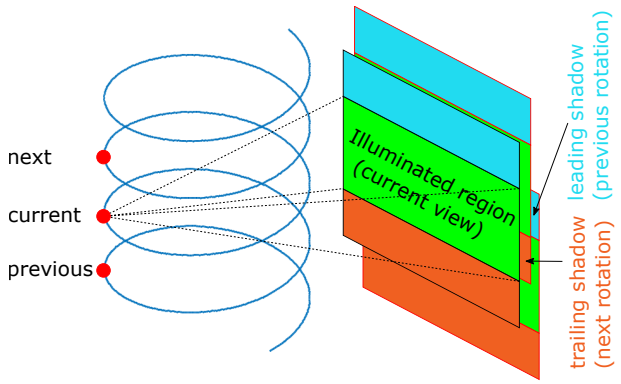


Fig. 1. Illustration of three projections of the same azimuth view angle from the previous, current and next rotations of a helical scan with pitch 0.5. We omitted penumbra for simplification purpose.

We introduce some convention and notation. Let λ be the angle of the current view. We refer to the views located at $\lambda - 2\pi$ and $\lambda + 2\pi$ as the previous and next rotation views, respectively; see Figure 1. The detector is divided into five regions axially. Following source moving direction, the five regions are named the trailing shadow region, the trailing penumbra, the illuminated beam region, the leading penumbra and the leading shadow region; see Figure 2 for illustration. The proposed method requires that at least one shadow region is available. Without loss of generality, we

will take the case with both shadow regions available as an example. Also, the method is in dependent on channel position, and we will only focus on the axial direction.

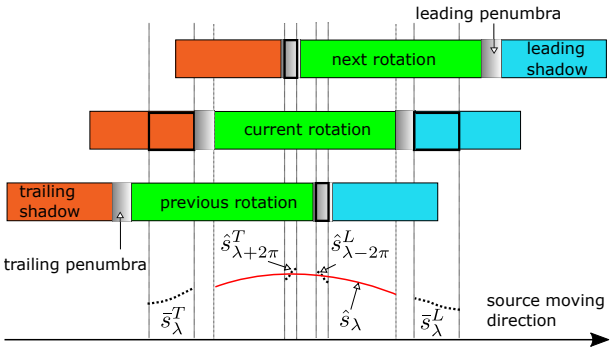


Fig. 2. Illustration of the constrained-curve-fitting (CCF) method using scatter information approximated from previous and next rotation views.

We denote the signals measured in the trailing and leading shadow regions of the current view as $\bar{s}^T_\lambda$ and $\bar{s}^L_\lambda$, respectively. We also denote $\hat{s}_\lambda$ as the final scatter estimation for the current view. Note that the signals here are gain and offset corrected, and their values are in $(0, 1]$.

B. Method

The CCF method can be achieved using the following steps; see Figure 2 for illustration.

- Step 1: For the current view, extract scatter measurements $\bar{s}^T_\lambda$ and $\bar{s}^L_\lambda$ (assume no off-focal radiation) from both shadow regions.
- Step 2: For the previous rotation view, estimate scatter located around the leading penumbra using the leading shadow by extrapolation. We denote this estimated scatter as $\hat{s}^L_{\lambda-2\pi}$.
- Step 3: For the next rotation view, estimate scatter located around the trailing penumbra using the trailing shadow by extrapolation. We denote this estimated scatter as $\hat{s}^T_{\lambda+2\pi}$.
- Step 4: Perform curve fitting using $\{\bar{s}^T_\lambda, \hat{s}^T_{\lambda+2\pi}, \hat{s}^L_{\lambda-2\pi}, \bar{s}^L_\lambda\}$.
- Step 5: Perform scatter estimation of the current view, $\hat{s}_\lambda$, using result of Step 4.

Some comments on the proposed method. First, regarding signal in the shadow regions, one should avoid contamination from the penumbra regions. Meanwhile, we should choose a limited number of shadow rows for scatter estimation, because the closer the location of measured scatter to the to-be-estimated scatter is, the more accurate scatter estimation it would result. In this work, we chose to use 50 shadow pixels (equivalent to 15 mm at isocenter) that are right next to the penumbra region.

Second, extrapolation is generally not preferred. However, in this case, because the scatter profile drops dramatically when transitioning from the illuminated region to the shadow region, directly using the shadow as scatter information will cause substantial scatter underestimation. On the other side, we also would like to keep the extrapolation as limited as possible. In this work, we used linear extrapolation and chose 20 pixels (equivalent to 6 mm at isocenter) about the penumbra region that was directly adjacent to the shadow pixels used for the extrapolation.

Third, the assumption of pitch smaller than 1 is to make sure that the extrapolated scatter information from neighbor rotations falls into the illuminated region of the current view.

Fourth, it can also be noted that the proposed method requires at least one shadow region to approximate scatter information in the illuminated region before curve fitting.

Finally, we used polynomial fitting of order 2 in Steps 4.

C. Experiments

To demonstrate the proposed method, we performed scatter estimation on projection data of an anthropomorphic skull phantom acquired on a benchtop system; see Figure 3. Two data sets were acquired: One with the skull phantom only, and the other with lead bead placed between the phantom and the detector. The lead beads were 2.5 mm in diameter and were placed on 2D grids of 25 mm pitch. The tube (G1092, Varex Imaging Corporation) used 1.2mm focal spot and 12° degree anode. The detector (4343RF, Varex Imaging Corporation) used gain mode 5, half panel readout and 3x3 binning (corresponding to 0.45 mm detector pixel size).

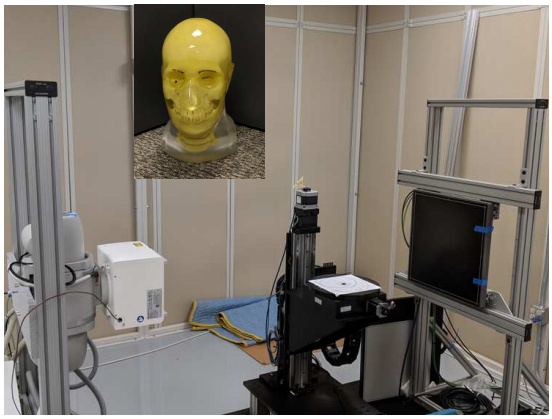


Fig. 3. Illustration of the skull phantom and the benchtop CBCT system.

Projection data were acquired using a 3-turn helical scan with 360 views per rotation. The helical pitch was 0.5 and the illuminated beam width at isocenter was 48 mm. The the source-to-iso and source-to-detector distances were 1104 mm and 1676 mm, respectively. Data were acquired using 125 kV tube potential, 125 mA tube current and 5 ms tube pulse.

D. Data Correction and Image Reconstruction

Gain, offset, bad-pixel and scatter corrections were applied before reconstruction. Images were reconstructed by a method similar to [6], which involves view-differentiation and Hilbert transform. The view-differentiation was implemented using the method presented in [7]. The resolution control parameter (ε) was set to 0.03. A Hanning window was used for the Hilbert transform. Image pixel size was $[0.4, 0.4]$ mm, and the slice thickness was about 1 mm.

III. Results

Estimated scatter profiles from the DCF method [4] and the proposed CCF were both compared to the projection profile obtained via BSA method [5]; see Figure 4 for illustration. Both the DCF and CCF scatter profiles were underestimated. The CCF results showed less underestimation than the DCF results.

We performed visual inspection of images that are reconstructed from projection data without scatter correction, with the DCF scatter correction and with the proposed CCF scatter correction. The reconstruction results are shown in Figure 5. Both DCF and CCF scatter corrections lifted the image profile noticeably. In most locations, the CCF correction resulted more CT number increase than the DCF correction. However, this observation is not conclusive, because of the complex nature of the phantom, the beam hardening artifacts and the streaking artifacts (caused by insufficient sampling of 360 views per rotation). More experiments is needed for further investigation.

IV. Discussion

In the current work, we have evaluated the proposed method using helical scans of a skull phantom with a small pitch of 0.5. The phantom is relatively small and the involved scatter contamination was also small. To thoroughly assess the performance of the CCF method, experiments with large phantoms are necessary and will be investigated in the future. In addition, the skull phantom is too complex for qualitative assessment, a quality phantom such as a CATPHAN phantom is planned as the next step.

Pitch plays a critical role for the performance of the CCF method. Note that, the scatter generated in the neighbor rotations are different from the scatter generated at the current rotation, even though the physical location for scatter estimation is the same. The smaller the pitch is, the closer the two scatter deposition would be. On the other hand, a larger pitch would pose more challenge for the DCF method. The larger the pitch is, the more the CCF would outperform the DCF method. Therefore, there is a balanced helical pitch (between 0 and 1) that would result in the most improvement of scatter estimation by the CCF method relative to the DCF method. This is a topic for future investigation.

Another important point is extrapolation. As it was mentioned earlier, extrapolation is not preferred in general, but useful in this case. However, how to optimize the extrapolation method and the amount of data to use are still topic deserves particular attention. In the current work, we used a linear extrapolation just to avoid any wild results. However, other extrapolation methods should be assessed. Although we used shadow signal of 50 pixels to extrapolate scatter for the adjacent 20 pixels, it is empirical and should be extensively investigated for different phantoms and scan configurations.

In addition, it was reported in [4] that the DCF method is not stable when some discontinuity occurs in the shadow region. The authors suggested to reduce the order of polynomial fitting to solve the problem. The same problem exists in the CCF method. The difference is that CCF may be more stable when the shadow of the current view presents discontinuity and the shadow of the neighbor views do not. However, it is a double-edge sword problem, because the the opposite could happen where the shadows of the neighbor rotation views present discontinuity, which degrades the scatter estimation of the current view. Addressing this discontinuity issue is another topic deserves future investigation.

We would like to point out that although we have restricted ourselves to helical scans of CBCT, the proposed method is also applicable to multi-circular scans, including CT. It is also worth pointing out that although we have limited ourselves to pitch smaller than 1, the CCF can be modified such that it estimates scatter information in the illuminated region of the current view directly using the shadow signal from neighbor rotations without extrapolation. Although this may be less effective as it will result in more underestimation, it would be more robust and conservative.

Finally, in image-guided radiation therapy, an off-centered flat panel detector is more of the case for imaging. We will investigate the CCF method in this scenario. Earlier we have reported that the off-centered detector will only allow pitch smaller than 1 [6], which perfectly falls in the pitch scope required by the proposed method.

V. Summary

We have proposed a constrained curve-fitting (CCF) method for scatter estimation of helical CBCT scans. This method uses scatter information approximated from neighbor rotations to improve the scatter estimation of the current view. We have conducted experiments using small pitch (0.5) helical scans of a skull phantom with both trailing and leading shadows available. The method was evaluated in both projection and image domains. The results showed that the CCF method is effective in scatter reduction and resulted in less scatter underestimation than the DCF method.

References

[1] M. Sun and J. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition," *Phys.Med.Biol.*, vol. 55, no. 22, p. 6695, 2010.

[2] A. Wang, A. Maslowski, P. Messmer, M. Lehmann, A. Strzelecki, E. Yu, P. Paysan, M. Brehm, P. Munro, J. Star-Lack *et al.*, "Acuros CTS: A fast, linear boltzmann transport equation solver for computed tomography scatter–part ii: System modeling, scatter correction, and optimization," *Med.Phys.*, vol. 45, no. 5, p. 1914, 2018.

[3] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep scatter estimation (DSE): Accurate real-time scatter estimation for x-ray ct using a deep convolutional neural network," *J. Nondestruct. Eval.*, vol. 37, no. 3, p. 57, 2018.

[4] J. Siewerdsen, M. Daly, B. Bakhtiar, D. Moseley, S. Richard, H. Keller, and D. Jaffray, "A simple, direct method for x-ray scatter estimation and correction in digital radiography and cone-beam CT," *Med.Phys.*, vol. 33, no. 1, p. 187, 2006.

[5] R. Ning, X. Tang, and D. Conover, "X-ray scatter correction algorithm for cone beam CT imaging," *Med.Phys.*, vol. 31, no. 5, p. 1195, 2004.

[6] Z. Yu, C. Bai, and D. Gagnon, "Double-helix trajectory for image guided radiation therapy: geometry and image reconstruction," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 11072, 2019, p. 110721Y.

[7] F. Noo, S. Hoppe, F. Dennerlein, G. Lauritsch, and J. Hornegger, "A new scheme for view-dependent data differentiation in fan-beam and cone-beam computed tomography," *Phys. Med. Biol.*, vol. 52, no. 17, p. 5393, 2007.

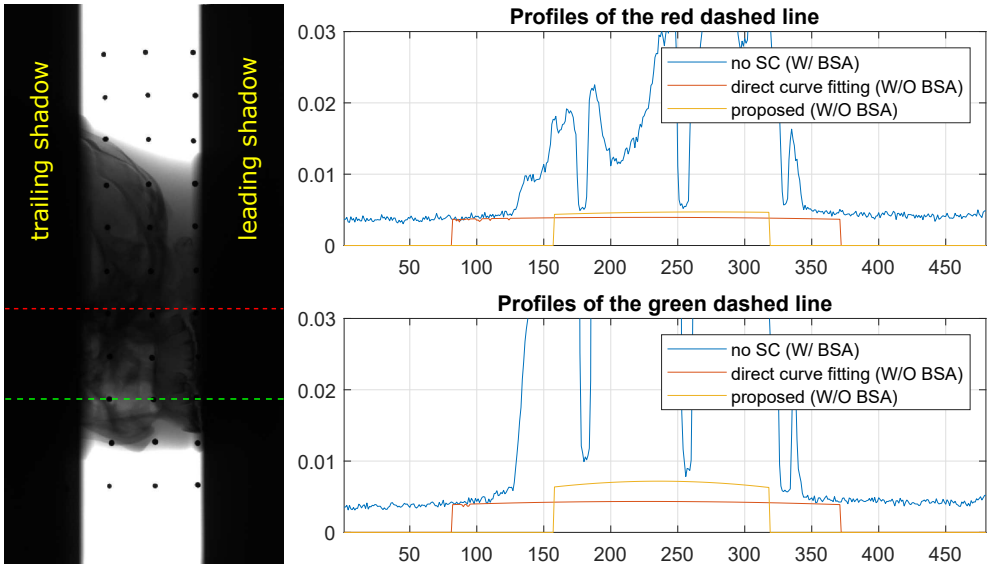


Fig. 4. Left: Projection of the skull phantom with beam stopping array (BSA). Right: Profiles of the red and green dashed lines in the left. The proposed method shows similar and better scatter estimation for the green and red lines, respectively.

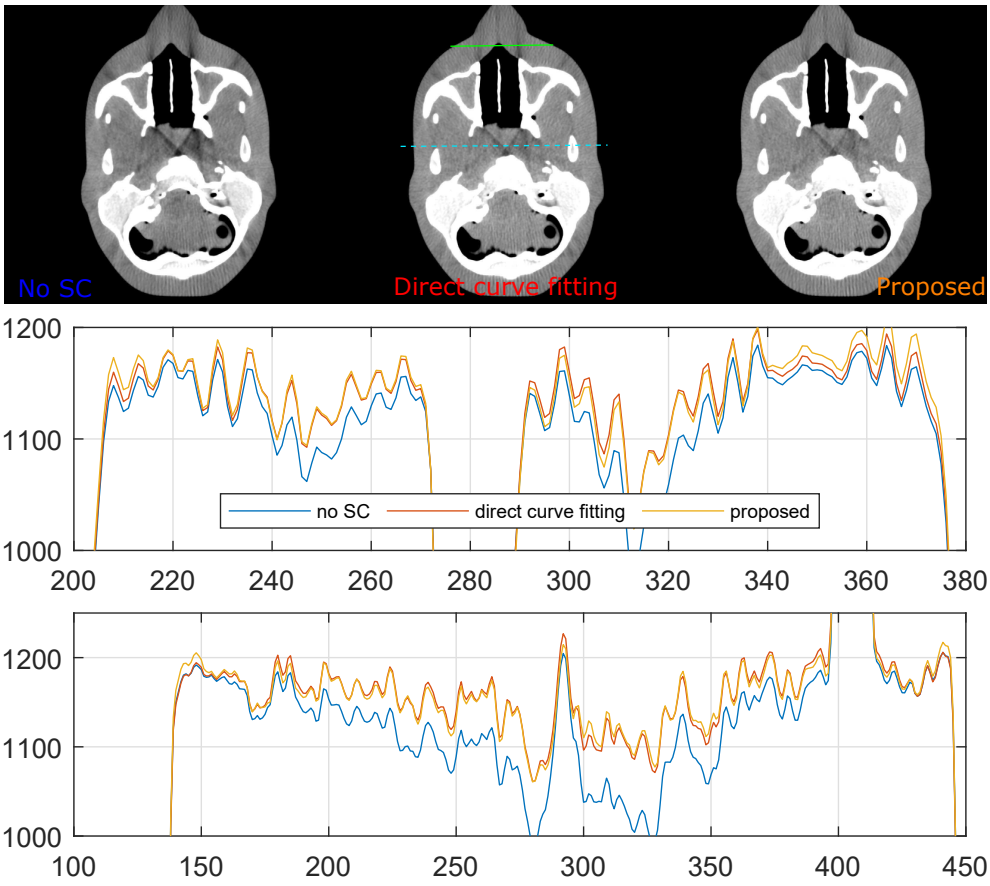


Fig. 5. Top: Reconstructed images. Middle: Profiles of the solid green line in the top images. Bottom: Profiles of the dashed green line in the top images. Display window: 600 HU.

Joint Reconstruction and Image Space Material Decomposition in Sparse View Spectral CT

Alexander Meaney, Jussi Toivanen, Samuli Siltanen, and Ville Kolehmainen

***Abstract*—The diagnostic capabilities of conventional computed tomography are limited by the fact that no definitive information on material composition is available: two separate tissues with different elemental compositions and densities may have the same HU value in the reconstructed image. More information on the material composition can be obtained with multi-energy imaging, which takes advantage of the nonlinear energy dependency of the attenuation coefficients. We propose a technique for joint reconstruction in multi-energy CT based on rapid kV switching and utilizing structural similarity in the reconstruction process. This facilitates low-dose multi-energy imaging using sparse views, as each reconstruction is able to utilize additional *a priori* structural information from the measurements at other energies. The CT slices at each energy are then used to compute an image space material decomposition into basis materials. Structural priors can be used to produce acceptable quality material decompositions even when the number of projections is reduced by over 80%. We demonstrate that our method produces reconstructions superior to computing sparse angle reconstructions independently for each energy.**

***Index Terms*—Multi-energy CT, material decomposition, iterative reconstruction, low-dose imaging, sparse view tomography.**

I. Introduction

Identification and quantification of tissues using computed tomography can be a highly ambiguous process: when examining a single CT image, two materials having different atomic numbers and mass densities can have the exact same grayscale value. More information on the material composition can be obtained by multi-energy computed tomography (MECT), which takes advantage of the distinct energy-dependencies of the attenuation coefficients of different elements [1]. It is then possible to differentiate the materials present in the sample into a set of basis materials, a process known as material decomposition (MD).

Different approaches to MECT exist. The multi-energy X-ray data acquisition can be performed by using energy sensitive detectors, or by varying the energy of the X-ray beam itself. The material decomposition step can be done pre-reconstruction (projection-based MD) or post-reconstruction (image-based MD). Projection-based MD requires that the projection directions for each energy overlap with each other.

In this work, we present an approach to low-dose MECT imaging that does not require overlapping projection directions for different energies, nor any type of energy-sensitive X-ray detectors. The multi-energy data is acquired using rapid switching of the X-ray tube potential (kV switching), and dose reduction is achieved by using sparse projection angles. The problems typical to sparse angle CT reconstruction, such as streak artefacts, are overcome by using structural priors, which enforce the reconstructions to be structurally similar, while allowing for different contrast between the energies. Our goal was to develop a joint reconstruction technique for low-dose multi-energy CT that is robust enough for quantitative analysis, *i.e.*, material decomposition into basis materials.

This work is an expansion of our earlier, more extensive work on joint reconstruction in multi-energy CT [2]. Here, we have combined a number of select techniques with material decomposition and a new image quality metric.

A. Meaney and S. Siltanen are with the Department of Mathematics and Statistics, University of Helsinki. E-mail: alexander.meaney@helsinki.fi.

J. Toivanen and V. Kolehmainen are with the Department of Applied Physics, University of Eastern Finland.

II. Methods and Materials

The workflow we used for MECT with image-based MD consists of three stages. Initially, X-ray projections are collected using two or more X-ray energies. This is followed by the reconstruction step, where a number of reconstructions equal to the number of different energies are computed. Finally, a material decomposition into basis materials is computed from the MECT reconstructions.

A. Data acquisition

The data acquisition process is based on a kV switching scheme, where the tube acceleration voltage cycles between *three* different acceleration voltages, from low to high, and then starting the cycle again. The imaging is performed during continuous rotation of the gantry, resulting in non-overlapping projection angles, *i.e.*, no two X-ray projections are acquired from the same position for any of the energies.

B. Multi-energy reconstruction

In MECT, the X-ray projections are collected using n different energies E_i. The forward model for the measurements is given by

$$m_i = A_i x_i + \epsilon_i, \tag{1}$$

where, for each E_i, $m_i \in \mathbb{R}^k$ is the projection data, $A_i \in \mathbb{R}^{k\times l}$ is the projection operator, $x_i \in \mathbb{R}^l$ is the true image we wish to recover, and $\epsilon_i \in \mathbb{R}^k$ models the measurement noise. It should be emphasized that the operators A_i for each energy can be different from each other.

In the conventional approach to optimization-based iterative reconstruction we form the objective function

$$F(x_i) = ||A_i x_i - m_i||_2^2 + \alpha_i R(x_i), \tag{2}$$

where $||A_i x_i - m_i||_2^2$ is the data fidelity term, $R(x_i)$ is the regularization term which incorporates *a priori* assumptions about the solutions, such as sparsity in a suitable basis, and α_i is the regularization parameter or weight. The solutions are found by minimizing the objective function for each E_i.

To provide a comparison to the reconstructions using structural priors, we computed sparse angle reconstructions independently for each energy using two methods. First, a simple least squares reconstruction where $R(x_i) = 0$ was used to compute reconstructions without adding any prior information to the inverse problem (**No prior**). Second, total variation regularization (**TV**), which promotes piecewise constant images via sparsity in the gradient of the reconstruction, was used [3].

In order to use structural priors we first form a joint reconstruction problem, where we combine all the datasets into a single problem which is then solved simultaneously for all energies. The objective function to be minimized now becomes

$$F(x_1, \ldots, x_n) = \sum_{i=1}^{n} ||m_i - A_i x_i||^2 + \alpha R(x_1, \ldots, x_n). \quad (3)$$

This work is based on the assumption that although the images will have different contrasts at different energies due to the energy-dependency of photon-matter interactions, the underlying structure remains the same. We have experimented with a number of regularization schemes which promote structurally similar reconstructions. Joint total variation (**JTV**), adapted from denoising of RGB images, favors sparsity of joint gradient, *i.e.*, nonzero gradients such as edges at the same locations [4]. We also developed a prior based on the structural similarity index (SSIM), which is used to quantify the similarity between two images using three components: luminance, contrast and structure [5]. This prior (**S**) is:

$$R(x_1, \ldots, x_n) = \big(\bar{S}(x_1, x_2) + \bar{S}(x_2, x_3) + \ldots + \bar{S}(x_n, x_1)\big)^{-1}, \quad (4)$$

where $\bar{S}(x_a, x_b)$ is a pairwise computed similarity measure based only on the SSIM structural component

$$\hat{S}(x_a, x_b) = \frac{\sigma_{x_a x_b} + C}{\sigma_{x_a}\sigma_{x_b} + C}, \quad (5)$$

where x_a and x_b are two images, $\sigma_{x_a x_b}$ is their cross-correlation, σ_{x_a} and σ_{x_b} are their standard deviations, and C is a small constant to avoid division by zero. $\bar{S}(x_a, x_b)$ grows when the two images become more structurally similar, corresponding to a decrease in (4). A full description of the prior is given in [2]. Finally, to cope with noise in the reconstructed images, we used a prior which combines the SSIM-based prior with total variation applied independently to each image (**S+TV**).

As reference images, we computed reconstructions independently for each energy using filtered backprojection (FBP) and densely acquired projections (**Ref.**) [6]. The FBP reconstructions and the forward models A_x for the iterative reconstruction techniques were created using the ASTRA Tomography Toolbox [7], [8], and the optimization problems in Eqs. (2) and (3) were minimized using the Polak-Ribière nonlinear conjugate gradient method [9].

C. Material decomposition

The linear attenuation coefficient $\mu(E)$ of a given material depends on the mass attenuation coefficient $u(E)$ and the mass density ρ:

$$\mu(E) = \frac{\mu(E)}{\rho}\rho = u(E)\rho. \quad (6)$$

The grayscale value of a given pixel in the CT image depends on the linear attenuation coefficient in the corresponding location of the object. $\mu(E)$ can be modeled as a linear combination of a set of basis materials and their densities:

$$\mu(E) = u_1(E)\rho_1 + u_2(E)\rho_2 + \ldots + u_N(E)\rho_N, \quad (7)$$

where N is the number of basis materials [10].

If the CT acquisition is performed using n X-ray energies, this leads to a system of linear equations:

$$\begin{aligned} \mu_1 &= u_{11}\rho_1 + u_{12}\rho_2 + \ldots + u_{1N}\rho_N \\ \mu_2 &= u_{21}\rho_1 + u_{22}\rho_2 + \ldots + u_{2N}\rho_N \\ &\vdots \\ \mu_n &= u_{n1}\rho_1 + u_{n2}\rho_2 + \ldots + u_{nN}\rho_N, \end{aligned} \quad (8)$$

or in, expressed in matrix form:

$$X = Up, \quad (9)$$

where $X \in \mathbb{R}^n$ contains the reconstructed linear attenuation coefficients, $U \in \mathbb{R}^{n \times N}$ contains the mass attenuation coefficients of the basis materials, and $p \in \mathbb{R}^N$ contains the densities of the basis materials. Eq. (9) was solved for p pixelwise by finding the least-squares solution of the system.

III. Results

The different reconstruction techniques were tested with X-ray data collected using a microtomography scanner (GE Phoenix nanotom 180 NF) at the Department of Physics at the University of Helsinki (Fig. 1). The scanner has an end-window X-ray tube with a tungsten target and a beryllium window. Three energies were used: 50 kV without additional filtration (E_1), 80 kV with a 1 mm Al filter (E_2), and 120 kV with a 0.5 mm Cu filter (E_3).

As a phantom, we used the chest of a small bird (a common quail, *Coturnix coturnix*) acquired from a local supermarket. The phantom was scanned three times consecutively, and 720 projections across 360° were collected at each energy and subsequently subsampled to provide a sparse geometry of non-overlapping angles. 30 projections were selected from each scan such that the angles were $0°, 12°, \ldots, 348°$ for E_1, $4°, 16°, \ldots, 352°$ for E_2, and $8°, 20°, \ldots, 356°$ for E_3.

A. Multi-Energy CT Reconstructions

Multi-energy reconstructions for E_1, E_2, and E_3 using the different techniques are shown in Fig. 2. The top 5 rows show the results obtained using the sparse angle geometry with 30 projections per energy, and the bottom row shows the reference reconstructions using FBP and 720 projections per energy.

It is visually obvious that the reconstructions computed with no prior contain severe streaking artefacts, and many of the

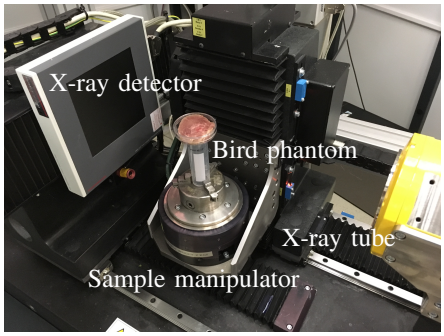


Fig. 1. Measurement setup for the MECT data.

TABLE I
HAAR WAVELET-BASED PERCEPTUAL SIMILARITY INDEX (HAARPSI) VALUES FOR THE MECT RECONSTRUCTIONS.

Reconstruction	E_1	E_2	E_3	*Mean*
No prior	0.376	0.386	0.387	0.383
TV	0.432	0.453	0.395	0.427
JTV	0.465	0.498	0.439	0.467
S	0.461	0.462	0.413	0.446
S + TV	0.509	0.536	0.439	0.494
Ref.	1.000	1.000	1.000	1.000

bone details are lost or smeared. The TV reconstructions do not contain the streaking artefacts, but the details in the images are still blurred in places. The JTV is clearly an improved on TV, while maintaining a similar type of noise structure and smoothing. The SSIM-based prior S gets rids of the streaking artefacts and maintains the structure of the smaller bones remarkably well, but the reconstructions are also very noisy. The combination of the SSIM-based prior with total variation, S + TV, maintains the bone structure detail seen using the S prior, while significantly reducing the noise.

When computing the reconstructions, we modelled the forward projections using the line integral model given in equation (1), although the attenuation process with polychromatic sources is actually nonlinear. Possibly due to the small size of the target, we did not encounter significant problems with beam hardening, although some cupping is detectable in the E_1 and - to a lesser extent - E_2 images. With larger targets some form of beam hardening correction may be required in combination with these reconstruction techniques, either using data preprocessing or an iterative correction method [11].

As an image metric, we computed the Haar wavelet -based perceptual similarity index (HaarPSI) for each reconstruction compared to the reference reconstruction of the same energy [12]. The values are shown in Table I. In all cases, the reconstructions using a structural prior (JTV, S, S + TV) produce higher values than the reconstructions computed independently of each other.

B. Material Decomposition

The bird phantom was considered to consist of only two basis materials: soft tissue and bone. We computed material decompositions into these bases using only the reconstructions from the E_1 and E_3 to minimize the spectral overlap between

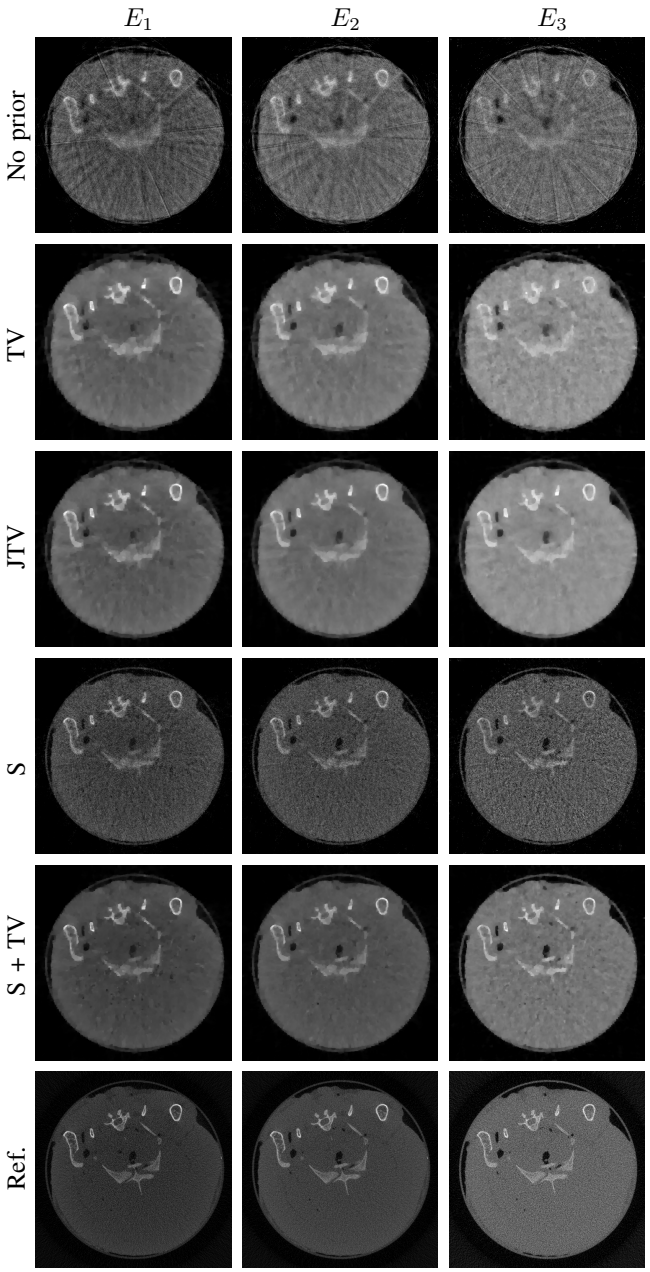


Fig. 2. MECT reconstructions for each energy (E_1, E_2, E_3) using the different regularization techniques.

the datasets, and because two measurements are sufficient to compute a MD into two bases. The X-ray mass attenuation coefficients were acquired from the NIST Standard Reference Database 126 [13], and the X-ray energies were estimated as the weighted averages of the estimated X-ray spectra.

The MD images are shown in Fig. 3. The soft tissue basis has been colour-coded as blue, and the bone basis as green. It is qualitatively clear from the images that the material decompositions computed from reconstructions using structural priors are superior to the ones using no prior or TV, although some artefacts still remain in the bone bases. Notably, in some of the material decomposition images some

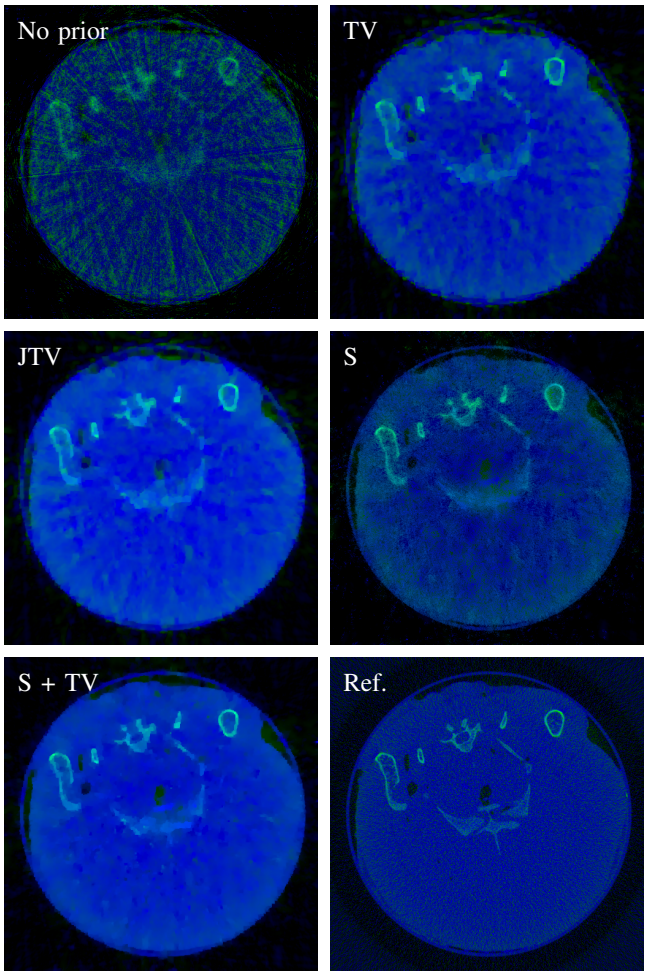


Fig. 3. Material decomposition images for each reconstruction technique. For visualization purposes, gamma correction was applied to the colour channels ($\gamma_{\text{blue}} = 0.6$, $\gamma_{\text{green}} = 0.8$).

TABLE II
HAAR WAVELET-BASED PERCEPTUAL SIMILARITY INDEX (HAARPSI) VALUES FOR THE MECT BASIS MATERIAL IMAGES.

Reconstruction	*Soft tissue*	*Bone*	*Mean*
No prior	0.322	0.309	0.316
TV	0.304	0.361	0.333
JTV	0.316	0.369	0.343
S	0.346	0.409	0.378
S + TV	0.309	0.389	0.349
Ref.	1.000	1.000	1.000

bone details are lost when compared to reconstructions for the individuals energies. This could be due to the fact that the basis material attenuation properties were modelled after human soft tissue and cortical bone. Bone composition in the quail may vary significantly, leading to some losses in the details.

We also computed the HaarPSI values of the tissue bases compared to the ones obtained from the reference reconstructions. The values are shown in Table II. In the bone basis, the reconstructions based on structural priors produce the best results. The HaarPSI values from the soft tissue basis are more ambiguous, but in both cases the S prior yields the best quantitative results.

IV. CONCLUSIONS

We computed low-dose MECT reconstructions using sparse angle sampling, reducing the total number of projections from 2160 to 90. The results demonstrate that using structural priors produces qualitatively and quantitatively superior sparse angle multi-energy reconstructions and material decomposition images compared to reconstructing the images independently with no prior or with total variation. Tissue boundaries are clearly delineated, and details of small bone details are preserved better.

Joint total variation has been used before in CT reconstruction, but to our knowledge, the type of SSIM-based structural prior we have used here has not previously been used in CT reconstruction, and these are also the first results of using this approach combined with material decomposition. We suggest that structural priors show significant potential for low-dose imaging in multi-energy computed tomography.

ACKNOWLEDGEMENTS

This work was supported by the Academy of Finland (Project 312343, Finnish Centre of Excellence in Inverse Modelling and Imaging), the Jane and Aatos Erkko Foundation, and Business Finland project 6614/31/2016.

REFERENCES

[1] Cynthia H McCollough, Shuai Leng, Lifeng Yu, and Joel G Fletcher. Dual- and multi-energy CT: principles, technical approaches, and clinical applications. *Radiology*, 276(3):637–653, 2015.

[2] Jussi Toivanen, Alexander Meaney, Samuli Siltanen, and Ville Kolehmainen. Joint reconstruction in low dose multi-energy CT. *Inverse Problems & Imaging*, 14(4):607–629, 2020.

[3] Samuli Siltanen, Ville Kolehmainen, Seppo Järvenpää, JP Kaipio, P Koistinen, M Lassas, J Pirttilä, and E Somersalo. Statistical inversion for medical x-ray tomography with few radiographs: I. General theory. *Physics in Medicine & Biology*, 48(10):1437–1463, 2003.

[4] Peter Blomgren and Tony F. Chan. Color TV: Total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3), 1998.

[5] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[6] Frank Natterer. *The Mathematics of Computerized Tomography*. John Wiley & Sons, 1986.

[7] Wim van Aarle, Willem Jan Palenstijn, Jan De Beenhouwer, Thomas Altantzis, Sara Bals, K Joost Batenburg, and Jan Sijbers. The astra toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy*, 157:35–47, 2015.

[8] Wim van Aarle, Willem Jan Palenstijn, Jeroen Cant, Eline Janssens, Folkert Bleichrodt, Andrei Dabravolski, Jan De Beenhouwer, K Joost Batenburg, and Jan Sijbers. Fast and flexible x-ray tomography using the astra toolbox. *Optics express*, 24(22):25129–25147, 2016.

[9] Elijah Polak and Gerard Ribiere. Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle. Série rouge*, 3(16):35–43, 1969.

[10] Kyung-Kook Park, William Pavlicek, Thomas Boltz, Robert Paden, Amy Hara, and Metin Akay. Image-based dual energy CT improvements using Gram-Schmidt method. *Proceedings of SPIE*, 7258, 2009.

[11] Tobias Würfl, Mathis Hoffmann, André Aichert, Andreas K. Maier, Nicole Maaß, and Frank Dennerlein. Calibration-free beam hardening reduction in x-ray CBCT using the epipolar consistency condition and physical constraints. *Medical Physics*, 46(12):e810–e822, 2019.

[12] Rafael Reisenhofer, Sebastian Bosse, Gitta Kutyniok, and Thomas Wiegand. A Haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication*, 61:33–43, 2018.

[13] John H. Hubbell and Stephen M. Seltzer. X-ray mass attenuation coefficients: NIST standard reference database 126. https://www.nist.gov/pml/x-ray-mass-attenuation-coefficients, Accessed May 22, 2018.

Scatter Correction for X-ray CT Using Spectral Modulator with Flying Focal Spot

Hewei Gao*, Tao Zhang, N. Robert Bennett, and Adam Wang

***Abstract*—For X-ray computed tomography (CT), spectral modulator with flying focal spot technology (SMFFS) could be a promising low-cost approach to accurately solving the X-ray scattering problem and physically enabling multi-energy imaging in a unified framework, without substantial sparsity or misalignment in data sampling. However, due to the different X-ray spectra during data acquisition, correcting for scatter using SMFFS needs careful consideration and its related algorithm highly depends on the design and implementation of SMFFS. In this work, we proposed two scatter estimation algorithms for SMFFS, where a two-dimensional modulator along with two focal spot positions are of interest. Specifically, a scatter estimation after residual spectrum linearization and another one based on spectral property of scatter are explored, respectively. Physical experiments using a Catphan©600 phantom on a tabletop cone-beam CT system demonstrate the effectiveness of our method with much better HU accuracy and less image artifacts.**

***Index Terms*—Scatter Correction, Computed Tomography, Spectral Modulator, Flying Focal Spot**

I. Introduction

X-RAY scatter is one of the major challenges for wide coverage computed tomography (CT) such as flat-panel detector based cone-beam CT (CBCT), as it could significantly degrade image contrast and cause severe artifacts in reconstructed CT images. In order to improve image quality and promote quantitative imaging in CBCT, numerous scatter correction methods have been developed in the literature [1].

Modulation of the X-ray source in computed tomography (CT) has been actively investigated in both academia and industry in recent years, such as the primary modulator for scatter correction and dual-energy CT [2], [3], the splitting filter in twin-beam dual-energy CT by Siemens [4], the dynamic bowtie filter [5], and the dynamic fluence field modulation [6] and the associated spatial-spectral modulator [7]. Recently, a new concept of multi-energy cone-beam CT using a spectral modulator and flying focal spot technology (SMFFS) has been proposed by our group [8], which has advantages of no significant sparsity or misalignment in data sampling and could potentially enable full-scale cone-beam CT spectral imaging and simultaneous scatter correction capabilities.

Due to the different X-ray spectra during data acquisition, correcting for scatter using SMFFS needs careful consideration and its related algorithm highly depends on the design and implementation of SMFFS. In this work, we proposed two scatter estimation methods for SMFFS, where a two-dimensional modulator along with two focal spot positions are of interest. Specifically, by taking advantage of X-ray physics in SMFFS, a scatter estimation after residual spectrum linearization and another one based on spectral property of scatter are explored, respectively.

Experiments using a Catphan©600 phantom on a tabletop cone-beam CT system are conducted with a copper modulator and by moving the X-ray tube by 1.235 mm in the X direction, to mimic the focal spot deflection. Preliminary results demonstrate the effectiveness of scatter correction from SMFFS.

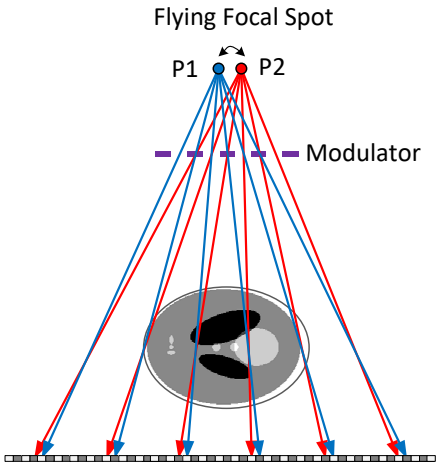


Fig. 1. Data sampling using spectral modulator with flying focal spot technology.

II. Spectral Modulator with Flying Focal Spot

As illustrated in Fig. 1, a spectral modulator is made of partially attenuating blockers and placed between the X-ray source and the scanned object. The modulator attenuates the intensities for some of the X-ray beam and filters their spectra to essentially generate low- and high-energy projection data as well. As a stationary filter with respect to the X-ray source, such a spectral modulator has advantages of simplicity in mechanical design and practical implementations. However, a stationary modulator alone suffers from data sparsity or misalignment of projection rays in terms of spectral data sampling, especially for circular scan trajectories such as cone-beam CT imaging. By moving the source point with flying focal spot technology, one can achieve much better

This project was supported in part by the New Faculty Startup Fund from Tsinghua University, Beijing, China. Hewei Gao and Tao Zhang are with the Department of Engineering Physics, Tsinghua University, Beijing 100084, China; N. Robert Bennett and Adam Wang are with the Department of Radiology, Stanford University, Stanford, CA 94305, USA.

*Author to whom correspondence should be addressed. Email: hwgao@tsinghua.edu.cn.

data sampling with no significant sparsity or misalignment as shown in Fig. 1 and analyzed in [8].

For SMFFS, when a focal spot FS_k whose position is labeled as k fires, the measured X-ray signal at a detector element (i,j) can be expressed as:

$$I_{t_k}(i,j) = I_{p_k}(i,j) + I_{s_k}(i,j)., \tag{1}$$

where, $I_{p_k}(i,j)$ and $I_{s_k}(i,j)$ denote the intensities of primary after modulation and the scatter, respectively, at source position k. The modulated primary, $I_{p_k}(i,j)$, can be further written as:

$$I_{p_k}(i,j) = \int S_k(E,i,j) e^{-u_m(E)T_{m_k}(i,j)} e^{-u_o(E)L_k(i,j)} \mathrm{d}E. \tag{2}$$

Here, $S_k(E,i,j)$ is the X-ray flux at energy E; $u_o(E)$ and $L_k(i,j)$ are the linear attenuation coefficients and the path-length of projection ray of the scanned object; and $u_m(E)$ and $T_{m_k}(i,j)$ are the linear attenuation coefficient and thickness of the modulator, respectively. In general, the modulation function of a spectral modulator in a polychromatic CT system is mainly determined by the material and thickness of the modulator, which can be defined as:

$$m_k(i,j) = \frac{I_{m_k}(i,j)}{I_{ori_k}(i,j)} = \frac{\int S_k(E,i,j) e^{-u_m(E)T_{m_k}(i,j)} \mathrm{d}E}{\int S_k(E,i,j) \mathrm{d}E}, \tag{3}$$

where, $I_{ori_k}(i,j)$ and $I_{m_k}(i,j)$ are the intensities of X-ray beam before and after passing through the modulator.

As investigated in the literature, scatter is mostly dominated by low-frequency components [2], [9]. In SMFFS, it is further assumed that there is a strong correlation between scatter distributions at varying focal spot positions when the flying distance is relatively small [8]. Based on our experience, it is also safe to assume that the low- and high-energy projection rays are aligned quite well (i.e., $L_1(i,j) = L_2(i,j) = L(i,j)$), and their scatter share the same distribution (i.e., $I_{s_1}(i,j) = I_{s_2}(i,j) = I_s(i,j)$). In the following section, two ideas are proposed to estimate $I_s(i,j)$ for each and every detector element (i,j). For simplicity, we omit the expression (i,j) from the variables, e.g., I_{t_k} represents $I_{t_k}(i,j)$ with the source point at position k.

III. Scatter Correction from SMFFS

A. Scatter estimation after Residual Spectrum Linearization (SE_{rsl})

Using the similarity of the scatter distributions between the two focal spot positions, one can eliminate the scatter term I_s in Eq. (1) by a subtraction, i.e.,

$$I_{t_2} - I_{t_1} = \int S_{res}(E) e^{-u_o(E)L} \mathrm{d}E, \tag{4}$$

where, $S_{\mathrm{res}}(E) = S_2(E)e^{-u_m(E)T_{m_2}} - S_1(E)e^{-u_m(E)T_{m_1}}$ is the residual spectrum after subtracting the high-energy spectrum at focal spot position FS_1 from the low-energy one at focal spot position FS_2. In Eq. (4), by assuming a single-material object being scanned, the path-length L can be obtained from a conventional linearization mapping based on the residual spectrum. Therefore, eventually a scatter distribution can be estimated like this:

$$\begin{aligned} I_s &= I_{t_1} - \int S_1(E) e^{-u_m(E)T_{m_1}} e^{-u_o(E)L} \mathrm{d}E \\ &= I_{t_2} - \int S_2(E) e^{-u_m(E)T_{m_2}} e^{-u_o(E)L} \mathrm{d}E. \end{aligned} \tag{5}$$

Thus SE_{rsl} algorithm can be implemented as follows:

STEP-1: Obtain the residual spectrum $S_{res}(E)$ from $S_1(E)$ and $S_2(E)$;
STEP-2: Compute path-length L based on the residual spectrum using Eq. (4);
STEP-3: Get scatter estimation through Eq. (5);

B. Scatter Estimation based on Spectral Compensation for Modulator (SE_{sps})

Based on a spectral property of scatter in Ref. [10], after applying a beam hardening correction (BHC) for $p_{t_k} = -\ln(I_{t_k}/I_{m_k})$ using a mapping function f_k at focal spot position k, when going back to the pre-logarithm domain, I_{t_k} will be essentially changed to I'_{t_k} as

$$\begin{aligned} I'_{t_k} &= I_{m_k} e^{-f_k(p_{t_k})} \\ &\approx I_{m_k} e^{-f_k(p_{p_k})} + I_s f'_k(p_{t_k}) e^{-f_k(p_{t_k}) + p_{t_k}}. \end{aligned} \tag{6}$$

Here, $p_{p_k} = -\ln(I_{p_k}/I_{m_k})$, f'_k is the first derivative of f_k. Therefore, as proposed in Ref. [10] the scatter term can be analytically approximated by

$$I_s \approx \frac{I_{m_1} I_{m_2} \left(e^{-f_1(p_{t_1})} - e^{-f_2(p_{t_2})}\right)}{I_{m_2} f'_1(p_{t_1}) e^{-f_1(p_{t_1}) + p_{t_1}} - I_{m_1} f'_2(p_{t_2}) e^{-f_2(p_{t_2}) + p_{t_2}}}. \tag{7}$$

Here, p_{t_k} is the total projection written as $-\ln \frac{I_{t_k}}{I_{m_k}}$, f_i is the beam hardening correction function and f'_i is the first order derivative of f_i. As a result, scatter estimation can be generated using Eq. (7) directly. Also, SE_{sps} can be implemented in an iterative manner, with corresponding pseudo codes demonstrated as follows:

STEP-1: initial I_s^0, $stop_criteria$ and N_{iter};
STEP-2: for $iter = 1 : N_{iter}$ do
STEP-3: $p_{t_1}^{iter} = -\ln\left(\frac{I_{t_1} - I_s^{iter-1}}{I_{m_1}}\right)$;
STEP-4: $p_{t_2}^{iter} = -\ln\left(\frac{I_{t_2} - I_s^{iter-1}}{I_{m_2}}\right)$;
STEP-5: Take corrected $p_{t_1}^{iter}$ and $p_{t_2}^{iter}$ into Eq. (7) to generate residual scatter ΔI_s;
STEP-6: $I_s^{iter} = I_s^{iter-1} + \Delta I_s$;
STEP-7: if $|\Delta I_s| < stop_criteria$ then stop early and return final $I_s = I_s^{iter}$.
STEP-8: maximum iterations reached and return the final $I_s = I_s^{N_{iter}}$.

After scatter estimation using either SE_{rsl} or SE_{sps}, a smoothing can be applied to the estimated scatter as a typical processing step in scatter correction. Once the estimated scatter is removed from the measured CT data, a spectral compen-

TABLE I
IMAGING PARAMETERS IN OUR EXPERIMENT

	Parameters	Values
X-ray source	Voltage	120 kVp
	Current	10 mA
	Pulse width	13 mS
Geometry	Detector size	0.388 mm
	Source-to-modulator distance	243.9 mm
	Source-to-isocenter distance	860.9 mm
	Source-to -detector distance	1419.6 mm
	Deflection of focal spot	1.235 mm
Copper Modulator	thickness	0.42 mm
	Spacing	0.889 mm

sation for modulator can also be implemented to remove the spectral non-uniformity [11].

It is worth noting that for a modulator type approach the gantry vibration might be an issue in practical applications. The algorithms developed here included an adaptive approach to getting the effective modulator thicknesses across all detector pixels, which can be used to compensate for such a vibration if repeatable. In the literature, a modulator-database based registration and interpolation approach has been developed to resolve the gantry vibration as well [12], [13].

IV. RESULTS

To validate our proposed scatter correction, we conducted experiments on a tabletop cone-beam CT system, where a Catphan©600 phantom (The Phantom Laboratory) was scanned at two focal spot positions separated by 1.235 mm in the X direction. Some imaging parameters are listed in Table I and our experiments were conducted in the following order: at each focal spot position, fan-beam scans with a narrow collimator were first done without the modulator (FB) and with the modulator ($\mathrm{FB_M}$), respectively. Then, cone beam scans with a fully opened collimation were done without modulator (CB) and with the modulator ($\mathrm{CB_M}$). The standard Feldkamp-Davis-Kress (FDK) algorithm was used for reconstruction. The reconstructed CT images are presented in Hounsfield units (HU), with a CT number of -1000 for air and 0 for water-equivalent material. The overall system spectrum with and without modulator for each detector pixel was obtained using the same method as proposed in [11].

It is worth noting that SE_{rsl} and SE_{sps}, are close mathematically, as demonstrated in Fig. 2. However, from practical application and algorithm implementation point of view, they are quite different. The SE_{rsl} method is derived by assuming a "dual-energy" SMFFS data acquisition. It requires an additional linearization using the residual spectrum but can easily extended to triple-energy or more. So it is compatible with material decomposition if doable. On the contrary, the SE_{sps} method is derived based on a spectral property of scatter. It needs no additional linearization beyond the conventional beam hardening correction function.

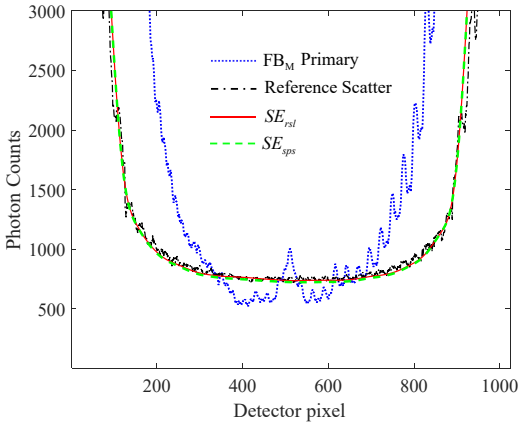


Fig. 2. Scatter profiles of the Catphan©600 phantom. The SE_{rsl} and SE_{sps} methods are used to estimate scatter from cone-beam scan using SMFFS, respectively. The reference scatter is obtained by subtracting the $\mathrm{FB_M}$ from that of the corresponding $\mathrm{CB_M}$, with an assumption that fan-beam scan is free from scatter.

The Catphan©600 phantom reconstruction is shown in Fig. 3. The FB images can be considered to be artifact free and are taken as reference. It is seen that the image quality of $\mathrm{CB_M}$ CT image is greatly improved after scatter correction by the SE_{rsl} and SE_{sps} methods. To quantify the correction performance, in Table II we measure the average CT number for the selected regions of interest (ROIs) within a uniform region indicated in Fig. 3(a). In addition, the root mean squared error (RMSE) for the CT number is calculated:

$$E_{\mathrm{RMSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(u_i - \bar{u}_i)^2} \tag{8}$$

where, i is the index of the selected ROI, u_i is the mean value in HU inside the ROI i and $\bar{u}_i$ is the corresponding value in the reference CT image from the FB scan. As listed in Table II, the E_{RMES} is reduced from 303.2 HU of CB reconstruction to 13.0 HU of $\mathrm{CB_M}$ reconstruction with SE_{rsl}, and to 3.4 HU of $\mathrm{CB_M}$ reconstruction with SE_{sps}.

V. CONCLUSIONS

In this work, we presented two scatter estimation algorithms for X-ray CT using a spectral modulator and the flying focal spot technology. Preliminary experiments using a Catphan phantom validated our method to be effective with better HU accuracy and less image artifacts, with the mean squared error of the selected ROIs reduced from 303.2 HU to 13.0 HU and 3.4 HU, respectively.

REFERENCES

[1] E.-P. Rührnschopf and K. Klingenbeck, "A general framework and review of scatter correction methods in x-ray cone-beam computerized tomography. part 1: Scatter compensation approaches," *Med. Phys.*, vol. 38, no. 7, pp. 4296–4311, 2011.

[2] L. Zhu, N. R. Bennett, and R. Fahrig, "Scatter correction method for x-ray ct using primary modulation: Theory and preliminary results," *IEEE transactions on medical imaging*, vol. 25, no. 12, pp. 1573–1587, 2006.

[3] M. Petrongolo and L. Zhu, "Single-scan dual-energy ct using primary modulation," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1799–1808, 2018.

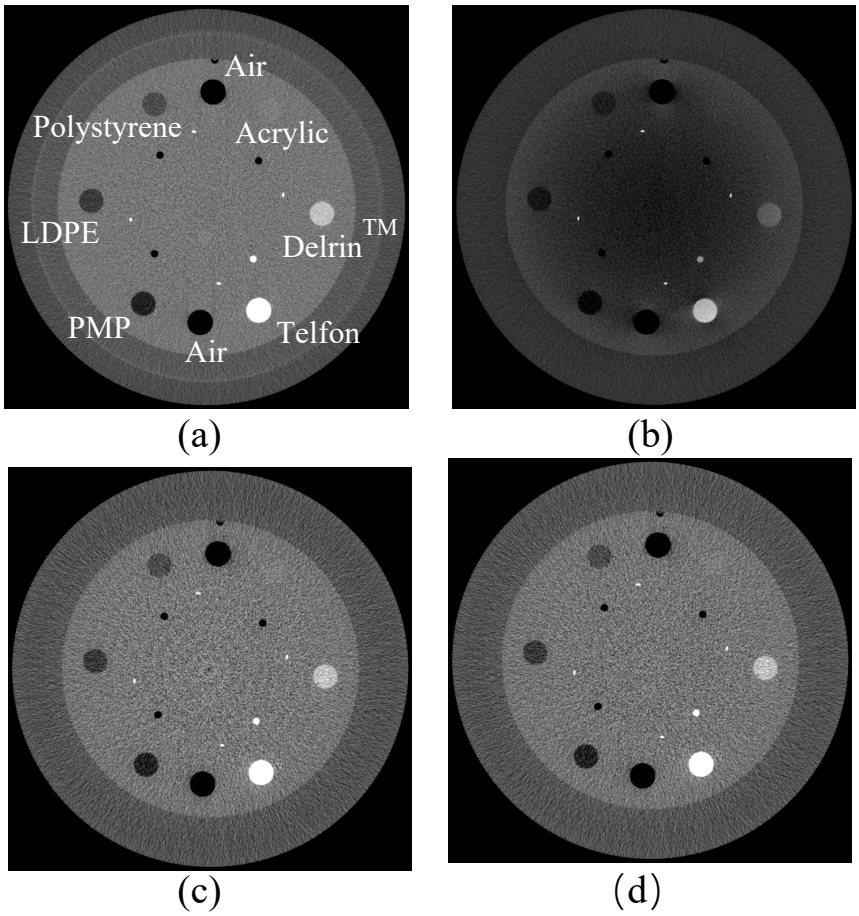


Fig. 3. Reconstructed CT images of Catphan phantom at 120 kVp: (a) FB reconstruction as reference; (b) CB reconstruction; (c) CB_M reconstruction after scatter correction by SE_{rsl} followed by spectral compensation for modulator to get rid of the spectral non-uniformity; (d) CB_M reconstruction after scatter correction by SE_{sps} followed by spectral compensation for modulator to get rid of the spectral non-uniformity; Display window: [-300 HU,500 HU].

TABLE II
AVERAGED CT NUMBER (HU) OF SELECTED ROIS OF THE CATPHAN PHANTOM AT 120 KVP IN FIG. 3.

ROIs	Air	PMP	LDPE	Polystyrene	Acrylic	Delrin^{TM}	Teflon	E_{RMSE}
FB	-953.3	-197.1	-116.9	-61.2	93.3	306.4	904.3	\
CB	-699.5	-282.3	-231.7	-207.4	-123.8	-10.8	280.3	303.2
CB_M + SE_{rsl}	-964.3	-201.2	-120.0	-60.9	94.2	314.3	935.6	13.0
CB_M + SE_{sps}	-950.7	-201.1	-120.7	-63.4	88.6	304.0	907.6	3.4

[4] I. P. Almeida, L. E. J. R. Schyns, M. C. Öllers, W. van Elmpt, K. Parodi, G. Landry, and F. Verhaegen, "Dual-energy ct quantitative imaging: a comparison study between twin-beam and dual-source ct scanners," *Medical Physics*, vol. 44, no. 1, pp. 171–179, 2017.

[5] P. Shunhavanich, S. S. Hsieh, and N. J. Pelc, "Fluid-filled dynamic bowtie filter: Description and comparison with other modulators," *Medical Physics*, vol. 46, no. 1, pp. 127–139, 2019.

[6] G. J. Gang, A. Mao, W. Wang, J. H. Siewerdsen, A. Mathews, S. Kawamoto, R. Levinson, and J. W. Stayman, "Dynamic fluence field modulation in computed tomography using multiple aperture devices," *Phys. Med. Biol.*, vol. 64, no. 10, p. 105024, 2019.

[7] M. Tivnan, S. T. II, and J. W. Stayman, "Physical modeling and performance of spatial-spectral filters for ct material decomposition," *Proc. SPIE*, vol. 10948, 2019.

[8] H. Gao, H. Zhou, L. Zhu, N. Pelc, R. Bennett, and A. Wang, "Spectral modulator with flying focal spot for cone-beam ct: a feasibility study," in *Medical Imaging 2020: Physics of Medical Imaging*, vol. 11312. International Society for Optics and Photonics, 2020, p. 113121P.

[9] H. Gao, R. Fahrig, N. R. Bennett, M. Sun, J. Star-Lack, and L. Zhu, "Scatter correction method for x-ray ct using primary modulation: Phantom studies," *Medical physics*, vol. 37, no. 2, pp. 934–946, 2010.

[10] T. Zhang, Z. Chen, H. Zhou, R. Bennett, A. Wang, and H. Gao, "A spectral property of scatter in x-ray ct," *(submitted)*.

[11] H. Gao, L. Zhang, R. Grimmer, and R. Fahrig, "Physics-based spectral compensation algorithm for x-ray ct with primary modulator," *Physics in Medicine & Biology*, vol. 64, no. 12, p. 125006, 2019.

[12] B. Bier, K. Müller, M. Berger, J.-H. Choi, R. Fahrig, and A. Maier, "Scatter correction for C-arm CT using primary modulation," in *Proc. 4th Intl. Mtg. on image formation in X-ray CT*, 2016, pp. 383–6.

[13] B. Bier, M. Berger, A. Maier, M. Kachelrieß, L. Ritschl, K. Müller, J.-H. Choi, and R. Fahrig, "Scatter correction using a primary modulator on a clinical angiography c-arm ct system," *Medical Physics*, vol. 44, no. 9, pp. e125–e137, 2017.

A phantom Study on Exploring the Accuracy of the Weighted CT Dose Index for Multi-detector CT

Jingwen Zhuang, Yaguang Lu, Mei Bai

***Abstract*—The main objective of the current research was to investigate the accuracy of CT dose index (CTDI) for MDCT with an enlarged beam width using polymethylmethacrylate (PMMA) phantom and to explore integration length necessary to assess radiation dose delivered to patients preliminarily. Dose profiles in PMMA phantom for three different CT scanners were derived in helical mode using 100, 120,140 tube voltages and 4.8mm~57.6mm collimations. The CTDIs for center, periphery and weighted with 100mm, 150mm and 200mm integration lengths and the corresponding CTDI efficiencies were calculated. $CTDI_{100,w}$ efficiencies of head phantom are 88.32%±0.82%, 88.26%±1.76% and 82.59%±1.04% for 0~20mm, 20mm~40mm and above 40mm collimations, while the $CTDI_{100,w}$ efficiencies of body phantom are 88.15%±1.47%, 87.45%±1.29% and 78.28%±1.08%. The error of $CTDI_{100,w}$ for body phantom can even reach 22.69% for 57.6mm beam collimation. The weighted $CTDI_{150}$ efficiencies of head and body phantoms are 95.72%~98.64% and 93.37%~99.83%, respectively, while the weighted $CTDI_{200}$ efficiencies can reach 99.04%~100.00% and 98.87%~100.00%, respectively. The $CTDI_{100,w}$ efficiencies decrease when beam collimations are increased and the error of $CTDI_{100,w}$ rises when beam collimations are widened. For MDCT with wider beam collimations, $CTDI_{100,w}$ should be modified. Extending the integration length of CTDI could improve the accuracy of CTDI. After initial exploration, for MDCT with collimation shorter than 60mm, using CTDI with 200mm integration length to evaluate the radiation dose would be more appropriate.**

***Index Terms*—MDCT; radiation dose; CTDI**

I. Introduction

As the continuously upgrade, CT has been widely used in medical procedures and has become an indispensable tool for medical diagnosis and treatment [1]. However, the rising radiation dose to the patient population from CT examinations has drawn more and more attention of the public [2]. Therefore, it is important to do further research on reducing radiation dose from CT examinations [3]. To reduce radiation dose delivered to patients, first of all, convenient and accurate dose measurement is essential.

CT dose index (CTDI), as a widely used metric for the dose performance of a CT scanner, is defined to characterize the radiation dose properties [4]. CTDI was originally defined as the integral of the dose profile with an integration length from -∞ to +∞ along z axis[5], i.e. $CTDI_{\infty}$. $CTDI_{\infty}$ can evaluate the radiation dose accurately but is obtained difficultly. In order to simplify measurement, $CTDI_{100}$ was introduced and its integration length is normalized to 100mm along z axis, and it currently becomes one of the recognized indicators which can reflect the radiation dose of CT scans [6]. In addition to $CTDI_{100}$, weighted CTDI ($CTDI_w$) and volume CTDI ($CTDI_{vol}$), both calculated using $CTDI_{100}$, were also frequently used to reflect the radiation dose of CT scans.

$CTDI_{100}$, as one of the most important indicators to reflect the level of radiation dose, its accuracy has been discussed by some studies [7-10]. Mori et al found that $CTDI_{100}$ was 60% of $CTDI_{\infty}$ at the center of phantom and 80% of $CTDI_{\infty}$ at the periphery of phantom, respectively, while the body phantom was scanned using 10mm beam width on a prototype 256-slice CT scanner[11]. The study of Nakonechny et al indicated that the ratio of $CTDI_{\infty}$ to $CTDI_{100}$ is about 1.3 at the center of body phantom while using 3–20 mm beam widths [12]. Gancheva et al reported that for CT imaging system，CTDI with an integration length of 100mm underestimated the results, and the underestimation of $CTDI_{vol}$ is about 14 % while beam width was 32mm [13]. These studies have made great efforts to explore the accuracy for CT dose measurement. So far as we concerned there are much less studies which investigate the accuracy of $CTDI_{100}$ systematically by experiment through using different scan conditions and different CT scanners.

In the current study, $CTDI_{\infty}$ and CTDI with different integration lengths were obtained using polymethylmethacrylate (PMMA) phantom with different scan parameters and CT scanners. To investigate the accuracy of CTDI with different integration length for multi-detector CT, the CTDI with different integration lengths was compared with the corresponding $CTDI_{\infty}$. In this way, the necessary integration length would also be acquired preliminarily. The accuracy of radiation dose assessment would therefore be improved in the following clinical research and applications.

. This work is supported by the National Natural Sciences Foundation of China (NSFC) grant no.81372923.

Jingwen Zhuang is with the Xuanwu Hospital, Capital Medical University, Beijing, 100053 (e-mail: jswei65@163.com).

Yaguang Lu is with the Xuanwu Hospital, Capital Medical University, Beijing, 100053 (e-mail: xwhosp@ sina.com).

Mei Bai is with the Xuanwu Hospital, Capital Medical University, Beijing, 100053 (e-mail: baimei@xwhosp.org).

II. MATERIALS AND METHODS

Measurements were performed on three different CT scanners: Siemens SOMATOM Definition Flash CT scanner (Siemens AG, Forchheim, Germany) with 64-row-detector, and Siemens SOMATOM Force CT scanner with 96-row-detector (Siemens AG, Forchheim, Germany). To obtain dose profiles the Piranha x-ray multi-meter equipped with a CT-SD 16 radiation detector (RTI Electronics AB, Mölndal, Sweden) was used in the current study. The CT-SD 16 detector is shaped like a CT pencil ionization chamber, and a submillimeter-thick solid-state radiation detector was placed in the middle of it. Using CT-SD 16 to measure CTDI is equivalent to using ionization chamber [14]. Two standard dosimetry PMMA phantoms with a length of 150mm were used: one head phantom with a diameter of 160mm and one body phantom with a diameter of 320mm. Both phantoms contain five holes: one is located in the center and others are around the periphery at "12 o'clock", "3 o'clock", "6 o'clock", "9 o'clock", respectively.

A. *Dose profiles acquisition*

The PMMA phantom was placed on the examining table of the CT scanner with the CT-SD 16 detector inserted into one of the center or peripheral holes, the CT dose data were collected by a computer and used to generate a dose profile, and 'corrected' option on the computer software was selected to make the dose profile corrected with respect to the angular dependence. The dose profile of x-ray beam was derived by scanning the whole phantom in a helical mode. The dose profiles of the central position and the four peripheral positions using the same exposure parameters were measured successively.

Some scan parameters were unchanged for all of the three CT scanners in the whole process: 160mA tube current, 1.0 pitch (or approximately equal to 1.0), 1.0s rotation time and 5mm slice thickness, while others were altered. For GE Discovery CT750 HD scanner the dose profiles were obtained using the following altered parameters: 100, 120 and 140 tube voltages, 20mm and 40mm beam collimations. For Siemens SOMATOM Definition Flash CT scanner, the dose profiles were derived using 100, 120 and 140 tube voltages, 16×0.3mm, 16×0.6mm, 20×0.6m, 40×0.6mm, 32×1.2mm and 128×0.6mm nominal beam collimations. And for Siemens SOMATOM Force CT scanner the dose profiles were derived using 100, 120 and 140 tube voltages, 64×0.6m, 128×0.6mm, 48×1.2mm and 192×0.6mm nominal beam collimations.

B. *CTDI and $CTDI_w$ with different integration lengths*

CTDI with different integration lengths ($CTDI_L$) were determined by [15]

$$CTDI_L = \frac{1}{T}\int_{-L/2}^{L/2} D(z)dz \qquad (1)$$

Where L is the dose integral length whose unit is mm, T is the beam collimation width at the isocenter, and D(z) is the dose profile along z axis. $CTDI_{100}$, $CTDI_{150}$, and $CTDI_{200}$ of different CT scanners with different scan parameters were calculated by the corresponding dose profiles using the equation above. $CTDI_\infty$ were derived by computing the integration, and the integration length was the length of the whole dose profile.

$CTDI_w$ with different integration lengths ($CTDI_{L,\,w}$) were determined by

$$CTDI_{L,w} = \frac{1}{3}CTDI_{L,c} + \frac{2}{3}CTDI_{L,p} \qquad (2)$$

where $CTDI_{L,\,c}$ and $CTDI_{L,\,p}$ represent CTDI with integral length L measured at the center and periphery of the PMMA phantoms, respectively. $CTDI_{100,\,w}$, $CTDI_{150,\,w}$, $CTDI_{200,\,w}$ and $CTDI_{\infty,\,w}$ of different CT scanners with different scan parameters were calculated using the equation above in the current study.

C. *Comparison between $CTDI_{100}$, $CTDI_{150}$, $CTDI_{200}$ and $CTDI_\infty$*

To compare CTDI100, CTDI150, CTDI200 with CTDI∞, respectively, CTDI100 efficiency, CTDI150 efficiency and CTDI200 efficiency were calculated as [16]:

$$CTDI_{L,j}efficiency = \frac{CTDI_{L,j}}{CTDI_{\infty,j}} \qquad (3)$$

Where j= c for the center, p for the periphery and w for the weighted, and L is the dose integral length (i.e. 100, 150, 200). CTDI100,w efficiency, CTDI150,w efficiency and CTDI200,w efficiency were calculated, meanwhile CTDI100 efficiency, CTDI150 efficiency and CTDI200 efficiency at the center and periphery of the PMMA phantoms were also obtained according to equation 3 in the current study.

III. RESULTS

The dose profiles of position A (i.e. isocenter) determined using head phantom at 120kV tube voltage and four different nominal collimations for Siemens SOMATOM Force CT scanner are displayed in figure 1. As is shown in figure 1, the dose profiles reach peak values at the point of 114mm, 118mm, 125mm and 113mm along z axis, and the lengths of the dose profiles along z axis are approximately 174mm, 199mm, 223mm and 224mm for 64×0.6m, 128×0.6mm, 48×1.2m, 192×0.6mm nominal collimations, respectively. It indicates that the length of the dose profile is increasing with the nominal collimation widening.

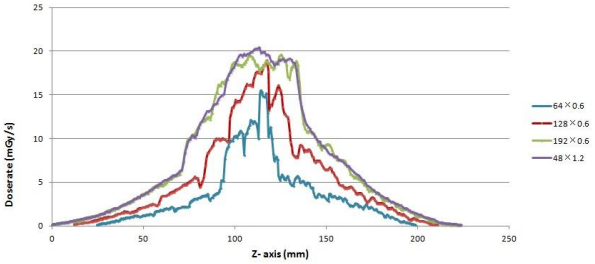


Fig.1 The dose profiles of position A determined using head phantom at 120kV tube voltage, and different collimations for Siemens SOMATOM Force CT scanner

The $CTDI_{100,\,w}$ efficiencies, the $CTDI_{100,\,c}$ efficiencies and the $CTDI_{100,\,p}$ efficiencies acquired using the head phantom at 120kVp are illustrated in Figure 2(a). According to the figure, the values of the $CTDI_{100,\,w}$ efficiencies for 4.8mm, 6.0mm, 12.0mm, 19.2mm, 38.4mm and 57.6mm beam collimations are 88.81%, 88.27%, 88.20%,88.59%, 86.54% and 81.76%. Figure 2(b) shows the $CTDI_{100,\,w}$ efficiencies, the $CTDI_{100,\,c}$

efficiencies and the $CTDI_{100,\,p}$ efficiencies for the body phantom at 120kVp.

The $CTDI_{100,\,w}$ efficiencies are decreasing with increased beam collimation. As shown in table 1, the values of the $CTDI_{100,\,w}$ efficiencies determined with the head and the body phantom are 88.81%~89.59% and 88.01%~ 91.55% for 4.8mm beam collimation, 86.54%~87.61% and 86.57%~ 88.92% for 38.4mm beam collimation. Remarkably, these values are only 81.76%~ 83.75% and 77.31% ~79.45% for 57.6mm beam collimation. Moreover, the $CTDI_{100}$ efficiencies for the center are lower than the $CTDI_{100}$ efficiencies for the periphery, while the weighted $CTDI_{100}$ efficiencies are intermediate.

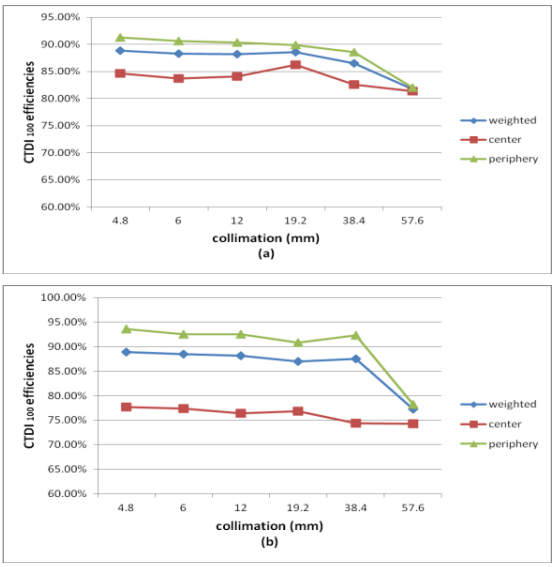


Fig.2 For the Siemens CT scanners, the CTDI100, w efficiencies are decreasing with increased beam collimation. Graphs show the $CTDI_{100,\,w}$ efficiencies, $CTDI_{100,\,c}$ efficiencies and $CTDI_{100,\,p}$ efficiencies determined (a) using head phantom, (b) using body phantom at different beam collimations

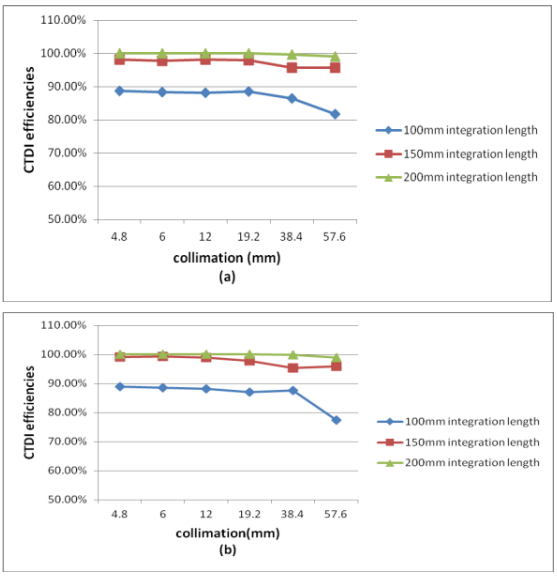


Fig.3 Graphs show the $CTDI_{100,\,w}$ efficiencies, $CTDI_{150,\,w}$ efficiencies and $CTDI_{200,\,w}$ efficiencies for Siemens CT scanners determined with 120kVp and different beam collimations (a) for head phantom, (b) for body phantom.

TABLE I

THE $CTDI_{100,\,w}$ EFFICIENCIES DETERMINED USING DIFFERENT TUBE VOLTAGES MEASURED IN HEAD AND BODY PHANTOMS FOR THE SIEMENS CT SCANNERS.

Beam collimation (mm)	$CTDI_{100,\,w}$ efficiency (%)					
	100KV		120KV		140KV	
	head phantom	body phantom	head phantom	body phantom	head phantom	body phantom
4.8	89.59	91.55	88.81	88.93	88.82	88.01
6	89.11	90.06	88.27	88.44	87.70	87.69
12	88.74	89.34	88.20	88.18	87.63	87.76
19.2	88.69	87.52	88.59	87.03	88.14	85.83
38.4	87.61	88.92	86.54	87.53	86.69	86.57
57.6	83.75	79.45	81.76	77.31	82.26	78.08

Figure 3 shows the $CTDI_{100,\,w}$ efficiencies, $CTDI_{150,\,w}$ efficiencies and $CTDI_{200,\,w}$ efficiencies for Siemens CT scanners determined with 120kVp and different beam collimations for head and body phantom. It could be seen from figure 3, the values of $CTDI_{100,\,w}$ efficiencies for 120kV are 81.76%~88.81% and 77.31%~88.93% for head and body phantom. Table 2 and table 3 lists the $CTDI_{150,\,w}$ efficiencies and the $CTDI_{200,\,w}$ efficiencies determined using different tube voltages measured in head and body phantoms for the Siemens CT scanners. As can be seen from table 2 and table 3, the values of $CTDI_{150,\,w}$ efficiencies are found to be 95.72%~98.41% and 95.02%~99.83% for head and body phantom, while the values of $CTDI_{200,\,w}$ efficiencies can reach 99.04%~100% and 98.87%~100.00% for head and body phantom, which signifies that the error of $CTDI_{150}$ is 0.27%~4.98%, while the error of $CTDI_{200}$ is always below 1.13%.

TABLE 2 THE $CTDI_{150,\,w}$ EFFICIENCIES DETERMINED USING DIFFERENT TUBE VOLTAGES MEASURED IN HEAD AND BODY PHANTOMS FOR THE SIEMENS CT SCANNERS.

Beam collimation (mm)	$CTDI_{150,\,w}$ efficiency (%)					
	100KV		120KV		140KV	
	head phantom	body phantom	head phantom	body phantom	head phantom	body phantom
4.8	98.41	99.83	98.06	99.04	97.92	98.76
6	98.03	99.49	97.85	99.23	97.64	98.87
12	98.26	99.50	98.14	98.99	97.77	98.94
19.2	98.34	98.21	98.03	97.67	98.14	97.44
38.4	95.96	95.35	95.61	95.38	95.72	95.02
57.6	95.74	96.75	95.74	95.94	95.76	96.13

TABLE 3 THE $CTDI_{200,\,w}$ EFFICIENCIES DETERMINED USING DIFFERENT TUBE VOLTAGES MEASURED IN HEAD AND BODY PHANTOMS FOR THE SIEMENS CT SCANNERS.

Beam collimation (mm)	$CTDI_{150,\,w}$ efficiency (%)					
	100KV		120KV		140KV	
	head phantom	body phantom	head phantom	body phantom	head phantom	body phantom
4.8	100.00	100.00	100.00	100.00	100.00	100.00
6	100.00	100.00	100.00	100.00	100.00	100.00
12	100.00	100.00	100.00	100.00	100.00	100.00
19.2	100.00	99.96	100.00	100.00	100.00	100.00
38.4	99.51	99.58	99.58	99.88	99.71	99.79
57.6	99.37	99.18	99.04	98.87	99.06	98.99

IV. DISCUSSION

It's noticeable that the CTDI100, w efficiencies are decreasing with increased beam collimations. With the increase of the beam width of multi-detector CT (MDCT) system, CTDI100 with an integration length only 100mm will underestimate the level of radiation dose, because the beam has scatter tails that are larger than the total integration length [14]. In figure 1, the dose profiles of 48×1.2m and 192×0.6mm nominal collimations cover a length of approximately 225mm and have obvious scatter tails. The area under the dose profile, which is almost a part of the scatter tails and is not in the integration range of CTDI100 has been neglected. The wider the beam width is, the larger the area overlooked becomes, consequently, the CTDI100, w efficiencies are decreasing. The reduction of CTDI100, w efficiencies means that the difference between CTDI100, w and CTDI∞, w is increasing with nominal collimation increasing. In other words, the accuracy of CTDI100 was declining with nominal collimation increasing, and using CTDI100 to evaluate the radiation dose of MDCT could underestimate the result. Moreover, according to the causes, the accuracy of CTDI is influenced by its integration length. Altering integration length could change the neglected area under the dose profile. Thus altering integration length has an impact on the accuracy of CTDI.

It could be seen that the extension of integration length of CTDI could improve the CTDI efficiency enormously and made CTDI more accurate. According to the results, the weighted CTDI efficiency determined using head phantom is growing by 8.81%~13.97% and 10.41%~17.28% with integration length extending from 100mm to 150mm and 200mm, respectively, and the weighted CTDI efficiency determined using body phantom is increasing by 6.42%~18.63% and 8.45%~21.56%, respectively. The increase of weighted CTDI efficiency signifies the raise of accuracy of CTDI. After initial exploration, for MDCT with collimation shorter than 60mm, using CTDI with 200mm integration length to evaluate the radiation dose would be reasonable. The results reveals that for wider collimations such as 57.6mm, 40mm and 38.4mm, the values of CTDI150, w efficiencies are more than 95%, and the values of the CTDI200, w efficiencies can exceed 99%. For other narrower collimations, the values of the CTDI150, w efficiencies can reach above 97%, while the values of the CTDI200, w efficiencies are almost 100%. It implies that while beam collimations are not exceeding 20mm, CTDI200, w is very accurate, and while beam collimations are between 20mm and 60mm the maximum error of CTDI200, w is lower than 1%.

Some studies have also investigated the influence of phantom length on CTDI measurement. The simulation results obtained by Li et al demonstrated that when the pencil ionization chamber was placed at the phantom center, the differences between using a 15cm length phantom and using a 175cm phantom to measure CTDI100 would not exceed 8% for 32cm diameter phantom, while the differences were only 0.9%~1.5% for 16cm diameter phantom [15]. Extending the length of the PMMA phantom will not only increase the costs but also make the quality assurance inconveniently. Therefore, phantoms with 15cm length were used in the current study. However, these results were obtained by simulation. We will validate the accuracy of these results by experiments in the future.

V. CONCLUSION

The accuracy of CTDI should be attached more importance to. In this work we detected that the $CTDI_{100,w}$ efficiencies are decreasing with increased beam collimations. These results are in line with the results reported previously. The consistency confirms the validity of the results in this work. Moreover, the accuracy of CTDI was found to be influenced by its integration length. For MDCT with wider beam collimations, especially exceeding 40mm, $CTDI_{100,w}$ should be modified. Extending the integration length of CTDI (i.e. extending the length of ionization chamber) could be an effective to improve the accuracy of CTDI. After preliminary exploration, for MDCT with collimation shorter than 60mm, using CTDI with 200mm integration length to evaluate the radiation dose would be more appropriate. Improving the accuracy of CT radiation dose assessment could make multi-detector CT scanners work more safely and effectively, prevent medical accidents and promote radiological diagnosis technology.

REFERENCES

[1] Weir VJ, Zhang J, Bruner AP(2015) Dosimetric characterization and image quality evaluation of the AIRO mobile CT scanner. J Xray Sci Technol 23:373-81.

[2] Brenner DJ, Hall EJ (2007) Computed tomography – an increasing source of radiation exposure. N Engl J Med 357: 2277-2284

[3] ICRP. ICRP Publication 87 (2001) Managing Patient Dose in Computed Tomography. Oxford：Pergamon Press.

[4] Boone JM (2007) The trouble with CTDI100. Med Phys 34:1364-1371.

[5] Shope TB, Gagne RM, Johnson GC (1981) A method for describing the doses delivered by transmission x-ray computed tomography. Med Phys 8:488-495.

[6] Li X, Zhang D, Liu B (2012) Estimation of the weighted CTDI∞ for multislice CT examinations . Med Phys 39: 901-905.

[7] Dixon RL (2003) A new look at CT dose measurement: beyond CTDI. Med phys 30:1272-1280.

[8] McCollough CH (2006) It is time to retire the computed tomography dose index (CTDI) for CT quality assurance and dose optimization. Med Phys 33:1190-1191.

[9] Perisinakis K, Damilakis J, Tzedakis A, Papadakis A, Theocharopoulos N, Gourtsoyiannis N (2007) Determination of the weighted CT dose index in modern multi-detector CT scanners. Phys Med Biol 52：6485-6495.

[10] Ruan C, Yukihara EG, Clouse WJ, Gasparian PB, Ahmad (2010) Determination of multislice computed tomography dose index (CTDI) using optically stimulated luminescence technology. Med Phys 37: 3560-3568.

[11] Mori S, Endo M, Nishizawa K.et al (2005) Enlarged longitudinal dose profiles in cone-beam CT and the need for modified dosimetry 32:1061-1069.

[12] Nakonechny KD, Fallone BG, Rathee S (2005) Novel methods of measuring single scan dose profiles and cumulative dose in CT. Med phys 32:98-109.

[13] Gancheva M, Dyakov I, Vassileva J, Avramova-Cholakova S, Taseva D (2015) Dosimetry methods for multi-detector computed tomography. Radiation Protection Dosimetry 165: 190–193.

[14] Bauhs JA, Vrieze TJ, Primak AN, Bruesewitz MR, McCollough CH (2008) CT dosimetry: comparison of measurement techniques and devices. Radiographics 28:245-53.

[15] Li X , Zhang D, Liu B (2011) A practical approach to estimate the weighted CT dose index over an infinite integration length. Phys Med Biol 56:5789-803.

Isotropic Ultra-High Resolution CBCT Extremity Imaging Using a Robotic Radiography System

Ludwig Ritschl, Marcel Beister, Magdalena Herbst and Steffen Kappler

In this paper we evaluate a new cone-beam CT scanning technique using a robotic radiography system equipped with a dynamic fluoroscopic detector. The method is capable of imaging the upper extremities – hand, wrist and elbow joint with quasi-isotropic ultra-high resolution.
The major focus during design of trajectory and acquisition protocol was maximizing the effectively achievable spatial resolution at clinically relevant patient dose levels. This was done by combining a fast, un-binned detector readout mode with an asymmetric scanning trajectory. We selected a larger X-ray tube to isocenter distance than detector to isocenter distance to balance the effect of the focal spot size being larger than the detector pixel size.
In measurements with technical phantoms we were able to demonstrate MTF values of up to 33 lp/cm (at 2% MTF). We also investigate 3D MTF and 3D NPS measurements supported by a theoretical model based on cascaded systems analysis to demonstrate the high isotropy of the system's spatial resolution capabilities.

***Index Terms*—CBCT, MSK imaging, high resolution, robotic system.**

I. Introduction

In the last decade cone-beam computed tomography (CBCT) gained an increasing role in diagnostic imaging. This was mainly driven by the release of several dedicated scanners for extremity imaging. Main arguments for CBCT imaging of the extremities instead of using MDCT (multi-detector computed tomography) are low radiation dose, high isotropic spatial resolution and its availability and cost efficiency [1,2].
Fieselmann et al. [3] showed that CBCT can also be performed on a radiography device which is equipped with a dynamic detector and combined with a robotic arm holding the detector (Figure 1). One realization of such a system is Siemens Multitom RAX (Siemens Healthineers AG, Forchheim, Germany). This system has the benefit of combining a standard radiography room with advanced CBCT functionality. From a user's perspective this setup might be used for optimized diagnostic pathways within the radiology department or releasing the workload of other CT scanners.
Early investigations showed, that diagnostic CBCT on Multitom RAX is very suitable for MSK imaging [4,5] especially for the extremities. Main clinical motivations are fracture detection in a trauma setup, surgery planning and the unique capability of imaging patients in the upright weight bearing position. Especially imaging of the wrist and carpal bones has a very high importance since fractures in these body regions have a very high incidence relatively to other fractures [6,7].
In this paper we investigate a new dedicated CBCT scan mode for the upper extremities, meaning the hand, wrist and elbow. When designing the acquisition mode both, ease of patient positioning, patient comfort and optimizing physical imaging performance have been accounted for.

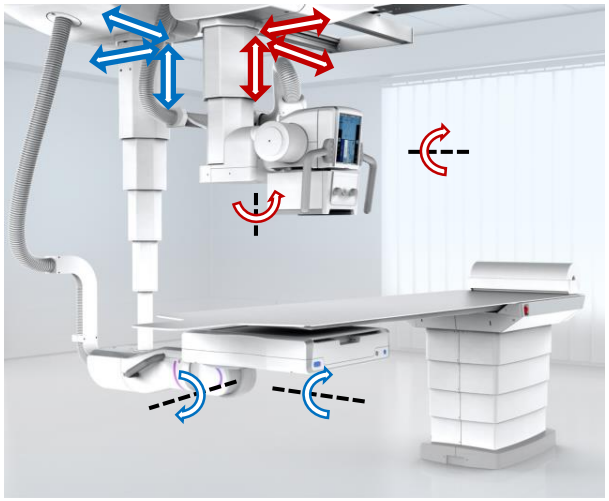


Fig. 1. Schematic illustration of the motion capabilities of Multitom RAX. The system consists of two robotic arms with 10 degrees of freedom.

II. Materials and Methods

A. Cascaded systems model

The system MTF is influenced by the focal spot, the detector and the reconstruction algorithm.

$$\mathrm{MTF}(\rho) = \mathrm{MTF}_{\mathrm{Focus}}(\rho) \cdot \mathrm{MTF}_{\mathrm{Detector}}(\rho) \cdot \mathrm{MTF}_{\mathrm{Reco}}(\rho), \quad (1)$$

where ρ is the spatial frequency.

Here

$$\mathrm{MTF}_{\mathrm{Focus}}(\rho) = e^{-\pi(b_F \rho)^2} \quad (2)$$

is the focus MTF assuming a Gaussian distribution of the focus with

$$b_{\mathrm{F}} = \sigma_F \frac{\mathrm{OID}}{\mathrm{SID}}$$

focus size projected to the isocenter. Here OID (object to imager distance) and SID (source to imager distance) define the magnification properties of the setup.

The detector MTF can accordingly be modeled by

$$\mathrm{MTF}_{\mathrm{Detector}}(\rho) = \left|\frac{\sin(\pi b_{\mathrm{D}}\rho)}{\pi b_{\mathrm{D}}\rho}\right| \quad (3)$$

with

$$b_{\mathrm{D}} = \mathrm{du} \cdot \frac{\mathrm{SOD}}{\mathrm{SID}}$$

being the aperture of a detector pixel of size du after scaling it to the isocenter. SOD is the source to object distance. Equation 3 shows the typical influence of the detector assuming an ideal rectangular aperture of width b. In the full abstract we will also consider the influence of the scintillator material on the detector MTF.
Here it can be seen, that by adjusting SOD and OID the influence of the single components on the final MTF can be adjusted. The full abstract will show a deeper analysis of parameter optimization with respect to optimization of spatial resolution.

B. *Acquisition protocol*

The acquisition is performed using a circular trajectory over 200° around an isocenter. The position of the isocenter is defined in such a way that the patient can put his arm on a rest mounted on the side of the patient's table. This enables also imaging of immobile patients. The detector is set to an active matrix of 1436x1436 un-binned pixels with a size 149µ. This leads to an effective detector size of 21x21 cm. The detector readout rate is 30 fps. The scan time is 12s, which leads to about ~350 projections for reconstruction.
The focal spot of the X-ray tube is 0.6 mm (according to IEC guidelines).
Following the ideas in Section A the source to object distance was set to $\mathrm{SOD} = 850$ mm and the object to imager distance to $\mathrm{OID} = 300$ mm.

C. *Image Reconstruction*

Due to the magnification of the system the reconstructed field of view is about 16 cm. For reconstruction a filtered backprojection based on the implementation of Feldkamp-Davis-Kress (FDK) [8] has been used.
To mitigate influence of smaller system vibrations which might become visible at such high spatial resolutions all reconstructed data were additionally corrected for motion with an entropy-based motion correction algorithm [9,10].

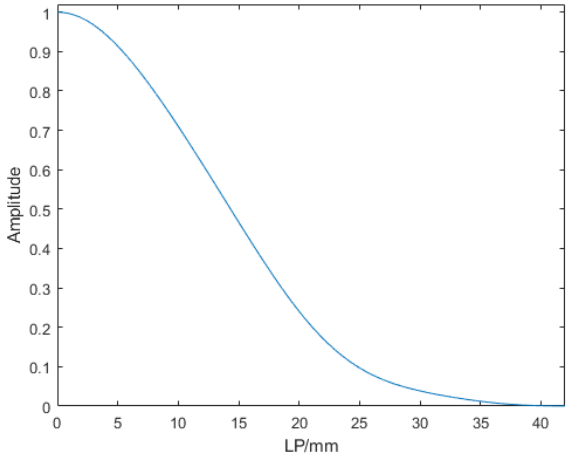


Figure2: Modulation transfer function (MTF) of the scan mode.

III. Results

A. *In-plane MTF measurements*

For MTF computation a wire phantom (QRM, Möhrendorf, Germany) consisting of a tungsten wire with a diameter of 50µ has been scanned with the proposed protocol. The derived system MTF is shown in Figure 2. A 2% MTF value of 33 LP/mm allows for reconstructing at small voxel sizes.

B. *3D MTF and 3D NPS measurements*

The full abstract will consist of measurements of the 3D system MTF and the 3D noise power spectrum. Both measurements will be motivated theoretically by a deeper analysis based on the cascaded systems analysis introduced in chapter 2.

C. *Anthropomorphic phantom measurements*

In addition, we performed scans of an anthropomorphic wrist phantom. Figure 3 shows reconstructions of axial, coronal and sagittal planes. The voxel size is 0.2 mm³. The visual appearance of the images indicates isotropic spatial resolution. In the full abstract this will be investigated quantitatively.

IV. Discussion and Conclusions

We evaluated a novel quasi-isotropic ultra-high resolution scanning mode dedicated for upper extremity imaging using a robotic radiography system. MTF measurements show values beyond 30 LP/mm. Scans of anthropomorphic phantoms and qualitative evaluation indicate highly isotropic properties of the system's spatial resolution.
The proposed method might have significant impact on the diagnostic pathway and the final treatment of patients who are affected by trauma of the upper extremities since it enables for fast tomographic imaging in the radiography suite with excellent image quality.

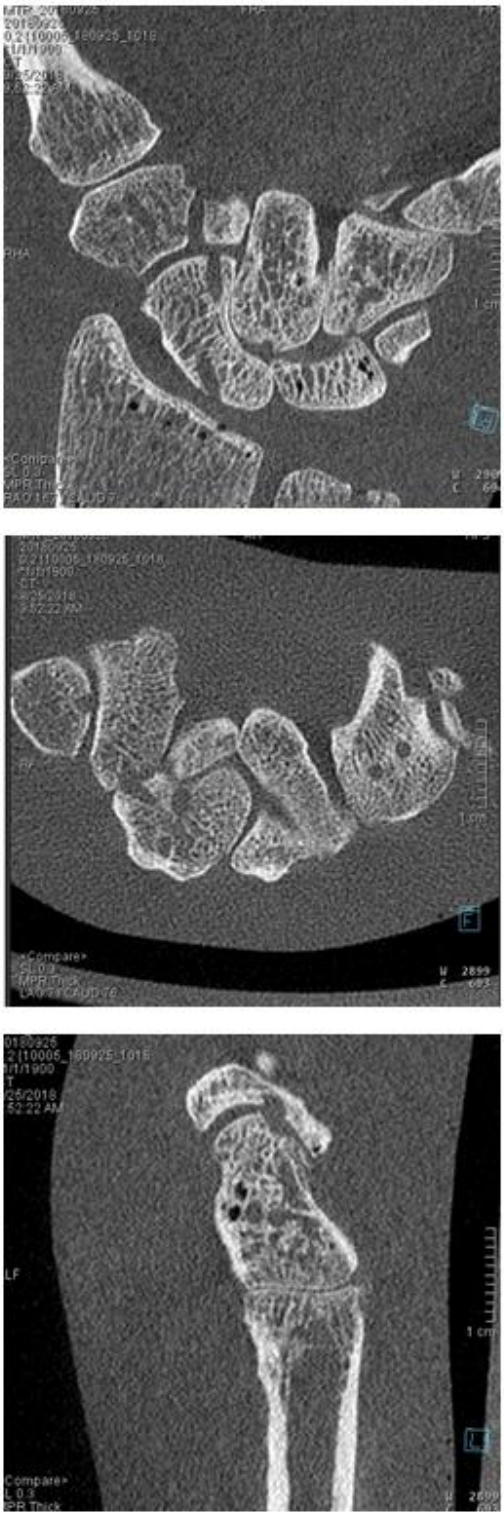


Figure 3: From top to bottom: Coronal, axial and sagittal slices of the reconstruction of an anthropomorphic wrist phantom. The details in the sagittal and coronal slice indicate a high z-resolution of the system.

V: DISCLAIMER

The presented method is not commercially available. Due to regulatory reasons, the Siemens Multitom RAX is not available in all countries, its future availability cannot be guaranteed.

REFERENCES

[1] Posadzy, Magdalena et al. “Cone beam CT of the musculoskeletal system: clinical applications.” *Insights into imaging* vol. 9,1 (2018): 35-45.

[2] A. Fieselmann et al., “Twin robotic x-ray system for 2D radiographic and 3D cone-beam CT imaging,” in *Proc. SPIE 9783, Medical Imaging 2016: Physics of Medical Imaging*, 2016, p. 97830G.

[3] Zbijewski, W et al. “A dedicated cone-beam CT system for musculoskeletal extremities imaging: design, optimization, and initial performance characterization.” *Medical physics* vol. 38,8 (2011)

[4] Grunz, J.-P. et al. "3D cone-beam CT of the ankle using a novel twin robotic X-ray system: Assessment of image quality and radiation dose," *European Journal of Radiology*, Volume 119, 2019.

[5] Grunz, J.-P. et al. "Twin Robotic X-Ray System for 3D Cone-Beam CT of the Wrist: An Evaluation of Image Quality and Radiation Dose." ,*American Journal of Roentgenology*, 1-6, 2019.

[6] Nellans, Kate W et al. “The epidemiology of distal radius fractures.” *Hand clinics* vol. 28,2 (2012): 113-25.

[7] Ensrud, Kristine E. " Epidemiology of Fracture Risk With Advancing Age", *The Journals of Gerontology: Series A*, 10,68 (2013)

[8] Feldkamp. L. A. et al. , "Practical cone-beam algorithm," *J. Opt. Soc. Am. A* 1, 612-619 (1984)

[9] Herbst, M. et al. " Misalignment compensation for ultra-high-resolution and fast CBCT acquisitions", *Proc. SPIE 10948, Medical Imaging 2019: Physics of Medical Imaging*, 109481M (2019)

[10] Wicklein, J. et al., "Image features for misalignment correction in medical flat-detector CT", *Medical Physics*. 39. 4918-31, (2012)

Fast statistical reconstruction algorithm for a CT scanner with flying focal spot

Robert Cierniak and Piotr Pluta

***Abstract*—This paper is related to the originally formulated 3D statistical model-based iterative reconstruction algorithm for tomography with flying focal spot. The conception proposed here is based on a continuous-to-continuous data model, and the reconstruction problem is formulated as a shift invariant system. The proposed approach resembles the FDK method, significantly improving the quality of the reconstructed images in comparison to that FBP referential algorithm, so allowing a reduction in the x-ray dose absorbed by a patient. The presented here approach avoids problems involved with nutating approaches usually applied in CT scanners with dual source x-ray tubes. Computer simulations have shown that the reconstruction method presented here outperforms standard FDK methods with regard to the image quality obtained and can be competitive in terms of time of calculation.**

***Index Terms*—dual source CT, flying focal spot, iterative reconstruction algorithm.**

I. Introduction

MANY years after the ivnention of the CT, new tomograph constructions are constantly being sought. In the beginning of 2004 medical spiral scanners were introduced that makes use of a flying focal spot (see e.g. [1]). The flying focal spot in the α-direction aims at inreasing the sampling density of the integral lines in reconstruction planes, and in the z-direction at increasing the sampling density just in z-direction. In practice, several approaches to reconstruction problem were formulated which allow for utilization of projections obtained from scanners with the flying focal spot. One of the most important among them is the adaptive multiple plane reconstruction (AMPR) algorithm, representing the class of nutating reconstruction method (see e.g. [2]). The AMPR conception is an extention of the advanced single slice rebinning (ASSR) method (see e.g. [1]). Despite the fact that the AMPR method is used commercially this rebinning approach has several serious drawbacks, mainly involved with some approximations and corrections needed to perform during realization of reconstruction procedure. Because of the linear nature of the operations carried out in the AMPR algorithm the ability to suppress noise in this approach is limited. It this context, an important circumstance is that recently the most significant problem in medical CT has been the development of image reconstruction methods which would enable the reduction of the impact of measurement noise on the quality of tomography images and thus decrease the dose of X-ray radiation absorbed by patients during examinations. Some of the most interesting research directions in this area are statistical reconstruction methods, especially those belonging to the MBIR (Model-Based Iterative Reconstruction) approach [3], where a probabilistic model of the measurement signals is taken into account. Unfortunately, those methods have some very serious drawbacks from the theoretical and practical point of view: for instance, if the image resolution is set to be $\mathrm{I} \times \mathrm{I}$ pixels, the calculation complexity of the problem is proportional to I^4, the statistical reconstruction procedure based on this methodology necessitates simultaneous calculations for all the voxels in the range of the reconstructed 3D image, the size of the forward model matrix $\mathbf{A}$ is huge, and this makes it often necessary to calculate them in every iteration of the reconstruction algorithm. In this case, the reconstruction problem is extremely ill-conditioned, and it is necessary to introduce an *a priori* term (often referred to in the literature as a regularization term) into the objective, and this leads to the use of the MAP model. The problems connected with the use of a methodology based on the D-D data model can be reduced by using a strategy of reconstructed image processing based on a continuous-to-continuous (C-C) data model. In previous papers we have shown how to formulate reconstruction problems consistent with the C-C mode and with the ML methodology for parallel scanner geometry [4], and finally for the spiral cone-beam scanner [5]. However, an approach to the reformulation of the reconstruction problem from parallel to real scanner geometries, called rebinning, was applied there. Much more convenient are FDK-type algorithms, and in this paper, we present a conception of the direct use of spiral cone-beam projections to a statistical reconstruction algorithm for dual source CT scanner.

Czestochowa University of Technology, Institute of Computational Intelligence, Armii Krajowej 36, Czestochowa, Poland, 42-200.

II. Reconstruction Algorithm

Our reconstruction method is based on the well-known maximum-likelihood (ML) estimation [7]. We propose here an optimization formula which is consistent with the C-C data model, in the following form:

$$\mu_{\min} = \arg\min_{\mu} \left(\sum_{i=1}^{\mathrm{I}} \sum_{j=1}^{\mathrm{I}} \left(\Delta_s^2 \sum_{\bar{i}=1}^{\mathrm{I}} \sum_{\bar{j}=1}^{\mathrm{I}} \mu^* \left(x_{\bar{i}}, y_{\bar{j}} \right) \cdot h_{\Delta i, \Delta j} - \tilde{\mu} \left(x_i, y_j \right) \right)^2 \right) \quad (1)$$

where I is a resolution of the reconstructed image, coefficients $h_{\Delta i, \Delta j}$ are precalculated according to the following relation:

$$h_{\Delta i, \Delta j} = \Delta_\alpha \sum_{\psi=0}^{\Psi-1} int \left(\Delta i \cos \psi \Delta_\alpha + \Delta j \sin \psi \Delta_\alpha \right), \quad (2)$$

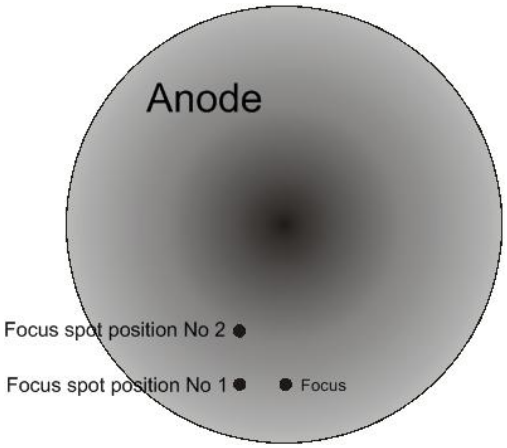


Fig. 1. The movement of focal spot with flying spot technique used in our experiments.

and $\tilde{\mu}(i,j)$ is an image obtained by way of a back-projection operation; Δ_s is a resolution of projections in a hypothetical parallel geometry of scanner; $int(\Delta s)$ is an interpolation function used in the back-projection operation; every projection is carried out after a rotation by Δ_α.

The presence of a shift-invariant system in the optimization problem (1) implies that this system is much better conditioned than the least squares problem present in the D-D approach [6].

In our considerations, we took into account the dual source CT scanner with two flying spot positions, as shown in Fig. 1. The focus spot position No. 1 is involved only with movement in α-direction, and the focus spot position No. 2 corresponds to the focal movement both in α-direction and z direction.

The conception presented here is a full 3D iterative reconstruction algorithm for spiral cone-beam scanner geometry. This algorithm is based on the one of the principal reconstruction methods devised for the cone-beam spiral scanner, i.e. the generalized FDK algorithm. The statistical reconstruction method proposed by us consists of two steps, namely: a back-projection operation described by the following relations:

$$\breve{\mu}^{'}(x,y) \approxeq \frac{1}{2}\int_0^{2\pi}\int_{-\beta_m}^{\beta_m} \frac{R_{fd}}{\sqrt{{R_{fd}}^2+z_k^2}} \cdot p^h\left(\beta, {\alpha'}^h, z_k\right) int_L(\Delta\beta)\, d\beta d{\alpha'}^h, \quad (3)$$

for the first focal spot position, where R_{fd} is the nominal focus-detector distance, ${\alpha'}^h = \alpha^h + \delta_{\alpha'}$; $\delta_{\alpha'}$ is the focal spot deviation in angular direction for the first focal spot position, and

$$\breve{\mu}^{''}(x,y) \approxeq \frac{1}{2}\int_0^{2\pi}\int_{-\beta_m}^{\beta_m} \frac{R''_{fd}}{\sqrt{{R''_{fd}}^2+z_k^2}} \cdot p^h\left(\beta, {\alpha''}^h, z_k\right) int_L(\Delta\beta)\, d\beta d{\alpha''}^h, \quad (4)$$

for the second focal spot position, where $R''_{fd} = R_{fd} + \delta_{R_f}$; δ_{R_f} is the focal spot deviation in radial direction, ${\alpha''}^h = \alpha^h + \delta_\alpha$; δ_α is the focal spot deviation in angular direction for the second focal spot position.

After the back-projection operation, an iterative reconstruction procedure according to formula (1) is performed using the image being a sum of two above described images, namely $\breve{\mu}^{'}(x,y)$ and $\breve{\mu}^{'}(x,y)$, i.e.:

$$\breve{\mu}(x,y) = \breve{\mu}^{'}(x,y) + \breve{\mu}^{''}(x,y). \quad (5)$$

Generally, it could be used any number of the focal spot positions, which could be moved both in z and in angular α directions.

It is worth noting that an interpolation function which is taken into account in relation (2) has the following form:

$$int_L(\Delta s) = \begin{cases} \frac{1}{\Delta_s}\left(1 - \frac{|\Delta s|}{\Delta_s}\right) & \text{for} \quad |\Delta s| \le \Delta_s \\ 0 & \text{for} \quad |\Delta_s| \ge \Delta_s \end{cases}, \quad (6)$$

where the optimal Δs can be determined using the following equation:

$$\Delta s = R_f \sin \Delta_\beta, \quad (7)$$

where R_f is the focus-isocenter distance.

There are some technical problems with the performing of the interpolation needed during the back-projection operation regarding the second focal spot position. A distribution of integral lines are no more equiangular for this position and the interpolation is performesd at the different resolution in every case. For details regarding method for determining parameters of geometry of scanner with the flying focal spot see e.g. [8] and [9].

Figure 1 depicts this algorithm after discretization and implementation of FFT which significantly accelerates the calculations.

III. Experimental results

In our experiments, we have used projections obtained from a Somatom Definition AS+ (helical mode) scanner with the following parameters: reference tube potential 120kVp and quality reference effective 200mAs, $R_{fd} = 1085.6mm$, $\mathrm{R}_f = 595\mathrm{mm}$, number of views per rotation $\Psi = 1152$, number of pixels in detector panel 736, detector dimensions were $1.09\mathrm{mm} \times 1.28\mathrm{mm}$. During the experiments, the size of the processed image was fixed at 512×512 pixels. A discrete representation of the matrix $h_{\Delta x,\Delta y}$ was established before the reconstruction process was started, and these coefficients were fixed (transformed into the frequency domain) for the whole iterative reconstruction procedure. The image obtained after the back-projection operations was then subjected to a process of reconstruction (optimization) using an iterative procedure. A specially prepared result of an FBP reconstruction algorithm was chosen as the starting point of this procedure (using projections obtained from the first focal spot position. It is worth noting that our reconstruction procedure was performed without any regularization regarding the objective function from (1). The iterative reconstruction procedure was implemented for a computer with 10 cores, i.e. with an Intel i9-7900X BOX/3800MHz processor (the iterative reconstruction

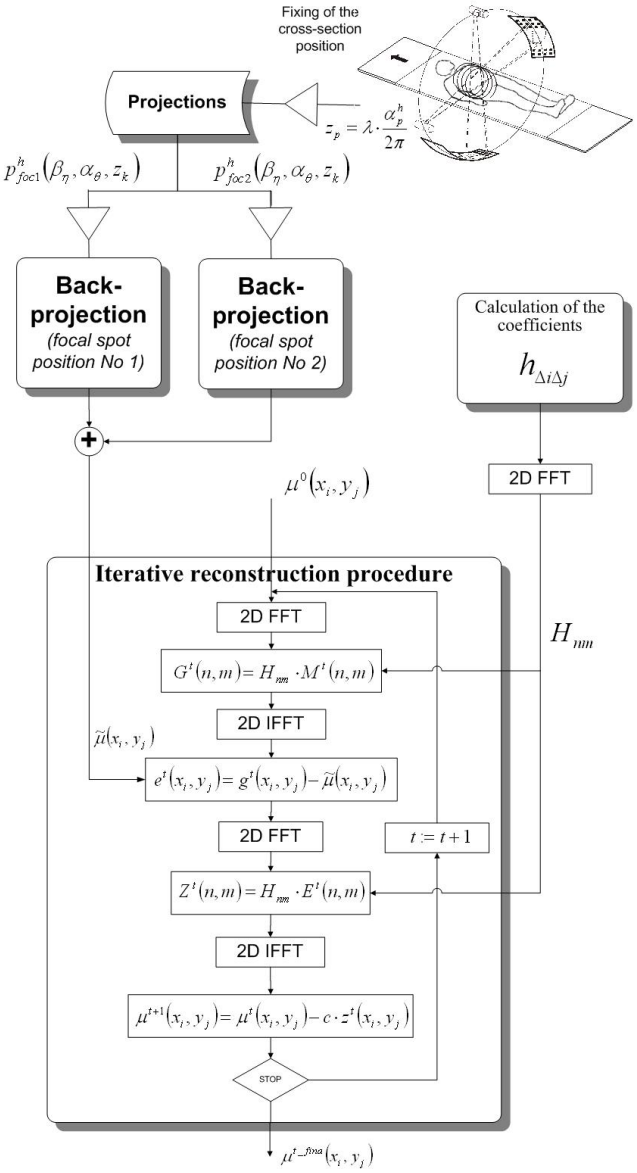


Fig. 2. Statistical reconstruction algorithm for spiral cone-beam scanner with flying focal spot.

procedure was implemented at assembler level), and using a GPU type nVidia Titan V. According to an assessment of the quality of the obtained images by a radiologist, 10000 iterations are enough to provide an acceptable image. The same results were achieved for both hardware implementations after about 9s, for the CPU and GPU implementations. One can compare the results obtained by assessing the views of the reconstructed images in Figure 3, where the full dose projections were used, and in Figure 4 where the quarter-dose projections were considered. In the both cases, Figures (A) depict reconstructed images obtained using the standard FDK algorithm (with linear interpolation function and Shepp-Logan kernel), Figures (B) and (C) present reconstructed images where the statistical approach presented in this paper was used: only measurements performed using the first focal spot position (B), all measurements were used (C).

IV. Conclusion

A statistical iterative reconstruction algorithm which can be used in practice for helical cone-beam scanners with flying focal spot has been shown above. We have conducted computer simulations, which proved that our reconstruction method is very fast, above all thanks to the use of FFT algorithms and efficient programming techniques, and it gives satisfactory results regarding the quality of the obtained images, and perspectivelly at a significantly reduced dose of x-rays absorbed by the patient. This algorithm is very easy to implement and open to use multisource technique, regardless the kind of the focal spot movement (both z and angle flying). If the image resolution is assumed to be $\mathrm{I} \times \mathrm{I}$ pixels, the complexity of the approach implemented here is proportional to $I^2 \log_2 I$. One can note that the iterative reconstruction procedure was performed without introducing any additional regularization term, using only an early stopping regularization strategy. It should be underlined that the price of the hardware used is relatively low (about 5000 USD in both cases).

Acknowledgment

The authors thank Dr. Cynthia McCoullough and the American Association of Physicists in Medicine for providing the Low-Dose CT Grand Challenge dataset.

This work was partly supported by The National Centre for Research and Development in Poland (Research Project POIR.01.01.01-00-0463/17).

References

[1] M. Kachelrieß, M. Knaup, C. Penßel, W. Kalender, *Flying focal spot (FFS) in cone-beam CT*, IEEE Transactions on Nuclear Science 53 (3), 1238-1247, 2006.

[2] Th. Flohr, K. Stierstofer, H. Bruder, J. Simon, A. Polacin, S. Schaller, *Image reconstruction and image quality evaluation for a 16-slice CT scanner*, Medical Physics 30 (5), 832-845, 2003.

[3] Y. Zhou, J.-B. Thibault, C.A. Bouman, J. Hsieh, K.D. Sauer, *Fast model-based x-ray CT reconstruction using spatially non-homogeneous ICD optimization*, IEEE Tran. Im. Proc. 20, 161–175, 2011.

[4] R. Cierniak, *An analytical iterative statistical algorithm for image reconstruction from projections*, Applied Mathematics and Computer Science 24, 7-17, 2014.

[5] R. Cierniak, *Analytical statistical reconstruction algorithm with the direct use of projections performed in spiral cone-beam scanners*,Proc. of the 5th International Meeting on Image Formation in X-Ray Computed Tomography, Salt Lake City, 293–296, 2018.

[6] R. Cierniak, A. Lorent, *Comparison of algebraic and analytical approaches to the formulation of the statistical model-based reconstruction problem for x-ray computed tomography*, Computerized Medical Imaging and Graphics 52, 19-27, 2016.

[7] C. Bouman, K. Sauer, *A unified approach to statistical tomography using coordinate descent optimization*, IEEE Tran. Image Processing 5, 480-492, 1996.

[8] T. G. Flohr, K. Stierstofer, S. Ulzhaimer, H. Bruder, A. N. Promak, C. H. McCollough, *Image reconstruction and image quality evaluation for a 64-slice CT scanner with z-flying focal spot*, Medical Physics 32 (8), 2536-2547, 2005.

[9] T. G. Flohr, H. Bruder, K. Stierstofer, M. Petersilka, B. Schmidt, C. H. McCollough, *Image reconstruction and image quality evaluation for a dual source CT scanner*, Medical Physics 35 (12), 5882-5897, 2008.

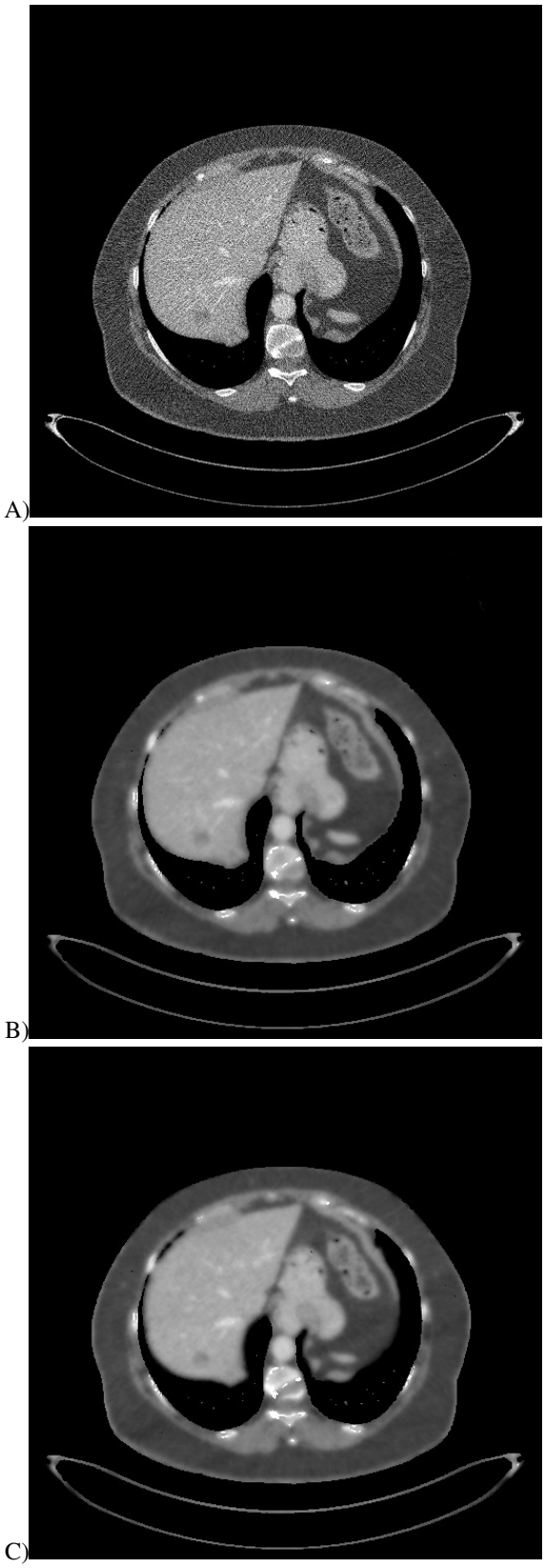


Fig. 3. View of the reconstructed image (a case with pathological changes in the liver) using full-dose projections with application of: the standard FDK algorithm (A); the statistical method presented in this paper (the first focal spot position) (B); the statistical method presented in this paper (both focal spot positions) (C).

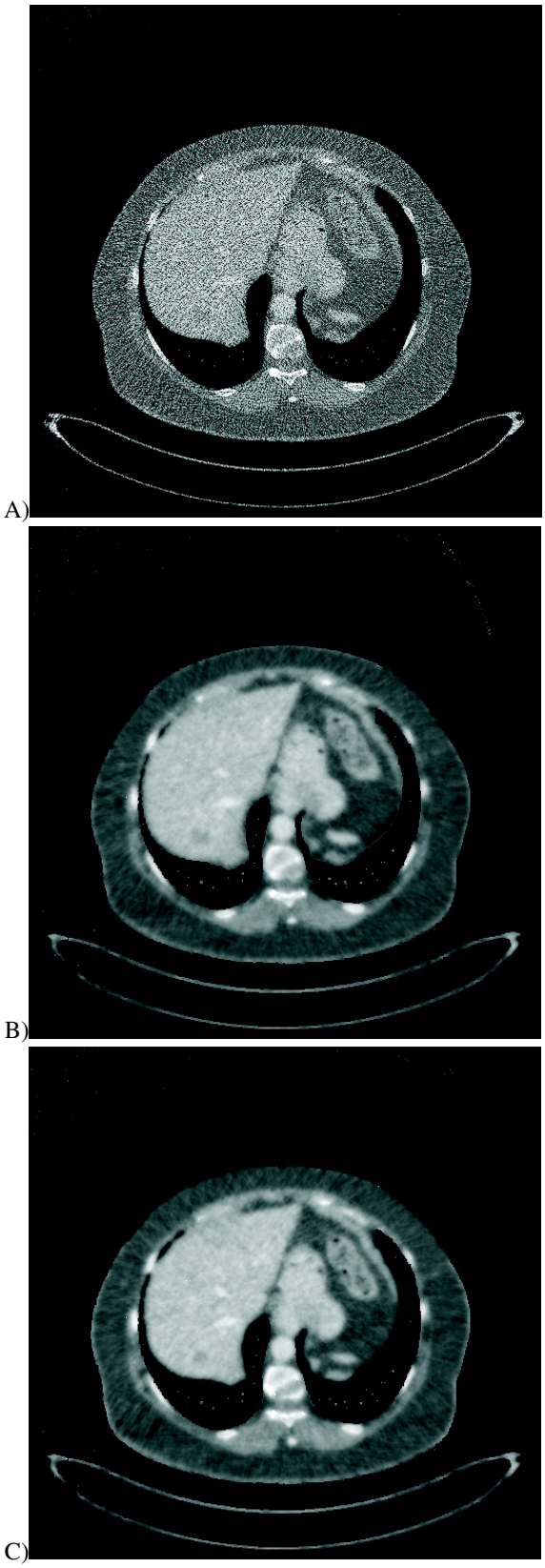


Fig. 4. View of the reconstructed image (a case with pathological changes in the liver) using quarter-dose projections with application of: the standard FDK algorithm (A); the statistical method presented in this paper (the first focal spot position) (B); the statistical method presented in this paper (both focal spot positions) (C).

Denoising of X-ray projections and computed tomography images using convolutional neural networks without clean data

Paolo Gnudi, Bernd Schweizer, Marc Kachelrieß, and Yannick Berker

***Abstract*—In X-ray projections and computed tomography (CT) imaging, image quality depends on the patient's exposure during the scan. Reducing the exposure reduces patients' health risks, but also reduces image quality due to higher noise in the image. A denoising technique preserving image features enables the acquisition of low-dose images without losing too much information. In this work, the performance of convolutional neural networks in denoising X-ray projections and CT images was evaluated. The peculiarity of the network is that it does not require the usage of clean data for training: instead, two or more independent noise realizations of each image are input during training. The network output was compared with that of a network trained using clean images, as well as images filtered using three conventional filters, namely Gaussian, Median and Bilateral. The results of this work indicate that the absence of clean images during training does not prevent the network from learning a good denoising model. In all examined cases, the performance of the deep-learning approaches provided better results than the selected filters. It is also shown that, whether clean images were used during training or not, each model has a denoising validity range related to the noise level of the training data.**

***Index Terms*—Convolutional neural networks, CT denoising, X-ray denoising, medical image denoising.**

I. Introduction

NOISE is present in virtually all types of acquired images, regardless of whether they are photographs or medical images, reducing the visibility of low-contrast objects. This loss of information can become a serious issue, especially in medical imaging. Some effort has been targeted at reducing noise by optimizing acquisition parameters: for example, using longer acquisition times usually increases the signal-to-noise ratio, but is not always practical in medical imaging due to issues such as patient comfort and the higher operating costs. In some medical imaging techniques, such as X-ray and CT imaging, a longer acquisition time also exposes the patient to a higher quantity of radiation, which should be avoided.

Various software techniques, such as image filtering, have been developed to reduce noise in medical imaging, so that noise removal has become an essential practice in medical imaging applications. Among all denoising techniques, the most recent ones are based on machine learning (ML), having shown very promising performance in image denoising, often superior to image filtering approaches [1, 2]. In this work, we compare the performance of different denoising approaches when applied to X-ray projections and CT images. These approaches comprise two ML techniques as well as three conventional denoising filters (Gaussian, Bilateral and Median). The first machine learning approach tested in this work follows the so-called "Noise2Clean" approach, requiring pairs of noisy and clean images during training in order to accomplish the denoising task. The second machine-learning approach ("Noise2Noise") was recently proposed by Lehtinen et al. [2]. The innovation of this approach is that it does not require clean images to learn how to restore noisy images.

The aim of the work is to evaluate which of these denoising tools is most suitable for X-ray applications. We investigated if, as expected, the ML approaches show superior performance than image filtering. The results of the noise-to-noise and noise-to-clean approaches were compared, and the impact of two loss functions (MSE and MAE) on the noise-to-noise results was investigated. Moreover, the range of validity of the learned denoising model, in dependence of the noise level of the training data, was investigated.

II. Theoretical Background

A. U-Net architecture

The machine-learning approaches used in this work are based on the U-Net, a particular kind of convolutional neural network (CNNs) that has been effective in noise removal [2, 3] although originally proposed for biomedical (semantic) image segmentation [4]. The architecture can be divided in an encoding and a decoding part.

In the encoding part, the information is encoded via spatial dimension reduction, while in the decoding part the information is decoded via spatial dimension increase. This structure produces levels where the inputs of the encoder part have the same dimension as the outputs of the decoder part.

One peculiarity of U-Nets are the skip connections between the encoding outputs and the decoding inputs of the same level. The concatenation between inputs and outputs gives U-

Manuscript received May 1, 2020.

P. Gnudi was, and B. Schweizer is with the Faculty of Engineering, RheinMain University of Applied Sciences, Am Brückweg 26, 65428 Rüsselsheim, Germany (e-mail: Bernd.Schweizer@hs-rm.de).

P. Gnudi and Y. Berker were, and M. Kachelrieß is with the Division of X-Ray Imaging and Computed Tomography, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany (e-mail: marc.kachelriess@dkfz.de).

Y. Berker is with the Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany (e-mail: yannick.berker@dkfz.de).

Nets the ability to maintain high-level as well as low-level features at the same time.

B. Noise2Clean and Noise2Noise approaches

A denoising model can be obtained by training a CNN regression model with a large number of training pairs $(\hat{x}; x)$ of noisy inputs $\hat{x}$ and clean images x. In such cases, the training inputs are the noisy images and the training targets are the clean images (Noise2Clean). The network is trained to output an image matching the clean input image as well as possible, as measured by a loss function, learning to denoise a noisy image by looking at both noisy and clean images. One of the problems in this approach is to find a sufficient number of clean targets to train the model.

Lehtinen et al. demonstrated that this denoising task could be accomplished also without clean images [2]. They trained a U-Net with noisy images as training instances and another noisy realization of the training instances as targets (Noise2Noise). In such a situation, the network is forced to learn to reproduce the second noisy realization, which is an impossible task because of the random behavior of the noise. Still, through a large number of training samples, the network learns to reproduce an average representation instead, which yields the desired denoising effect. In their work, the authors showed that, for noise removal in magnetic resonance imaging, the results produced with Noise2Noise are comparable with the ones produced with Noise2Clean.

The Noise2Noise technique is particularly suitable to train a denoising model for medical applications, because sometimes it can be challenging to obtain enough clean training targets. The Noise2Noise approach solves this problem not least because with three noisy realizations of the same image, six training pairs can be produced. In general, if N is the number of noisy realizations per image and I is the total number of available images, the number of possible training pairs is $I \cdot (N \cdot (N - 1))$.

III. Materials And Methods

For both X-ray and CT imaging data sets described in the following, we trained different Noise2Noise and Noise2Clean models by changing the loss function used during training and the targets in the training set, all the while using the same U-Net structure.

We used the peak signal-to-noise ratio (PSNR) as a metric to compare the output images to their clean references. On each test set, we also applied Gaussian, Bilateral and Median filters in order to compare their denoising capabilities with those of the ML approaches. For each filter, the best setting was determined image by image in order to ensure the best filter performance. To test the limits of the models, for both imaging techniques, we investigated the output noise level as a function of various input noise levels different from the noise level of the training data.

A. Simulated X-ray dataset

Clean images of the X-ray dataset were simulated projections, produced using the Extended Cardiac Torso (XCAT) simulation framework, a command line application capable of producing realistic digital human phantoms [5]. Noisy images were obtained from the simulated clean projections by computing two independent Poisson noise realizations. In total, we simulated 100 digital human phantoms with varying organ geometry and, for each phantom, projections at 63 different projection angles. In order to avoid different projections of the same phantom to be in the training set and the test set, the data were divided by phantom into 80 training and 20 test phantoms. This produced a training set composed by 10,080 images and a test set of 1,000 images. Table I shows the trained networks together with the adopted loss function for this dataset.

TABLE I
Networks Trained on X-ray Projections

Trained network	Loss function
Noise2Noise	MSE
Noise2Noise	MAE
Noise2Clean	MSE

Before being fed to the network, all noisy and clean slices were rescaled roughly into the [0, 1] range: clean slices were rescaled by dividing all pixels in an image by the maximum pixel value of the image itself. The noisy images were rescaled by dividing each pixel by the maximum of the respective clean image. This produced noisy images where some pixel values were greater than 1; negative pixel values were not present in this dataset due to the non-negativity of the Poisson distribution. For the construction of the training set to be used with the Noise2Noise approach, two different noisy realizations per clean slice were used. The baseline PSNR of this dataset was 29.91 dB.

B. Real CT dataset

Approximately clean CT slices were obtained from part of the DeepLesion dataset released by the National Institutes of Health (NIH) Clinical Center [6], which consists of 32,120 axial CT slices from 10,594 studies of 4,427 unique patients. The clean slices were rescaled by dividing all images by the highest pixel value of the entire dataset.

Noise was added after rescaling the clean images. Again, noisy images were artificially obtained starting from the clean slices. In CT images, noise is similar to Gaussian noise, because in back-projection of a Poisson-corrupted sinogram, all noise contributions are added over the entire image plane. We decided to add an approximation of this type of noise, namely, additive white Gaussian noise (AWGN) before feeding images to the network.

Two training and test sets at two different baseline PSNRs were created: one at 24.44 dB and another at 18.66 dB. Each training set consisted of 10,000 images, and each test set of 1,000 images. In addition, using the MSE loss function, we trained a Noise2Noise network for the training set at 18.66 dB. Table II summarizes the trained networks together with the adopted loss function for the CT dataset.

TABLE II
Networks Trained on CT Slices

Trained network	Noise level [dB]	Loss function
Noise2Noise	24.44	MSE
Noise2Noise	24.44	MAE
Noise2Clean	24.44	MSE
Noise2Noise	18.66	MSE

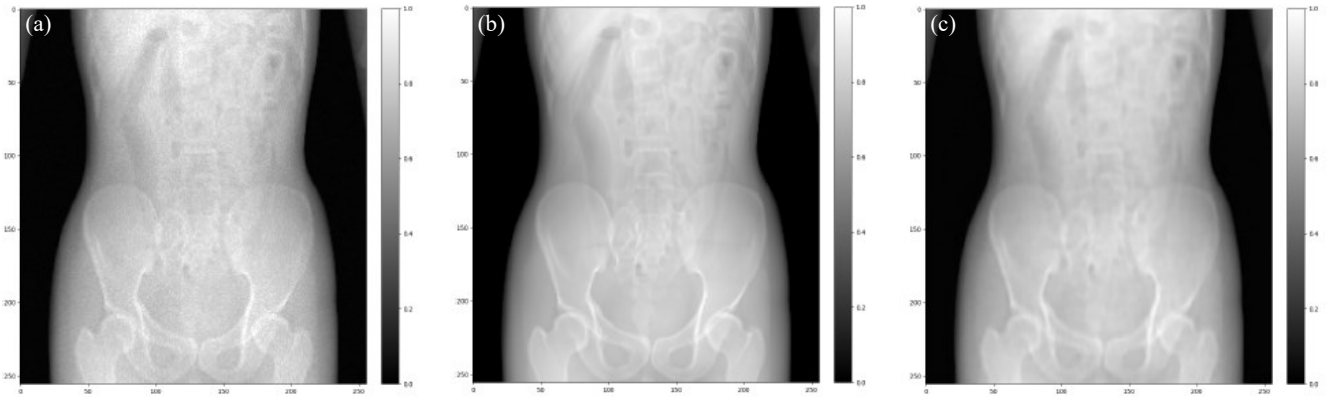


Fig. 1: Comparison between (a) noisy input (31.57 dB), (b) clean reference and (c) prediction of the Noise2Noise model trained with the MSE loss (38.36 dB).

C. *Network architecture*

The network used in this work is directly inspired by the one proposed previously [2], a U-Net with six levels. In the encoding part, except for the first layer, the dimensional reduction is performed by a series of 2D convolutions with 48 feature maps each and max pooling. In the decoding part, up-sampling is performed by an up-sampling layer, plus two consecutive 2D convolutions with 96 feature maps each. Each level of the encoding part is concatenated to the same level of the decoding part. Leaky ReLUs ($\alpha = 0.1$) were used as activation functions for all layers except for the output layer, where a linear activation function was adopted. The initial learning rate was set to 0.001 with a reduction to its half after two epochs in which the PSNR did not improve during testing.

IV. Results

A. *Denoised X-ray projections*

Table III shows the results in terms of PSNR obtained by applying different denoising tools on the X-ray projection test set. The experimental results show that the Noise2Noise network showed better performance when trained with the MSE loss instead of the MAE loss. Among all ML approaches the Noise2Clean showed the best performance. It is important to notice that independently of the approach and the adopted loss function, the ML methods are superior to all tested filters. Figure 1 shows the comparison between the noisy input, the clean reference and the prediction of the Noise2Noise model trained with the MSE loss.

The Noise2Noise and Noise2Clean trained networks were applied to different noisy realizations of the same input image in a PSNR range between 14.80 dB to 50.06 dB. Figure 2 shows the PSNR of the predictions by both models over the input PSNR, as well as the model gain (output – input PSNR).

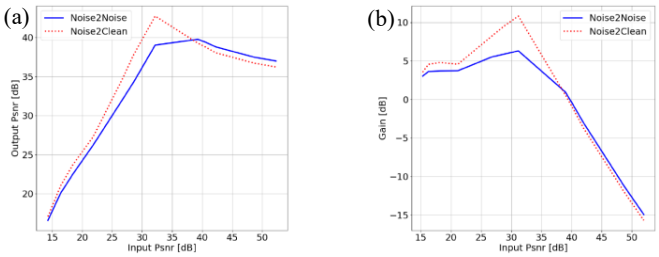


Fig. 2: Behavior of the Noise2Noise (solid blue) and Noise2Clean (dotted red) models trained with the MSE loss function for different noise levels of the same input image. (a) Output PSNR and (b) PSNR gain over input PSNR. Both models were trained on a baseline PSNR of 29.91 dB.

Both models show a similar behavior. The output PSNR increases until the input PSNR is around 35 dB, which is when it starts to decrease (Fig. 2a), indicating that the network's learned denoising model is valid in a high-to-moderate range of noise. When the input image PSNR is higher than 40 dB, both learned models start to deteriorate the image instead of restoring it. Figure 2b reveals a non-linear gain behavior in the region of very high-moderate noise level ([15, 32] dB). For an input PSNR in the range [15, 22] dB, the gain of both networks starts to saturate, suggesting that the learned model has the same denoising effect on very-high-level-noise images, regardless of the value of the input PSNR.

B. *Denoised CT slices*

Tables IV and V show the results of the various tools used to denoise the CT slices of the two test sets with the two different noise levels. Also in case of high-moderate noise level, the Noise2Noise model performs better compared to the three conventional filters. In case of moderate noise level, the mean predicted PSNR of the Noise2Noise model trained with the

TABLE III
Experimental Results on the X-ray Dataset (Baseline PSNR 29.91 dB)

Tool	Mean PSNR [dB]
Noise2Noise MAE loss	37.66
Noise2Noise MSE loss	39.34
Noise2Clean MSE loss	40.54
Gaussian filter	35.09
Bilateral filter	35.22
Median filter	35.71

TABLE IV
Experimental Results on the CT Dataset (Baseline PSNR 24.44 dB)

Tool	Mean PSNR [dB]
Noise2Noise MAE loss	41.80
Noise2Noise MSE loss	42.09
Noise2Clean MSE loss	40.63
Gaussian filter	34.25
Bilateral filter	35.09
Median filter	33.48

TABLE V
EXPERIMENTAL RESULTS ON THE CT DATASET (BASELINE PSNR 18.66 dB)

Tool	Mean PSNR [dB]
Noise2Noise MSE loss	30.01
Gaussian filter	26.94
Bilateral filter	27.22
Median filter	25.30

MAE loss is 41.80 dB, while the predicted PSNR of the model trained with the MSE loss is 42.09 dB. The performance of the Noise2Noise model, independently from the adopted loss function during training, is better than the performance of the Noise2Clean model.

Figure 3 shows the comparison between the noisy input, the clean reference and the prediction of the Noise2Noise model trained with the MSE loss. It can be noticed that the network was able to restore the image preserving the patient's anatomical structures without the introduction of artifacts in the predicted image. Figure 4 shows the behavior of the two Noise2Noise models trained with the MSE loss function at two different noise levels (24.44 dB and 18.66 dB) when applied to different noise realizations of the same clean image in a PSNR range between 3.61 dB and 46.51 dB.

Both models present a similar behavior. While the input PSNR is in the range [3, 30] dB (Fig. 4a), the output PSNR of both models increases, indicating a denoising effect. For input PSNRs lower than 16 dB, the model trained at 18.66 dB performs better than the model trained at 24.44 dB; the inverse is true for the input PSNR range [16, 30] dB. Around 30 dB the output PSNR of both models starts to saturate, leading to lower output than input PSNRs starting from 35 and 45 dB, respectively.

V. CONCLUSIONS

An ML approach is capable of denoising X-ray and CT images with superior performance compared to the selected conventional filters. In the case of the CT dataset, we found that the results of the Noise2Noise approach are even better than the ones of the more conventional Noise2Clean. Since we used an approximated noise model in CT images, a new training on CT slices with high noise is required to confirm the superiority of the Noise2Noise approach. The best performance of the Noise2Noise approach was reached using the MSE loss function. We have observed that the validity of each trained model is restricted to a range around the amount of noise in the training data. Results should be confirmed using additional metrics beyond PSNR.

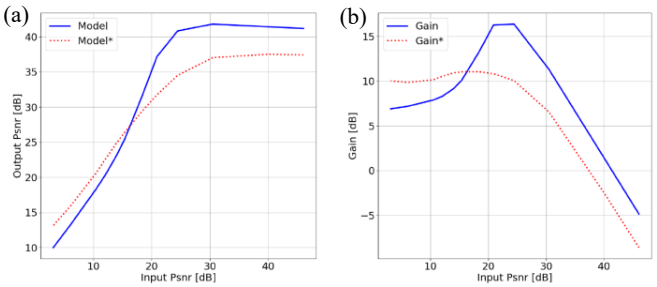


Fig. 4: Behavior of the Noise2Noise network trained at 24.44 dB (solid blue) and trained at 18.66 dB (dotted red) for different noise levels of the same input image. (a) Output PSNR and (b) PSNR gain over input PSNR.

More generally, the experimental results, especially for the CT dataset, suggest a possible application of ML methods in reducing patient exposure during image acquisition without loss of important information. The Noise2Noise method is particularly suitable to train denoising algorithms for all those imaging modalities which are inherently noisy (e.g. positron emission tomography, ultrasound) and where clean targets are not available. Since each scanner and reconstruction mode is characterized by their own noise, a different denoising model trained on images of each combination may be required.

REFERENCES

[1] A.S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI". Zeitschrift für Medizinische Physik, vol. 29, no. 2, pp. 102-127, 2019. https://doi.org/10.1016/j.zemedi.2018.11.002

[2] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. "Noise2Noise: Learning Image Restoration without Clean Data". arXiv:1803.04189v3, 2018. https://arxiv.org/abs/1803.04189v3

[3] M.G. Asante-Mensah and A. Cichocki, „Medical Image De-noising Using Deep Networks". IEEE International Conference on Data Mining Workshops, 2018. https://doi.org/10.1109/ICDMW.2018.00052

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation". Medical Image Computing and Computer-Assisted Intervention, 2015. https://doi.org/10.1007/978-3-319-24574-4_28

[5] P.W. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B.M. Tsui, "4D XCAT phantom for multimodality imaging research". Medical Physics, vol. 37, no. 9, pp. 4902-4915, Sep. 2010. https://doi.org/10.1118/1.3480985

[6] K. Yan, X. Wang, L. Lu, and R. M. Summers, "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning". Journal of Medical Imaging, vol. 5, no. 3, 036501, 2018. https://doi.org/10.1117/1.JMI.5.3.036501

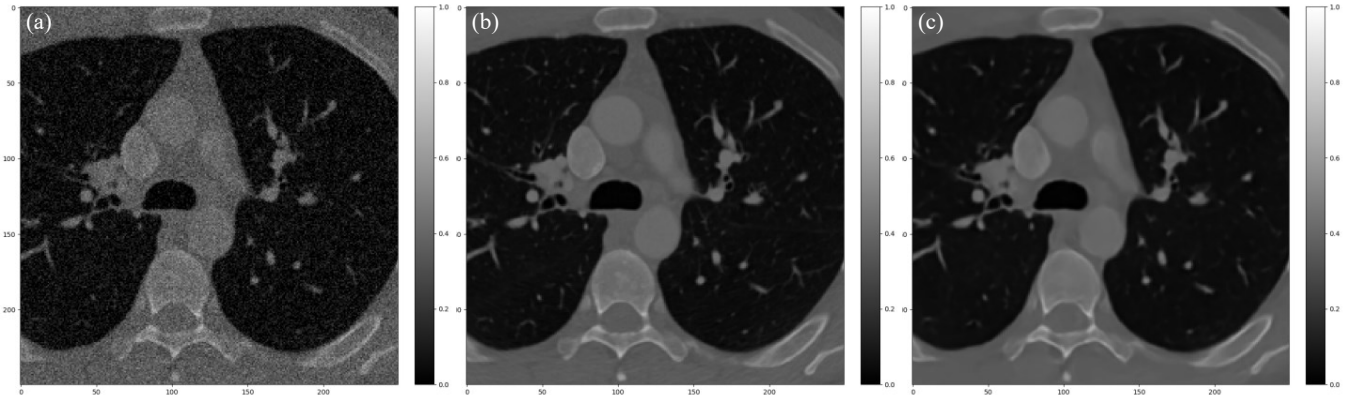


Fig. 3: Comparison between (a) noisy input (24.44 dB), (b) clean reference, and (c) prediction of the Noise2Noise model trained with the MSE loss (40.22 dB).

Projection-Domain Iteration to Estimate Unreliable Measurements

Gengsheng L. Zeng

***Abstract*—Due to the beam-hardening effect of the broad energy spectrum of the x-ray source in computed tomography (CT), the reconstructed images usually suffer from severe artifacts when metallic objects are being imaged. Metal artifact correction methods are usually sophisticated and not practical, especially in some non-medical applications, in which the linear attenuation coefficients are unknown. This paper suggests a simple and effective algorithm to estimate the unreliable measurements. The proposed algorithm is an iterative algorithm, in which the iteration is performed in the projection domain, while the objective function is set up in the image domain. The final image is reconstructed with the conventional filtered backprojection (FBP) algorithm. The feasibility of the proposed method is verified with airport bags that contain some unknown metals.**

***Index Terms*—CT artifacts, metal artifact reduction, beam-hardening effects, X-ray CT**

I. Introduction

In airport bag CT scans, almost every bag contains some unknown metallic objects. The x-ray tubes in the CT scanners have a broad energy spectrum. The linear attenuation coefficients of the metallic materials vary dramatically within the energy spectrum; this dramatic variation in the linear attenuation coefficients is not properly handled in the reconstruction algorithms. As a result, severe streaking and shadow artifacts appear in the reconstructed images.

Many attempts have been made to battle the metal artifacts. For example, only the projections not affected by the metals are used in an iterative algorithm that also has an edge-preserving prior [1]. Other popular methods are to replace metal affected projections with estimated values. These types of methods are also referred to as projection completion or inpainting [2]. The inpainting methods can use many non-linear image processing techniques such as opening, closing, and segmentation [3]. Inpainting can also be implemented in the Fourier domain, which is essentially lowpass filtering [4]. It is noticed that inpainting can lose spatial resolution; the boundary information from the FBP reconstruction (before inpainting) can be useful [5]. Segmentation in the inpainting methods can be avoided by using histogram deformation [6]. Some authors try to model the polychromatic energy spectrum into the algorithm [7]. Other inpainting methods are also reported in [8-10]. Park *et al* proposed a method of correcting metal artifacts due to beam hardening based on the observation that streaking artifacts arise mainly from the geometry of the boundaries of the metallic objects [11]. Some methods require the knowledge of materials involved and their parameters at different energy windows [12-14]. Iterative algorithms are popular in image reconstruction when metal objects are involved [15-17].

The total variation (TV) minimization method was used in [18] to assist inpainting. In [18], the TV norm was evaluated in the projection domain. In this paper, the TV norm is also used, but the TV norm is evaluated in the image domain instead. Even though the TV norm is evaluated in the image domain, our proposed iteration procedure is carried out in the projection domain. This is the unique feature of our proposed algorithm.

In airport luggage scanning, almost every bag contains metallic objects, and their linear attenuation coefficients are unknown. Therefore, metal artifacts in bag scans are more unpredictable and more difficult to eliminate than in clinical studies, where the metals are usually known. This paper suggests a projection-domain iterative algorithm to estimate the unreliable metal-affected projection measurements. After this step is done, the conventional FBP algorithm is used to reconstruct the final image. Realistic bag scan examples are used to verify the feasibility of the proposed method.

II. Methods

A. Motivation

The metal artifacts are the beam hardening effects, which are nonlinearly dependent on the metallic materials. These nonlinear effects introduce errors to the line-integral model of the measurements. The line-integral amplitudes are distorted when the integration lines pass through metals. Usually, the distorted line-integral value is smaller than the true value. The distortion is nonlinear and difficult to estimate, because the metallic materials in the objects are unknown.

Instead of developing a mathematical model for the beam hardening effects, this paper will take a different approach to estimate the distorted measurements by minimizing an objective function evaluated in the image domain.

A conventional objective function typically has a data-fidelity term and a Bayesian term. This paper proposes an objective function that does not have a data-fidelity term. The proposed objective function only contains one Bayesian term, which the TV normal of the FBP reconstruction. Our philosophy of selecting this objective function is as follows. The distortion of the line-integral measurements causes extra structures in the FBP reconstruction. By minimizing the TV-norm of the FBP

G. L. Zeng is with the Department of Engineering, Utah Valley University, Orem, UT 84058 USA and Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, UT 84108 USA (e-mail: larry.zeng@uvu.edu).

reconstruction, the extra structures (i.e., artifacts) can be reduced.

In the proposed algorithm, the iteration is only performed for the distorted measurements. The reliable measurements are kept to their original values unaltered.

B. Algorithm Development

Let f be the FBP reconstruction. The image f is represented in a two-dimensional (2D) array and $f_{i,j}$ is its pixel value at the ith row and jth column. The TV norm of f is defined as

$$T = \sum_{i,j} \sqrt{(f_{i,j} - f_{i,j+1})^2 + (f_{i,j} - f_{i+1,j})^2}. \quad (1)$$

When the quantity under the square root in (1) is positive, the partial derivative of T with respect to pixel (i, j) is readily calculated as [19]

$$U_{i,j} = \frac{\partial T}{\partial f_{i,j}} = \frac{(f_{i,j} - f_{i,j+1}) + (f_{i,j} - f_{i+1,j})}{\sqrt{(f_{i,j} - f_{i,j+1})^2 + (f_{i,j} - f_{i+1,j})^2}} + \frac{f_{i,j} - f_{i,j-1}}{\sqrt{(f_{i,j-1} - f_{i,j})^2 + (f_{i,j-1} - f_{i+1,j-1})^2}} + \frac{f_{i,j} - f_{i-1,j}}{\sqrt{(f_{i-1,j} - f_{i-1,j+1})^2 + (f_{i-1,j} - f_{i,j})^2}}. \quad (2)$$

When the quantity under the square root is zero, the quantity already reaches its minimum and the penalty function is no longer needed. In this situation, it is safe to set the derivative to zero. One easy way to handle this situation in an actual implementation of (2) is to add a very small positive constant ε, say, 10^{-8}, under the square roots in all denominators. Thus (2) becomes

$$U_{i,j} \approx \frac{(f_{i,j} - f_{i,j+1}) + (f_{i,j} - f_{i+1,j})}{\sqrt{(f_{i,j} - f_{i,j+1})^2 + (f_{i,j} - f_{i+1,j})^2 + \varepsilon}} + \frac{f_{i,j} - f_{i,j-1}}{\sqrt{(f_{i,j-1} - f_{i,j})^2 + (f_{i,j-1} - f_{i+1,j-1})^2 + \varepsilon}} + \frac{f_{i,j} - f_{i-1,j}}{\sqrt{(f_{i-1,j} - f_{i-1,j+1})^2 + (f_{i-1,j} - f_{i,j})^2 + \varepsilon}}. \quad (3)$$

Thus, an iterative gradient descent algorithm to minimize the objective function T defined in (1) is

$$f_{i,j}^{new} = f_{i,j}^{old} - \lambda U_{i,j}. \quad (4)$$

Algorithm (4) is not useful, because it will make f converge to a constant (i.e., a flat image with $U_{i,j} = 0$). We will now make some modifications to (4) so that it can be useful for our purposes.

The first modification to (4) is taking the Radon transform on both sides of (4), obtaining

$$p_{t,\theta}^{new} = p_{t,\theta}^{old} - \lambda \Re\{U_{i,j}\}, \quad (5)$$

where λ is the relaxation parameter chosen to be 0.01 in this paper, $\Re$ is the Radon transform operator, and $p_{t,\theta} = \Re\{f\}$ with (t, θ) being the measurement space coordinates. In fact, (5) is not useful like (4), only driving $U_{i,j}$ to 0.

The second modification is to multiply a mask function on both sides of (5). In order to find this mask function, we first use the FBP algorithm to reconstruct an initial image, which may contain lots of metal artifacts. Select a threshold value (for example, 1/3 of the maximum image value), and use this value to segment the FBP reconstruction to create a metal object image f_{metal}. The mask, denoted as $mask_{metal}$, is a function in the projection-domain and is based on the Radon transform of f_{metal}. The mask function $mask_{metal}$ is 1 at the location that the Radon transform of f_{metal} is positive and is 0 at the location that the Radon transform of f_{metal} is 0. The mask function restricts (5) only on the region when $mask_{metal}$ is 1. Thus, (5) is further modified as

$$p_{t,\theta(metal)}^{new} = p_{t,\theta(metal)}^{old} - \lambda \times mask_{metal} \times \Re\{U_{i,j}\}. \quad (6)$$

Algorithm (6) only updates the line-integral measurements that the projection rays $p_{t,\theta(metal)}$ touch the metal objects.

Algorithm (6) is still not satisfactory, because it will smooth out all metallic objects while minimizing the TV norm (1). In order to keep the metallic objects in the image, the image-domain masking function f_{metal} can be used to hide the metals in TV gradient image $U_{i,j}$, that is, replacing $U_{i,j}$ by $(U_{i,j})\times(1- f_{metal})$. At a pixel with metal, $(1- f_{metal}) = 0$. At a pixel without metal, $(1- f_{metal}) = 1$. The final algorithm proposed in the paper is, therefore, expressed as

$$p_{t,\theta(metal)}^{new} = p_{t,\theta(metal)}^{old} - \lambda \times mask_{metal} \times \Re\{U_{i,j} \times (1 - f_{metal})\}. \quad (7)$$

The proposed algorithm can be implemented in the following steps:

i) Obtain a raw FBP reconstruction.
ii) Use a threshold to segment a metal image, f_{metal}, from the raw BPF reconstruction.
iii) Set the raw FBP reconstruction as the initial image and perform iterative updates for the unreliable measurements $p_{t,\theta(metal)}$ according to Algorithm (7).
iv) Obtain the final FBP reconstruction with the measurements, the unreliable portions of them have been revised by step iii.

C. Airport Bag Experiments

The original projections of airport bags were acquired with an Imatron C300 clinical CT scanner, which was a fifth generation CT scanner based on a scanning electron beam X-ray source manufactured in the late 1990s. The original projection data was rebinned and downsized for our reconstruction in this paper. The number of views was 180 over 180°. The number of channels (i.e., the detection bin at each view) was 597. The projections used the parallel-beam imaging geometry. The reconstructed image size was 420 × 420.

III. RESULTS

Five airport-bag examples are presented here to illustrate the feasibility of the proposed algorithm. For each example, the following 5 images are shown: 1) the raw FBP reconstruction, 2) the segmented metal image f_{metal}, 3) the projection-domain mask $mask_{metal}$, 4) the final FBP reconstruction using the restored projections, and 5) the reconstructed image using the image-domain TV iterative algorithm. The image-domain TV iterative algorithm is similar to (5), by adding a data fidelity term. The relaxation parameter for the data fidelity term was 0.001 and the relaxation parameter for the TV constraint term was 0.0005. The number of iterations was 400 for both the proposed projection-domain iteration algorithm and for the image-domain iterative TV algorithm. The image-domain TV

iterative reconstructions are presented here for the comparison purposes.

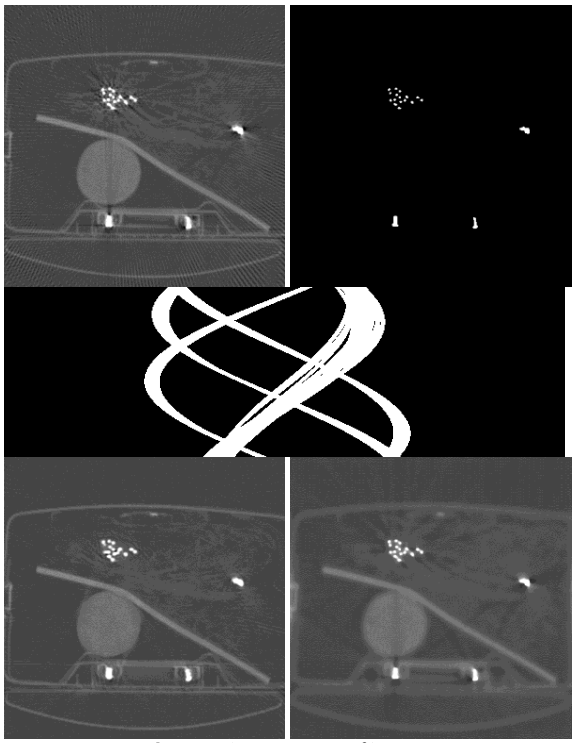


Figure 1. Airport bag #1. (Upper row, Left) raw FBP reconstruction, (Upper row Right) metal segmentation, (Middle row) projection mask, (Lower row, Left) final proposed FBP reconstruction, and (Lower, Right) image-domain TV reconstruction.

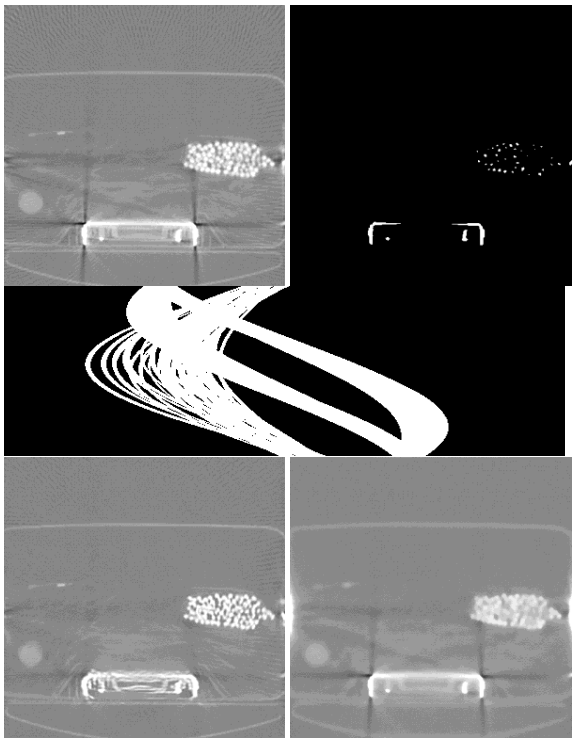


Figure 2. Airport bag #2. (Upper row, Left) raw FBP reconstruction, (Upper row Right) metal segmentation, (Middle row) projection mask, (Lower row, Left) final proposed FBP reconstruction, and (Lower, Right) image-domain TV reconstruction.

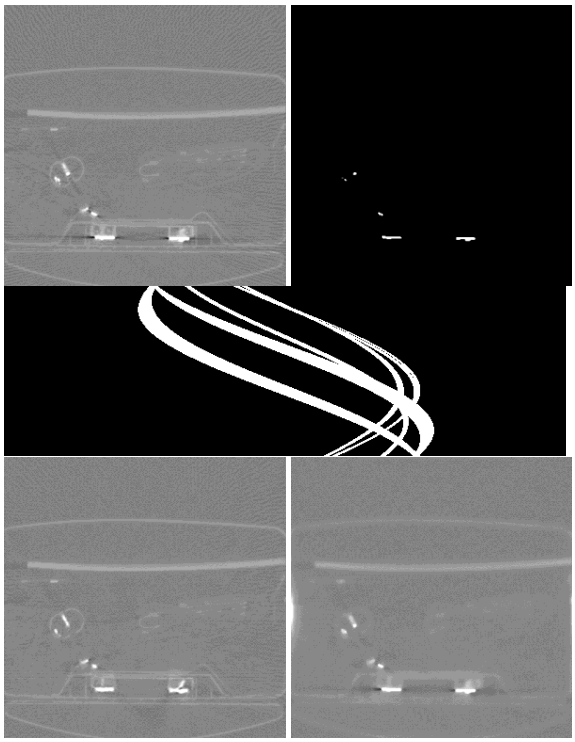


Figure 3. Airport bag #3. (Upper row, Left) raw FBP reconstruction, (Upper row Right) metal segmentation, (Middle row) projection mask, (Lower row, Left) final proposed FBP reconstruction, and (Lower, Right) image-domain TV reconstruction.

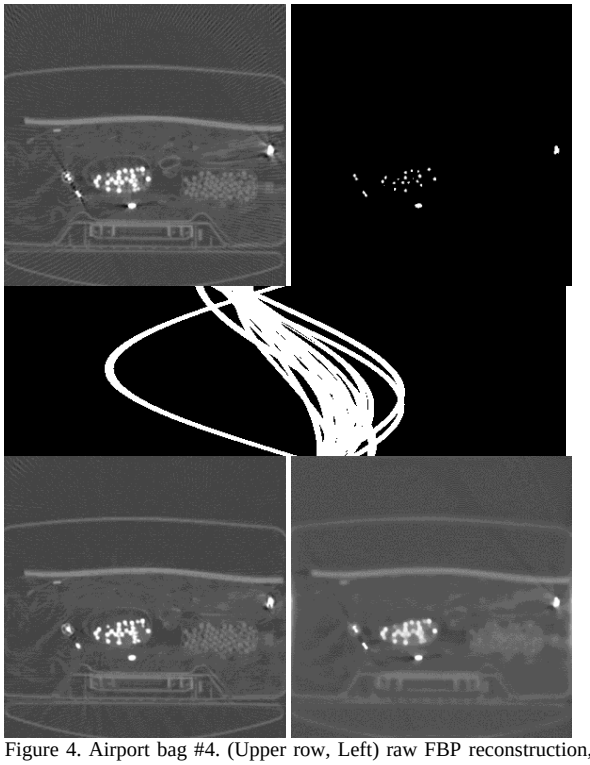


Figure 4. Airport bag #4. (Upper row, Left) raw FBP reconstruction, (Upper row Right) metal segmentation, (Middle row) projection mask, (Lower row, Left) final proposed FBP reconstruction, and (Lower, Right) image-domain TV reconstruction.

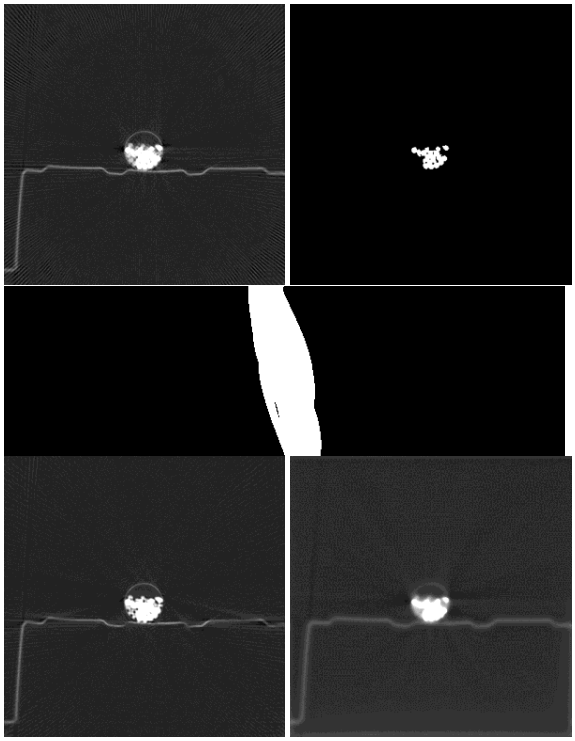


Figure 5. Airport bag #5. (Upper row, Left) raw FBP reconstruction, (Upper row Right) metal segmentation, (Middle row) projection mask, (Lower row, Left) final proposed FBP reconstruction, and (Lower, Right) image-domain TV reconstruction.

The reconstructions are displayed from the minimum image value to 0.3 times the maximum image value. The results from these 5 examples are shown in Figs. (1)-(5), respectively.

The metal artifacts are the dark shadows in these examples, and they are effectively reduced by the proposed algorithm. However, the conventional FBP and TV algorithms are not effective in removing those artifacts.

IV. Conclusion

The proposed projection-domain iterative algorithm minimizes the image-domain TV norm of the FBP reconstruction. This algorithm does not use any models for the unreliable projections. Therefore, this algorithm can be applied to many applications other than removing beam-hardening artifacts. It can be applied to any imaging geometries. It can find applications, where the detector has some bad detection cells, or some detection cells are missing. Of course, no algorithm is perfect. This algorithm may fail, for example, when there are too many unreliable projections.

Acknowledgment

The airport bag data was provided by the U.S. Department of Homeland Security, Science and Technology Directorate, under Task Order Number HSHQDC-12-J-00056. In this paper, we downgrade the spatial resolution of the original data on purpose. The views and conclusions are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

References

[1] X. Zhang, J. Wang, and L. Xing, "Metal artifact reduction in x-ray computed tomography (CT) by constrained optimization," *Med. Phys*., vol. 38, pp. 701-711, 2011.

[2] R. M. S. Joemai, P. W. de Bruin, W. J. H. Veldkamp, and J. Geleijns, "Metal artifact reduction for CT: Development, implementation, and clinical comparison of a generic and a scanner-specific technique," *Med. Phys*., vol. 39, pp. 1125-1132, 2012.

[3] S. Karimi, P. Cosman, C. Wald, and H. Martz, "Segmentation of artifacts and anatomy in CT metal artifact reduction," *Med. Phys*., vol. 39, pp. 5857-5868, 2012.

[4] B. Kratz, I. Weyers, and T. M. Buzug, "A fully 3D approach for metal artifact reduction in computed tomography," *Med. Phys*., vol. 39, pp. 7042-7054, 2012.

[5] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Frequency split metal artifact reduction (FSMAR) in computed tomography," *Med. Phys*., vol. 39, pp. 1904-1916, 2012.

[6] S. Schüller, S. Sawall, K. Stannigel, M. Hülsbusch, J. Ulrici, E. Hell, and M. Kachelrieß, "Segmentation-free empirical beam hardening correction for CT," *Med. Phys*., vol. 42, pp. 795-803, 2015.

[7] K. Van Slambrouck and J. Nuyts, "Metal artifact reduction in computed tomography using local models in an image block iterative scheme," *Med. Phys*., vol. 39, pp. 7080-7093, 2012.

[8] M. Abdoli, M. R. Ay, A. Ahmadian, R. A. Dierckx, and H. Zaidi, "Reduction of dental filling metallic artifacts in CT-based attenuation correction of PET data using weighted virtual sinograms optimized by a genetic algorithm," *Med. Phys*., vol. 37, pp. 6166–6177, 2010.

[9] M. Bazalova, L. Beaulieu, S. Palefsky, and F. Verhaegen, "Correction of CT artifacts and its influence on Monte Carlo dose calculations," *Med. Phys*., vol. 34, pp. 2119–2132, 2007.

[10] J. C. Roeske, C. Lund, C. A. Pelizzari, X. Pan, and A. J. Mundt, "Reduction of computed tomography metal artifacts due to the Fletcher-Suit applicator in gynecology patients receiving intracavitary brachytherapy," Brachytherapy, vol. 2, pp. 207–214, 2003.

[11] H. S. Park, D. Hwang, and J. K. Seo, "Metal artifact reduction for polychromatic X-ray CT based on a beam-hardening corrector," IEEE Trans. Med. Imag., vol. 35, pp. 480-487, 2016.

[12] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography*," Phys. Med. Biol*., vol. 21, p. 733, 1976.

[13] L. Lehmannet et al., "Generalized image combinations in dual kVp digital radiography," *Med. Phys*., vol. 8, pp. 659–667, 1981.

[14] L.Yu, S. Leng, and C. H. McCollough, "Dual-energy CT-based monochromatic imaging," *Am. J. Roentgenol*., vol. 199, pp. S9–S15, 2012.

[15] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "An iterative maximum-likelihood polychromatic algorithm for CT," *IEEE Trans. Med. Imag*., vol. 20, pp. 999–1008, 2001.

[16] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag*., vol. 21, no. 2, pp. 89–99, 2002.

[17] J. F. Williamson *et al*., "Prospects for quantitative computed tomography imaging in the presence of foreign metal bodies using statistical image reconstruction," *Med. Phys*., vol. 29, pp. 2404–2418, 2002.

[18] X. Duan, L. Zhang, Y. Xiao, J. Cheng, Z. Chen, and Y. Xing, "Metal artifact reduction in CT images by sinogram TV inpainting," in IEEE Nucl. Sci. Symp. Conf. Rec., pp. 4175–4177, 2008.

[19] V. Y. Panin, G. L. Zeng and G. T. Gullberg, "Total variation regulated EM algorithm," *IEEE Trans. Nucl. Sci*., vol. 46, pp. 2202-2210, 1999.

First Commercial Dual-energy MeV CT for Container Imaging: Experimental Results

Wei Fang, Liang Li, Zhiqiang Chen, Chunguang Zong

Abstract*–The MeV X-ray radiography technology has been widely used for cargo/container imaging in customs all over the world. The major limitation for this technology is the overlapping problem. When complicated materials exist along the X-ray beam, their projections are overlapped on the radiography image, making it hard for identification. Computed tomography (CT) technology can generate the cross-section image of an object with the projection data measured from different angle of views by applying certain reconstruction algorithm. So, it is of great significance and interest to try to build a CT system for cargo imaging. The main difficulties for building such a system include the manufacturing of the giant CT gantry and the rotation system of the bulky accelerators and detectors. In this paper, we report our efforts for building world's first dual-energy MeV CT for cargo imaging. We will introduce the system geometric design, the accelerators and detectors and the reconstruction and material decomposition algorithms used in the system. Experimental reconstruction and material decomposition results are also presented. These results show the advantageous performance of MeV CT system for more detailed cargo-imaging.

I. Introduction

HUNDREDS of millions of cargo transactions are performed annually in airports, stations, and harbors all over the world. The size of these cargos is so big that their width and height can be up to a few meters. The contents of these cargos are very complicated and the density of a full-loaded cargo can be up to a few thousand kilograms per cubic meter. The cargos and containers can be illegally used for concealing illicit materials, such as explosives, drugs, nuclear materials, smuggled goods and even stowaways. For the protection of homeland security, it is necessary to perform routine security screenings on these cargos. At present, high-energy X-ray radiography with energy up to 9 MeV is the dominant technique for this kind of inspection [1-3]. A few well-respected companies have presented commercially available products [4-8]. However, X-ray radiography technique has a big limitation and that is the overlapping problem [6]. X-ray radiography image is a projection of the 3D object, the structures along the X-ray lines are overlapped during the projection process. One possible way to alleviate this problem is to take a radiography image from a second view, such as a view perpendicular to the first view. While the overlapping problem still exists in the second image. To thoroughly solve this problem, computed tomography (CT) technique is needed. CT imaging can provide the attenuation coefficients map of a cross section of the cargo without material overlapping. The MeV CT for large cargo imaging is of big interest but has not been feasible for so many years due to a number of constraints, including the large size of the required CT gantry and the heavy weight of the MeV X-ray accelerator, which need to be rotated around the scanned object [9, 10]. In this paper, we introduce the world's first MeV CT cargo imaging system after years of hard work. The system consists of two accelerators with fast 6/9 MeV switching spectra, two arc detector arrays and a rotating system. GPU-accelerated reconstruction and dual-energy material decomposition algorithms are developed for this system.

Manuscript received January 16, 2020. This work was supported in part by National Key R&D Program of China (No. 2017YFC0109103) and NSFC 11775124.

Wei Fang, Liang Li and Zhiqiang Chen are with the Department of Engineering Physics, Tsinghua University, Beijing, 100084, China & Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education, 100084, China. Chunguang Zong is with Nuctech Company Limited, Beijing, China.

Corresponding author: lliang@tsinghua.edu.cn

II. Methods

A. The System Geometric Design

Fig. 1 shows the schematic drawing of the MeV cargo CT imaging system. Two accelerators and corresponding two arc detector arrays are fixed on a big gantry. The two accelerators rotate around the axis for 114 degrees at the same time. The angle between the two accelerators is also 114 degrees, which means after the rotation the ending point of the Accelerator I is the beginning point of the accelerator II. Thus, the whole scanning covers overall 228 degrees, which equals the sum of 180 degrees and the fan angle (48 degrees), making a sufficient condition for image reconstruction. The diameter for the gantry of the MeV cargo CT system is 12.6m and the diameter of the field of view is 5.12 m, which is big enough for truck loaded with cargo or container passing by.

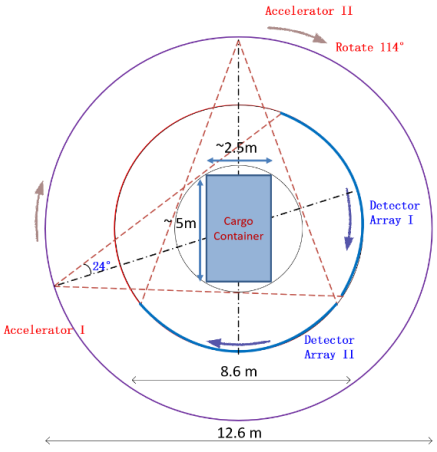


Fig. 1. The schematic drawing of the CT system.

B. *The Practical System*

Fig. 2 presents the whole image of the inspection system. The truck loaded with container is drugged through the CT gantry by the transport system and get scanned. The inspection system provides the radiography image from the horizontal and vertical view. If something suspicious is found in the radiography image, the inspection system can perform CT scanning to the specific location to check it out.

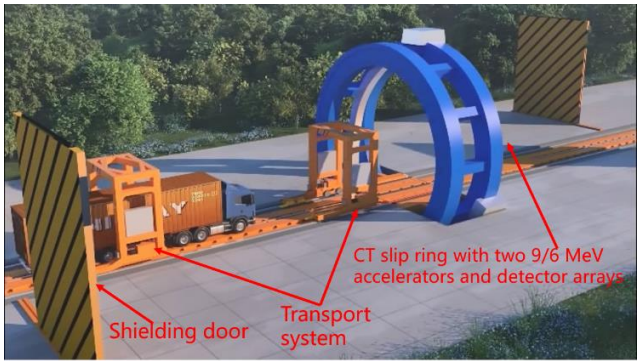


Fig. 2. The Cargo/Vehicle CT Inspection System.

Fig. 3 shows the practical CT system. The huge blue ring is the CT gantry. The CT gantry is a whole circle that can rotate around the scanned cargo. The bottom part is hidden under the yellow iron platform. Two linear accelerators are respectively fixed at the left and top of the gantry. Two arc detector arrays are fixed at the opposite of the corresponding accelerators. The detector arrays are composed of energy-integrating detectors that are regularly used in cargo X-ray radiography. The yellow iron platform can move back and forth to get different slice of the cargo scanned.

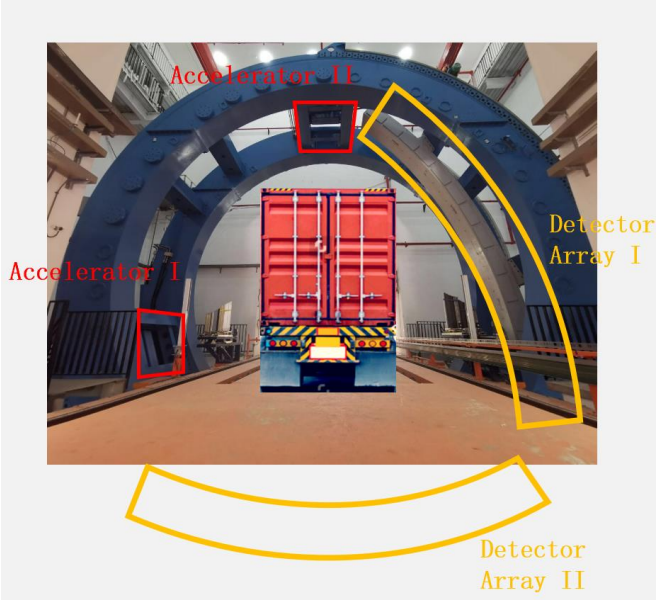


Fig. 3. The actual CT gantry.

The linear accelerators can generate very fast switching X-ray beam of 6MeV and 9MeV as the source for low and high energy imaging. The spectrums of the accelerator including the simulated and estimated ones are shown in Fig. 4.

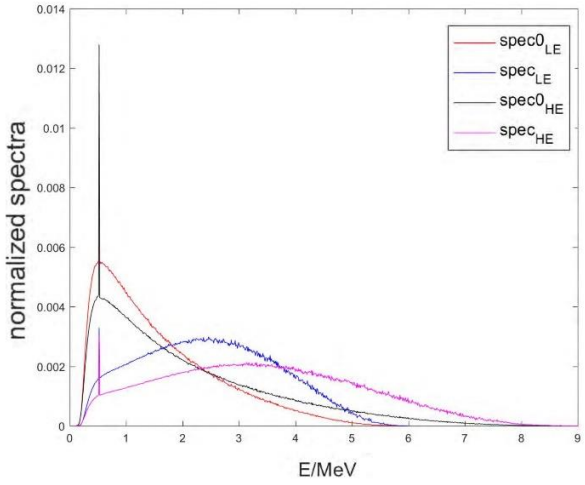


Fig. 4. Normalized spectra used in MeV CT for cargo imaging. ''$spec0_{LE}$'' and ''$spec0_{HE}$'' denote the simulated 6 and 9 MeV spectra by Monte Carlo method. ''$spec_{LE}$'' and ''$spec_{HE}$'' denote the estimated 6 and 9 MeV spectra by experimental measurements, respectively.

C. *The Test Phantom*

Fig. 5 shows the phantom we used for evaluating the spatial and density resolution. Materials including graphite, Mg-Al alloy, resin and Teflon with different densities are used for evaluating the density resolution. Six iron line pairs with different intervals are used for spatial resolution evaluation. In Fig. 6, the test module is put together with other things like fire extinguisher, steel pipe to simulate the real situation in container.

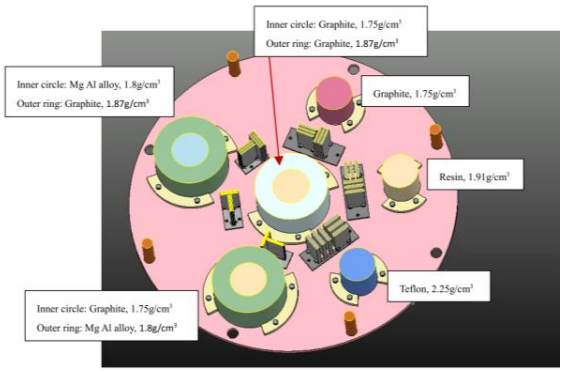


Fig. 5. The test phantom we used in the experiment.

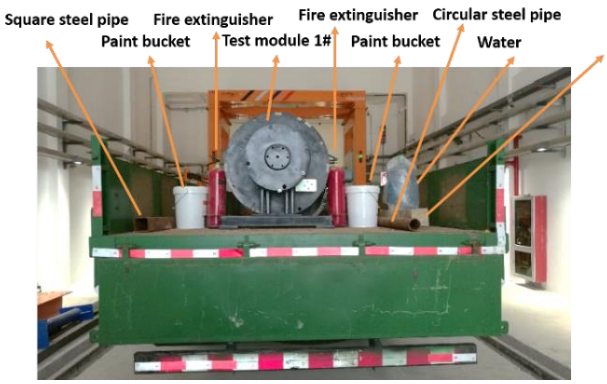


Fig. 6. Different objects are placed on the truck for imaging.

D. The Reconstruction

The size of the field of view is very big, whose diameter is up to 5.12m. In order to ensure the spatial resolution, a lot of projection data should be measured. The algorithm we used is FBP method, for its high efficiency and low computational burden. Since the X-ray source's rotation didn't cover the whole 360 degrees, the projection data we acquired is short-scan fan-beam data. Parker smooth windowing function is applied to solve the projection data redundancy problem [11]. GPU acceleration is used for the reconstruction. Some specific parameters are shown below in Table I.

TABLE I
SOME PARAMETERS OF THE RECONSTRUCTION

Parameters	Value
Number of views	5701
Number of detectors	3392
Size of the reconstructed image	[2048, 2048]
The GPU we used	Titan Xp
Time consumed on reconstruction	9.28s

E. The Material Decomposition

Based on the consideration of algorithm's efficiency and robustness, we choose classic post-processing approach as the material decomposition method [12, 13], which can be illustrated as:

$$\begin{pmatrix}\mu_{\mathrm{H}}\\ \mu_{\mathrm{L}}\end{pmatrix}=\begin{pmatrix}\mu_{1\mathrm{H}}(x,y) & \mu_{2\mathrm{H}}(x,y)\\ \mu_{1\mathrm{L}}(x,y) & \mu_{2\mathrm{L}}(x,y)\end{pmatrix}\begin{pmatrix}b_1\\ b_2\end{pmatrix}, \quad (1)$$

where $\mu_{\mathrm{H/L}}$ denotes the attenuation coefficients under the low and high energy spectrum. The notation $\mu_{\mathrm{iH/L}}$ denotes the attenuation coefficients of basis material $i(i=1,2)$ measured at low and high energy spectrum. The basis materials we chose for material decomposition are carbon (C) and iron (Fe). The attenuation coefficients of them are experimentally measured in the practical system using the pure material. However, the spectrum of the x-ray source of different directions of the fan-beam is not uniform. This results into the non-uniformity of the attenuation coefficients of the basis materials at different locations of the field of view. The initial strategy we applied here for solving this problem is to get the attenuation coefficients of the basis materials at different locations experimentally measured. And for each pixel, the basis material attenuation coefficients used for decomposition are specified, which is the nearest measured ones to that pixel. That's why $\mu_{\mathrm{iH/L}}$ is also the function of pixel location (x,y) in formula (1). Fig. 7 shows the attenuation coefficients map of the basis materials at low and high energy spectra.

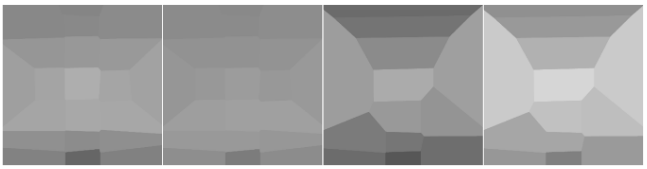


Fig. 7. The attenuation coefficients map of basis materials (C and Fe) we used at different locations of the imaging field. From left to right: C at low energy spectrum; C at high energy spectrum; Fe at low energy spectrum; Fe at high energy spectrum.

The direct matrix inversion of equation (1) gives the decomposition coefficients of two materials:

$$\begin{pmatrix}b_1\\ b_2\end{pmatrix}=\frac{1}{\mu_{1\mathrm{H}}(x,y)\mu_{2\mathrm{L}}-\mu_{2\mathrm{H}}\mu_{1\mathrm{L}}}\begin{pmatrix}\mu_{2\mathrm{L}} & -\mu_{2\mathrm{H}}\\ -\mu_{1\mathrm{L}} & \mu_{1\mathrm{H}}\end{pmatrix}\begin{pmatrix}\mu_{\mathrm{H}}\\ \mu_{\mathrm{L}}\end{pmatrix}. \quad (2)$$

After we get the decomposition coefficients, the electron density ρ_e and equivalent atomic number Z_{eff} can be calculated by:

$$\rho_e = 2(b_1\rho_1\frac{Z_1}{A_1}+b_2\rho_2\frac{Z_2}{A_2}), \quad (3)$$

$$Z_{eff}=\left[b_1\rho_1\frac{(Z_1)^2}{A_1}+b_2\rho_2\frac{(Z_2)^2}{A_2}\right]\Big/\left(b_1\rho_1\frac{Z_1}{A_1}+b_2\rho_2\frac{Z_2}{A_2}\right), \quad (4)$$

where ρ_i is the mass density of the basis material i, and Z_i, A_i respectively represent the atomic number and mass number of basis materials i. More detailed derivation of formula (3) and (4) is illustrated in paper [14].

III. RESULTS

Fig. 8 shows the reconstruction and decomposition results of the truck in Fig. 6. Image (a) and (b) are respectively low and high energy reconstruction image. The test phantom, fire extinguisher, paint bucket and steel pipe can be clearly seen in the reconstructed image. The image quality of high energy reconstruction image is obviously better than the low energy reconstruction image due to less beam hardening artefacts.

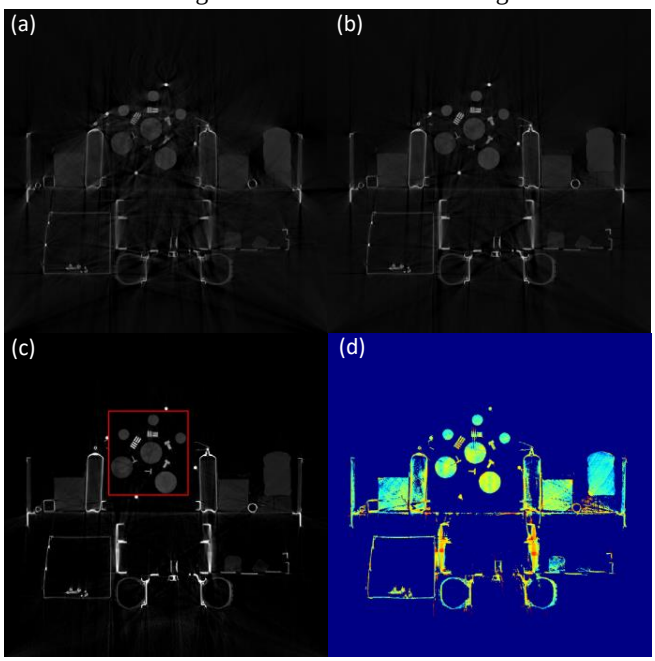


Fig. 8. Experimental reconstruction and decomposition results of a truck slice: (a) low energy reconstruction image. (b) high energy reconstruction image. (c) electron density image. (d) equivalent atomic number image.

Image (c) and (d) are respectively the calculated electron density image and equivalent atomic number image. The error

of ρ_e is acceptable. But the error of Z_{eff} is not very ideal. The decomposition results may be affected by different artefacts, including scattering, beam hardening and some serious photon starvation streak artefacts caused by high Z material in the reconstruction image.

The enlarged image of the red rectangle in (c) is shown in Fig. 9. We can see that the line pair with the parameter of 1.42 lp/cm can be clearly seen. Judging from the cylinder near L1 line pair, a density resolution of 2.7% is achieved.

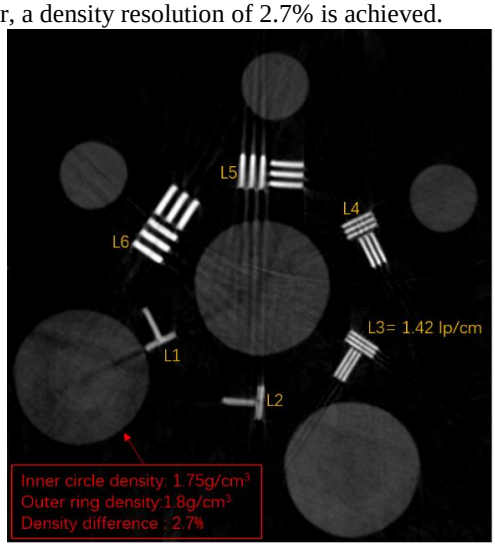


Fig. 9. The enlarged electron density image of the testing phantom.

TABLE II
PERFORMANCE OF MEV CT SYSTEM

Parameters	Value
Spatial resolution	1.42lp/cm
Density resolution	2.7%

Fig. 10 shows another imaging case of a van, we can clearly see the engine and a fire extinguisher in the image.

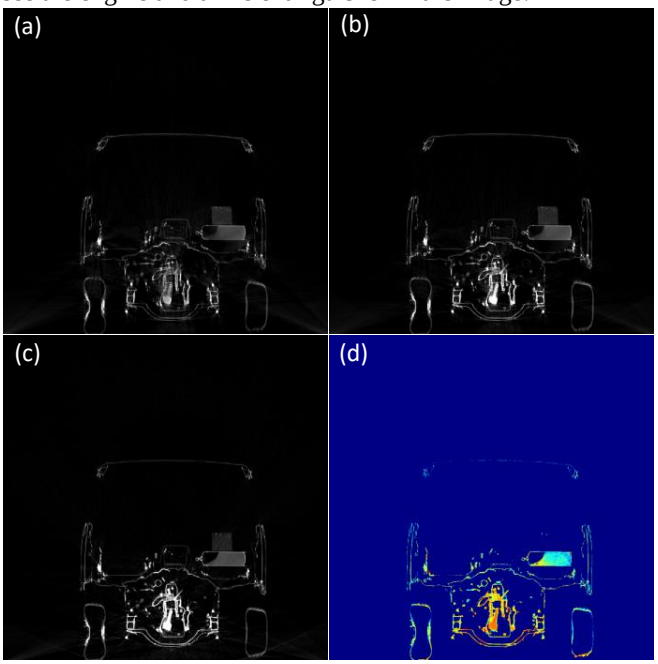


Fig. 10. Imaging results of a van. (a) and (b) are respectively low and high energy reconstruction image. (c) electron density image. (d) equivalent atomic number image.

IV. CONCLUSION AND DISCUSSION

In this paper, we present the experimental results of world's first dual-energy MeV CT for cargo imaging. The significance of building cargo imaging CT is to solve the overlapping problem existed in the traditional X-ray radiography image. We introduce the system geometric design, the reconstruction and material decomposition algorithm of the practical system. This system uses two linear accelerators with two fast switching spectra of 6MeV and 9 MeV as the X-ray source, two arc detector arrays and a rotation system for scanning. The diameter of the field of view is up to 5.12 m and is big enough for the truck loaded with container to pass by. GPU accelerated reconstruction algorithm is implemented for this system. The imaging results of a truck loaded with a testing phantom and a van are presented. The spatial and density resolution of this CT system is evaluated, which is 1.42lp/cm and 2.7%. Post-processing decomposition method is applied. The decomposition results including electron image and equivalent atomic number image are presented. We will focus more on this part in the future to make the decomposition results more accurate. Overall speaking, the dual-energy MeV CT system gives a brand-new perspective on visualising and quantitatively inspecting the container or cargo.

REFERENCES

[1] G. Chen, "Understanding X-ray cargo imaging," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 241, pp. 810-815, 2005.
[2] G. Chen, G. Bennett, and D. Perticone, "Dual-energy X-ray radiography for automatic high-Z material detection," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 261, pp. 356-359, 2007.
[3] P. M. Shikhaliev, "Megavoltage cargo radiography with dual energy material decomposition," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 882, pp. 158-168, 2018.
[4] *http://www.saic.com*.
[5] *http://www.rapiscan.com*.
[6] *http://www.nuctech.com*.
[7] "http://www.L-3com.com."
[8] *http://www.heimannsystems.com*.
[9] P. M. Shikhaliev, "Large-scale MV CT for cargo imaging: A feasibility study," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 904, pp. 35-43, 2018.
[10] L. Li, T. Zhao, and Z. Chen, "First Dual MeV energy x-ray CT for container inspection: design, algorithm, and preliminary experimental results," *IEEE Access*, vol. 6, pp. 45534-45542, 2018.
[11] D. L. Parker, "Optimal short scan convolution reconstruction for fan beam CT," *Medical physics*, vol. 9, pp. 254-257, 1982.
[12] W. A. Kalender, W. Perman, J. Vetter, and E. Klotz, "Evaluation of a prototype dual‐energy computed tomographic apparatus. I. Phantom studies," *Medical physics*, vol. 13, pp. 334-339, 1986.
[13] T. Zhao, L. Li, and Z. Chen, "K-edge eliminated material decomposition method for dual-energy X-ray CT," *Applied Radiation and Isotopes*, vol. 127, pp. 231-236, 2017.
[14] Y. Xing, L. Zhang, X. Duan, J. Cheng, and Z. Chen, "A reconstruction method for dual high-energy CT with MeV X-rays," *IEEE Transactions on Nuclear science*, vol. 58, pp. 537-546, 2011.

A new empirical dual-energy calibration (EDEC) method for material decomposition in x-ray cargo inspection

Sunho Lim, Donghyeon Lee, Hwanhui Jo, Seungryong Cho

***Abstract*— Dual-energy x-ray imaging provides material decomposed information that is very useful for x-ray inspection systems. Our focus in this work is on developing an efficient dual-energy calibration method for linac-based high-energy x-ray cargo inspection systems. Conventional calibration methods have the disadvantages of being time-consuming and labor-intensive. CT imaging-domain empirical dual-energy calibration (EDEC) method resolves such problems, and it has recently been applied to linac-based 2D inspection system. We note, however, that the performance is limited in circumstances that often occur in linac-based inspection system. The challenges include 1) large objects beyond the calibration length, and 2) unstable polychromatic x-ray spectrum over time. In this paper, we propose a new calibration method to address this problem. A simulation study shows that the proposed method improves material decomposition performance substantially in such challenging situations.**

***Index Terms*—X-ray cargo inspection, Dual-energy, Material decomposition, Monte Carlo simulation.**

I. INTRODUCTION

Dual-energy x-ray imaging is a technique for obtaining material-specific information of an object by use of projection data acquired at two different x-ray spectra. A method of material classification using dual-energy x-ray imaging was first proposed in 1976 by Alvarez and Macovski. [1] Recently, with increasing interest in the security field, studies have been actively conducted on baggage or container inspection systems. [2-5]

Dual-energy material-decomposition method can be classified into two categories: the calibration process in the projection domain and in the computed tomography (CT) image domain. In general, projection-based material decomposition is more accurate, free from beam hardening artifacts, and requires a matched projection dataset measured on the same physical line in both spectra. Projection-data based material decomposition calibration can be directly used for 2D container inspection. This study focuses on projection-data based material decomposition methods.

The conventional calibration methods in the field use projection data of calibration phantoms with various combinations of thicknesses and materials. However, this method is usually time-consuming and labor-intensive. A method called empirical dual-energy calibration (EDEC) has been developed that uses a calibration phantom that includes various length combinations of two materials depending on the projection angle and uses resulting CT images that are reconstructed out of the projection data acquired over 360 degrees of scan-angle range. [2] The EDEC method enables a fast and efficient calibration. However, this method also has the limitation of poor material discrimination in two specific situations. First, there is a high dependence on calibration phantoms, which results in poor material decomposition for objects with lengths beyond the maximum length contained in the calibration phantom. Second, the EDEC method is quite sensitive to the spectral change of the x-rays over time since the decomposition typically involves higher order polynomial terms of projection data. It has been observed that the LINAC output x-ray may have an unstable operation characteristics or a drift in time. It has been reported that the LINAC system may have an uncertainty of about 3% for maximum energy. [6]

To address above problems, this study proposes a new empirical dual-energy calibration method. Instead of using higher-order polynomial decomposition function in the existing EDEC method, we use the piecewise linear decomposition function. In order to compare the performance of the proposed method with that of the existing method, material decomposition of the test phantom is performed by Monte Carlo simulation. In order to identify the two problems of the conventional method, a comparative study was conducted using test phantoms of various lengths and spectrums of various energy as operational variables. We found that the proposed method outperforms the existing ones in both cases.

II. METHODS

A. Conventional EDEC method

The calibration phantom is scanned at two different spectra. Let q_j be the resulting polychromatic raw data and p_i be the desired monochromatic, material-selective raw data we want to get, where the subscripts i and j denote the type of material and the x-ray spectrum, respectively. Then, the purpose of the calibration process is to find a decomposition function D_i that yields the p_i from q_j.

$$p_i = D_i(q_1, q_2)$$

The conventional EDEC assumes that the unknown decomposition function is a linear combination of the basis function $b_n(q_1, q_2)$, a function of two polychromatic raw data. We then need to determine the type of the basis function and the coefficients corresponding to each basis function to find the decomposition function. For example, conventional EDEC uses a polynomial basis function that includes the higher order terms as following.

$$b_n(q_1, q_2) = q_1^k q_2^l \ (\mathrm{k} = 0,1,2,\dots,\mathrm{K}, \mathrm{l} = 0,1,2,\dots,\mathrm{L})$$

In this case, a total of $(K+1) \times (L+1)$ basis functions are constructed. In our study, we constructed 16 basis functions with $\mathrm{K} = \mathrm{L} = 3$ for each material. The CT image reconstructed from the raw data q_1 and q_2 is

called a basis image as below.

$$f_n(r) = R^{-1} b_n(q_1, q_2)$$

The final material decomposed image is obtained by selectively determining the decomposition function for each material. The material decomposed image $F_i(r)$ expressed as a linear combination of basis images by the linearity of the radon transform R as in the following equation.

$$\begin{aligned} F_i(r) = R^{-1} D_i(q_1, q_2) &= R^{-1} \sum_n c_{i,n} b_n(q_1, q_2) \\ &= \sum_n c_{i,n} R^{-1} b_n(q_1, q_2) = \sum_n c_{i,n} f_n(r) \\ &= \mathbf{c} \cdot \mathbf{f}(\mathbf{r}) \end{aligned}$$

B. Proposed EDEC method

In the conventional EDEC method, relatively high order polynomial is used as the basis function with the maximum order being determined through empirical assessment. In this work, the proposed method uses a piecewise linear basis function. The proposed EDEC method uses the basis function by dividing q_1 and q_2 by the linear attenuation coefficients. The basis function and decomposition function are expressed as follows.

$$b_n(q_1, q_2) = q_{i,h},\ D(q_1, q_2) = \textstyle\sum_h c_{i,h}\, q_{i,h}\ (\mathrm{i} = 1,2, \mathrm{h} = 0,1,2, \dots, 8)$$

, where h denote the histogram bin index of sinogram data. The raw data q_1 and q_2 are divided based on the linear attenuation coefficient and each is used as a basis function. In our study, 16 basis functions are constructed by dividing each raw data into eight bins for a fair comparison with the conventional EDEC method.

C. Optimization procedure

The corresponding coefficients should be obtained after the basis function is constructed according to the above method. To obtain the coefficient c, we compare the material decomposed image with the template image of the calibration phantom. More specifically, we calculate a coefficient set $\mathbf{c}$ by minimizing E^2, which is the least square deviation between the material decomposed image $F_i(r) = \mathbf{c} \cdot \mathbf{f}(\mathbf{r})$ and template image $t(r)$.

$$E^2 = \int d^2r\, w(r)\, (f(r) - t(r))^2$$

The two images consisting of 1 and 0, template image and weight image, are used to determine the coefficients. The template image is the ground truth image of the material decomposed image for each material. The weight image prevents the coefficient determination from interfering with materials other than materials 1, 2 and air.

$$t(r) = \begin{cases} 1\ for\ r \in material\ 1 \\ 0\ for\ r \in material\ 2\ or\ air \end{cases}$$

$$w(r) = \begin{cases} 1\ for\ r \in material\ 1\ or\ 2\ or\ air \\ 0\ for\ r \in unkown\ materials \end{cases}$$

Differentiating E^2 by c_n yields a linear system $a_n = B_{nm} \cdot \mathbf{c}$. Then we can simply get the coefficient set $\mathbf{c}$ by $\mathbf{c} = {B_{nm}}^{-1} \cdot a_n$ where a_n and B_{nm} are expressed as follows.

$$a_n = \int d^2r\, w(r)\, f_n(r) t(r)$$

$$B_{nm} = \int d^2r\, w(r)\, f_n(r) f_m(r)$$

III. SIMULATIONS AND MEASUREMENTS

We performed the simulation of the existing EDEC method and the proposed EDEC method. To demonstrate that the proposed EDEC method solves the two limitations of the existing EDEC, two simulations were performed to compare the performance of the two methods. First, we performed material decomposition on the objects beyond the range of calibration length. Second, we performed material decomposition for various spectra. All simulations were performed using Gate_v8.0.

A. Calibration process

The EDEC scheme uses an object with different length combinations of two materials for each angle as a calibration phantom. In this study, we used a double-cylinder phantom consisting of a 150 mm radius polymethyl methacrylate (PMMA) cylinder and a 26.5 mm radius iron cylinder as a calibration phantom. The calibration phantom was scanned by two spectra to obtain angle-specific projection images, which were used as calibration data. Detailed parameters are summarized in Table A.

B. Simulation for beyond range of calibrated length

The test phantoms, which are larger than the calibration phantom, were used to compare the material fractionation performance for objects over the calibration length. We used a PMMA wedge phantom of 500mm on one side and an iron wedge phantom of 80mm on one side. In this case, the maximum length of the path through the test phantom is approximately three times longer than that for the calibration phantom. The test phantoms were scanned by the same dual energy x-ray spectrum as the calibration procedure, and the material decomposition was compared with the conventional EDEC and the proposed EDEC method.

C. Simulation for spectrum uncertainty problem

To compare the performance of material decomposition in case of changing spectra, test phantoms of size within the calibration length were scanned by various spectra. PMMA wedge phantom with 200mm of one side and iron wedge phantom with 40mm of one side were used, and x-ray spectra were 6 ± 0.2MV and 9 ± 0.3MV, reflecting an uncertainty of about 3% at maximum energy. In this case, we also compared the decomposed results of the conventional EDEC method with the proposed EDEC method.

IV. RESULTS

The double-cylinder phantom and wedge phantom were used as a calibration phantom and test phantom, respectively.

The calibration was performed in two ways: the conventional EDEC and the proposed EDEC. Thereafter, the results of the simulations were analyzed visually and quantitatively.

A. Calibration process

We performed a calibration procedure according to the methods section before evaluating the two EDEC methods. Fig. 1 shows the reconstruction image, weight image and template image of a double-cylinder phantom required for the EDEC process. Fig. 2, Fig. 3 show the basis image set of the conventional and the proposed EDEC method.

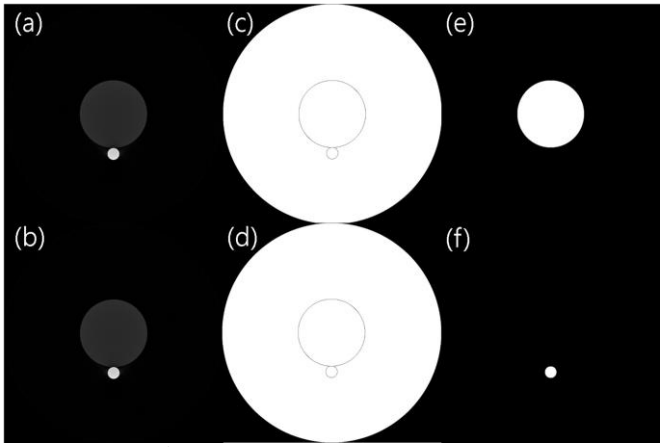


Fig. 1. Images of the calibration phantom in calibration process. (a) reconstruction image using 6MV, (b) reconstruction image using 9MV, (c) weight image of PMMA, (d) weight image of iron, (e) template image of PMMA, (f) template image of iron.

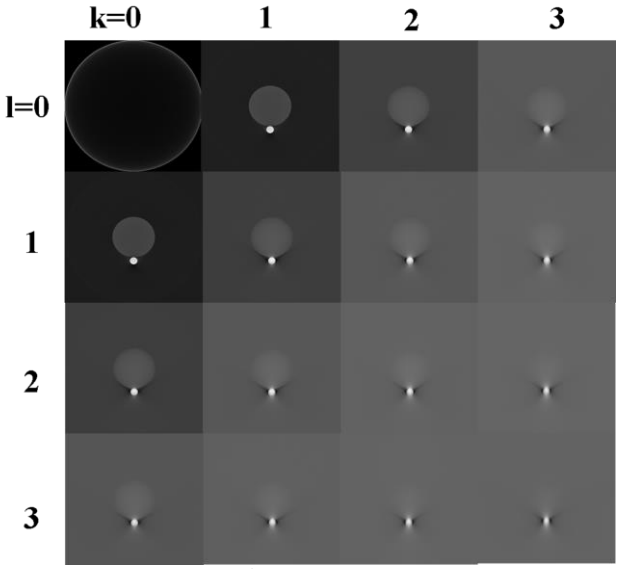


Fig. 2. Basis images obtained from the conventional EDEC method. Note that the window level is chosen for each image to display the range mean±1sigma from black to white.

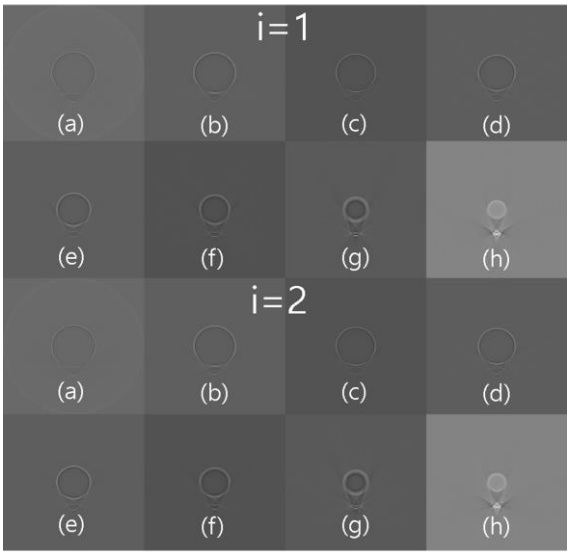


Fig. 3. Basis images obtained from the proposed EDEC method. Note that the window level is chosen for each image to display the range mean±1sigma rom black to white.

B. Result for beyond range of calibrated length

Fig. 4 shows the result images of material decomposition of each materials using two methods. Because the size of the test phantom is beyond the calibration length range, both results degrade the performance of material decomposition within the phantom. However, we can see that there are less dark parts inside the phantom in the proposed method than in the conventional method. For accurate comparisons, we calculated the root mean square error (RMSE) of the ground truth image from the results of both methods. Fig. Figure 5 shows that in all cases, the RMSE is smaller in the proposed method than in the conventional method.

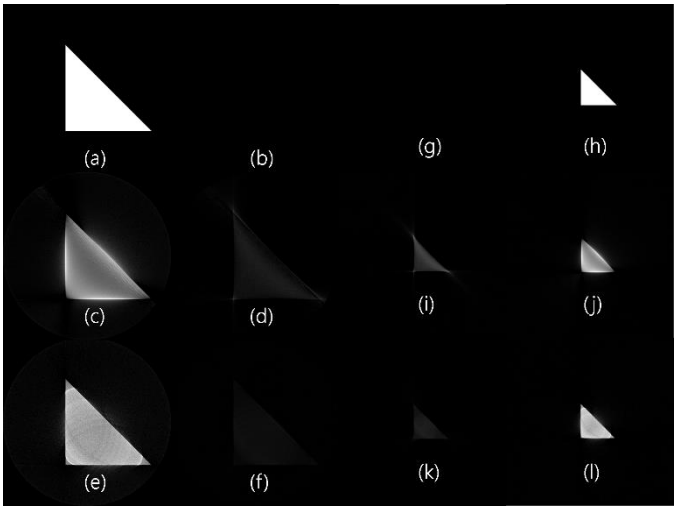


Fig. 4. The first line is the ground truth image, the second is the result of the conventional method, and the third is the result of the proposed method. The first two column represents the PMMA and iron image of PMMA phantom. The next two column represents the PMMA and iron image of iron phantom.

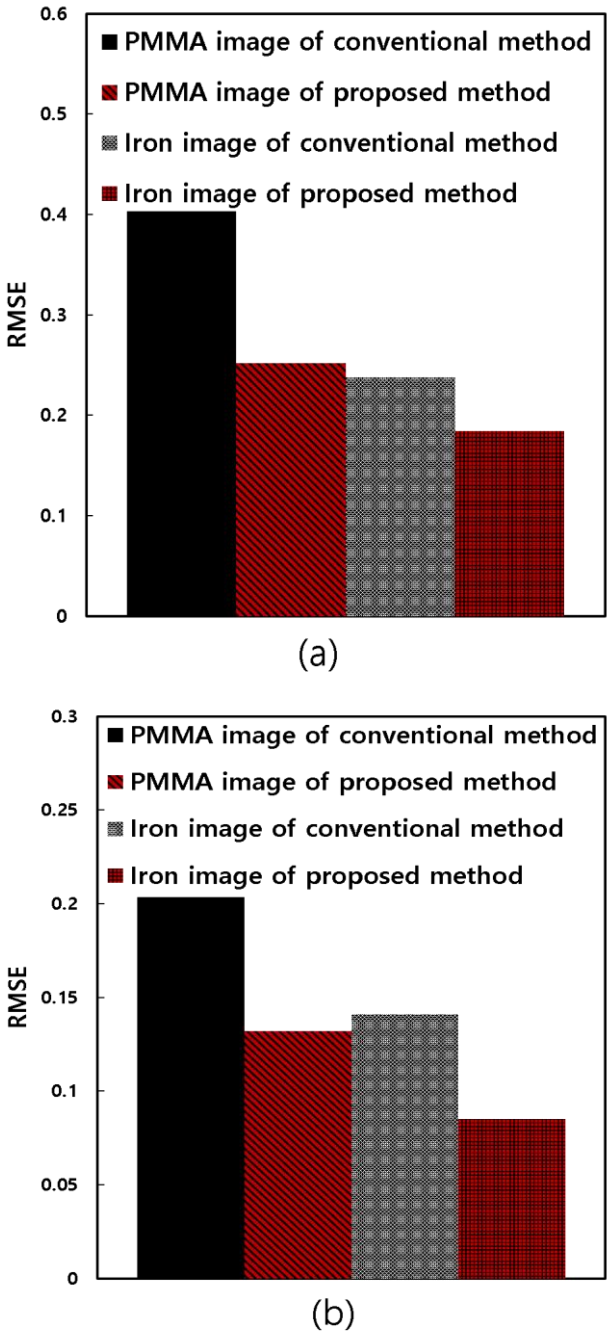


Fig. 5. RMSE comparison of PMMA and iron wedge phantom result between two EDEC methods. (a): PMMA wedge phantom result, (b): Iron wedge phantom result.

C. Result for spectrum uncertainty problem

Fig. 6 shows the results of material decomposition using the two methods for the irradiating spectrum of 6MV and 9MV, 5.9MV and 8.85MV, and 5.8MV and 8.7MV. To see the results clearly, we compared the PMMA images of the iron wedge phantom. Ideally, there should be nothing in the image, but some portion of the wedge phantom appears. In both methods, the results of 6MV and 9MV spectra were relatively successful, but the performance of material decomposition decreased as the spectrum moved away from 6MV and 9MV. For accurate comparisons, we calculated the RMSE of the ground truth image and the results of both methods. Fig. 7 shows that the RMSE is smaller in the proposed method than in the conventional method except for 6MV and 9MV.

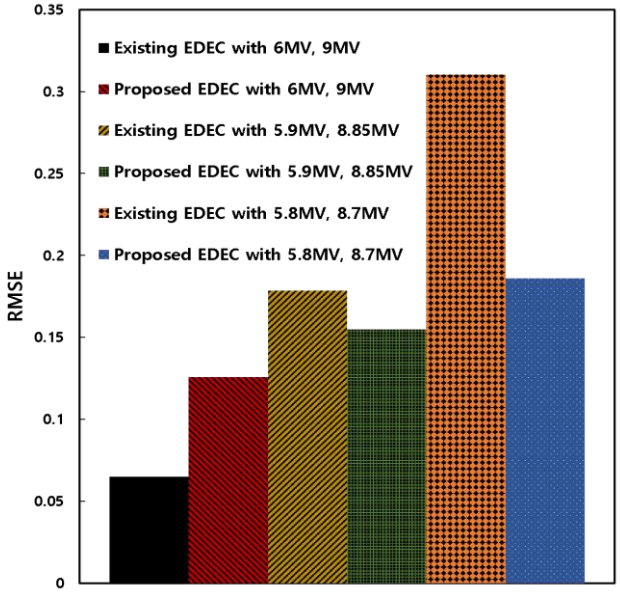


Fig. 7. RMSE values in two methods with three different spectrum combinations.

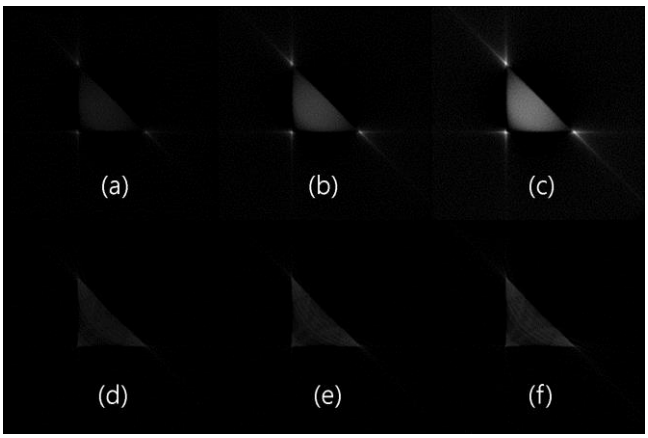


Fig. 6. PMMA image of iron wedge phantom result using conventional EDEC method with (a) 6MV and 9MV, (b) 5.9MV and 8.85MV, (c) 5.8MV, 8.7MV and proposed EDEC method with (d) 6MV and 9MV, (e) 5.9MV and 8.85MV, (f) 5.8MV and 8.7MV

V. DISCUSSION AND CONCLUSIONS

We have proposed a new empirical dual-energy calibration method using a piecewise linear decomposition function in the material decomposition process. The proposed method is compared with the existing method from two perspectives using Monte Carlo simulation. (1: beyond the range of calibrated length problem, 2: x-ray spectrum uncertainty problem) The proposed method turns out to provide more accurate material decomposition for the objects whose path-length exceeds the calibration length. When using a high order polynomial type basis function, the optimization procedure tends to yield an overfitting coefficient for the calibration phantom. It remains as our future work to theoretically approach this finding through a rigorous estimation framework. In conclusion, we showed that the proposed EDEC method successfully manages the two limitations of the existing EDEC method. The proposed method is expected to play an important role as a calibration method for material decomposition in dual energy x-ray inspection applications.

REFERENCES

[1] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray CT," *Phys. Med. Biol.* 21(5), 733– 744 (1976).10.1088/0031-9155/21/5/002

[2] P. Stenner, T. Berkus, and M. Kachelrieß, "Empirical dual energy calibration (EDEC) for cone-beam computed tomography," *Med. Phys.* 34(9), 3630– 3641 (2007).10.1118/1.2769104

[3] Lee, D., Lee, J., Min, J., Lee, B., Lee, B., Oh, K., Kim, J., and Cho, S., Efficient material decomposition method for dual-energy x-ray cargo inspection system, *Nucl. Instrum. Methods Phys. Res.,Sect. A*, 2018, vol. 884, pp. 105–112.

[4] Lee J, Lee Y, Cho S and Lee B-C 2012 A dual-energy material decomposition method for high-energy x-ray cargo inspection J. Korean Phys. Soc. 61 821–4

[5] C. Maaß, M. Baer, and M. Kachelrieß, " Image-based dual energy CT using optimized precorrection functions: A practical new approach of material decomposition in image domain," *Med. Phys.*0094-2405 36, 3818– 3829 (2009).10.1118/1.3157235

[6] Electron Energy Distribution for a Research Electron LINAC Heuijin Lim, Manwoo Lee, Jungyu Yi, Sang Koo Kang, Me Young Kim, Dong Hyeok Jeong Research Center, Dongnam Institute of Radiological & Medical Sciences, Busan, Korea

X-ray Phase-Contrast Tomography of Heart Tissue

Marius Reichardt, Jasper Frohn, Amara Khan, Frauke Alves and Tim Salditt

***Abstract*—In this work, we demonstrate a three-dimensional (3d) imaging approach using x-ray phase-contrast tomography to characterize the structure of small animal hearts. To this end, CT scans of mouse hearts were acquired at a home-built phase-contrast μ-CT setup installed at a liquid metal jet source. This laboratory setup allows to image the entire organ with a resolution below 10 μm and resolves the structure of the arrangement of cardiomyocytes within the tissue for different sample preparations. Based on the reconstructed volumes, the 3d organization of the cardiac structure was analyzed in terms of structural orientation, degree of alignment and dominant fiber thickness using a sub-volume-based algorithm. In order to corroborate the laboratory results, the cardiac structure was also analyzed at synchrotron radiation facilities in a multi-scale approach covering scans of the whole organ and region of interest scans with effective pixel sizes down to less than 200 nm.**

***Index Terms*—x-ray, phase-contrast tomography, cardiac structure, fiber orientation, anisotropy**

I. Introduction

HEART contractility relies on an intricate structure spanning multiple length scales: starting with the molecular scales of the myosin motors and the actomyosin assembly, the formation of the sarcomeric units and myofibrils, to the cellular scale of single cardiomyocytes, arrangement of myocyte chains embedded in connective tissue, all the way to the aggregation of cardiomyocyte chains forming a complex cardiac mesh and finally the scale of the entire organ. While the macroscopic architecture of heart muscle is known for more than hundred years from classical anatomy and histology, the detailed three-dimensional (3d) arrangement of muscle fibers and myofibrils cannot be reconstructed from histology alone, due to the invasive nature of sectioning [1]. Different models for the description of the heart are therefore still under intensive debate [2], [3]. With recent advances in medical imaging techniques, most notably magnetic resonance imaging (MRI) and ultrasound, the 3d structure of the heart can now even be imaged dynamically, however, not at a resolution to resolve the complete fiber network down to the level of a single muscle fiber [4], [5].

In this work we explore an approach to small animal heart imaging using x-ray phase contrast tomography at a compact laboratory instrumentation. Based on the reconstructed electron density, the orientation of the muscle fiber bundles can be determined by sub-volume based algorithms described in [6]. In particular, we show that the image quality and resolution is high enough to deduce the 3d (pseudo-)vector field of the orientation of the cardiac tissue. These results are further supported by a multi-scale imaging approach measurements acquired at synchrotron radiation facilities.

This work is based on: M. Reichardt et al., "Fiber orientation in a whole mouse heart reconstructed by laboratory phase-contrast micro-CT", Proc. SPIE 11113, Developments in X-Ray Tomography XII, 111130S (2019) and M. Reichardt et al., "Multi-scale x-ray phase-contrast tomography of murine heart tissue", Biomed. Opt. Express, (submitted, 2019).

M. Reichardt, J. Frohn and T. Salditt are with the Institute for x-ray physics, Georg-August-University Goettingen, Germany. e-mail: marius.reichardt@stud.uni-goettingen.de.

A. Khan and F. Alves are with the Max Planck Institute for Experimental Medicine, Group of Translational Molecular Imaging, Goettingen, Germany.

II. 3d structure of entire mouse hearts

In order to characterize the cardiac structures using a laboratory μ-CT setup, different sample preparations were analyzed. In specific the standard procedure of paraffin embedding, paraffin embedding with an iodine stain or with an phosphotungstic acid (PTA) stain an embedding in 100% ethanol solution and the evaporation of solvent method (EOS) were performed.

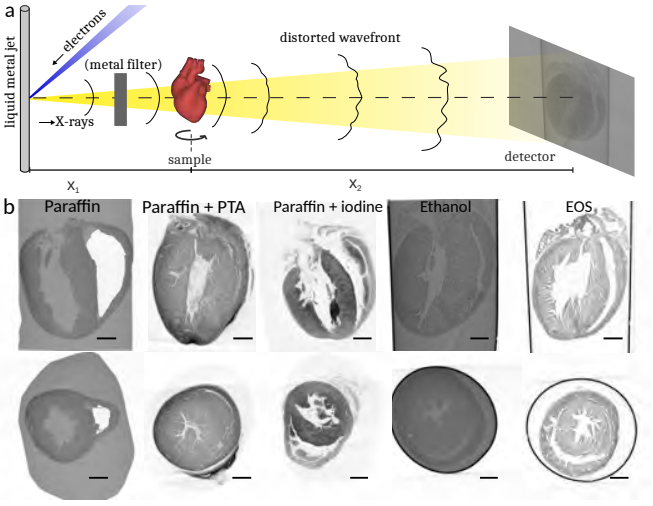


Fig. 1. Sketch of the laboratory liquid metal jet μ-CT setup with electron beam sketched in blue and X-ray beam in yellow. The emission spectrum can be filtered (i.e. pre-hardened) by metal foils. The sample is placed at a distance x_1 behind the source spot and projections are recorded with a flat panel detector at distance x_2. (b) Results of the tomographic reconstructions of the different sample preparations. For each column the a slice of the reconstructed volume along the long axis of the heart and a sagital slice are shown. Scale bars: 1 mm.

The laboratory imaging system for phase contrast tomography is sketched in Fig. 1a. A liquid metal jet μ-CT (Excillum Metaljet-D2) is combined with a flat panel flat panel detector (Dexela CMOS, PerkinElmer Inc., 1536 $\times$ 1944 pixels) with an isotropic pixel size of 75 μm. The cone beam geometry of the setup allows to customize zoom and field of view (FOV) by geometric magnification M. Due to a small effective source spot the effects of source blurring are relatively small and effective pixel sizes dx_{eff} of about 5-6 μm can be selected. In order to increase contrast a phase retrieval following the Bronnikov Aided Correction (BAC) was

performed [7]. Tomographic reconstructions were performed using the ASTRA toolbox [8]. Slices of the volumes are shown in Fig. 1b. The overall structure of the hearts can be seen in all reconstructions, but the visibility of detailed structures within the cardiac tissue differ based on the different preparations. In terms of contrast, metal-based staining of tissue samples result in highest signal-to-noise ratio (SNR). However, there are differences in the details of the tissue structure in both stains PTA can bind with a higher affinity to various cardiac tissue structures such as collagen, fibrin and other fibers of connective tissue. On the contrary, iodine seems to lack this specificity hence, results in blurry reconstructions and randomly clusters at a few regions of the tissue. For further analysis the EOS data set is used, because it has a high SNR and the structure of the tissue is not affected by any staining agent.

III. 3D ORIENTATION OF THE CARDIAC STRUCTURES

For the analysis of the heart structure an algorithm based on a local 3d Fourier transform was used [6]. It is based on the reconstructed electron density of the sample and delivers the fiber orientation, as well as the degree of fiber alignment and the local fiber thickness.

The algorithm is applied to small sub-volumes (probing volumes) of the tomographic reconstruction, which are then shifted through the entire heart. For each data point a cube with a defined edge length l is virtually cut from the volume. Fig. 2 shows (a) the volume rendering from the tomographic reconstruction, (b) an exemplary sub-volume and (c) a 2D slice of this sub-volume. An exemplary outcome of the Fourier transform is shown as an isosurface in Fig. 2d. It is similar to a triaxial ellipsoid and yields information about the spacial frequencies in the volume. In order to quantify the orientation and degree of alignment of this ellipsoid, a principal component analysis (PCA) is applied to the co-variance matrix C

$$C = \begin{pmatrix} \langle q_{xx} \rangle & \langle q_{xy} \rangle & \langle q_{xz} \rangle \\ \langle q_{yx} \rangle & \langle q_{yy} \rangle & \langle q_{yz} \rangle \\ \langle q_{zx} \rangle & \langle q_{zy} \rangle & \langle q_{zz} \rangle \end{pmatrix},$$

with

$$\langle q_{ij} \rangle = \frac{\int I(q_x, q_y, q_z)\ q_i q_j\ d\vec{q}}{\int I(q_x, q_y, q_z) d\vec{q}}$$

for each sub-volume of the reconstruction. The corresponding eigenvalue problem $C \cdot \vec{b_i} = \lambda_i \cdot \vec{b_i}$ yields three eigenvectors $\vec{b_i}$ with $i \in 1, 2, 3$, which form a new orthogonal basis. The eigenvalues $\lambda_i = \sigma_i^2$ correspond to the variance along the three principal directions. Two eivenvectors are sketched in the 2D slice of the Fourier transform in Fig. 2e.

The fiber direction is defined by the eigenvector with the smallest eigenvalue. In addition, the tissue anisotropy, i.e. local degree of orientation, is quantified in terms of an order parameter Ω defined as

$$\Omega = \sqrt{\frac{(\frac{1}{\lambda_1} - \frac{1}{\lambda_2})^2 + (\frac{1}{\lambda_2} - \frac{1}{\lambda_3})^2 + (\frac{1}{\lambda_3} - \frac{1}{\lambda_1})^2}{2(\frac{1}{\lambda_1}^2 + \frac{1}{\lambda_2}^2 + \frac{1}{\lambda_3}^2)}}.$$

Without preferable direction we get $\Omega = 0$, while $\Omega = 1$ is the limit of perfect structural alignment. Ω describes the

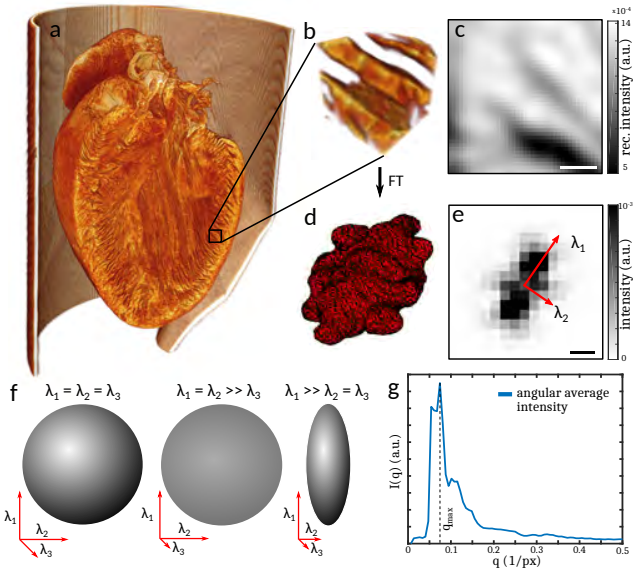


Fig. 2. Workflow of the fiber tracking algorithm. (a) Volume rendering of the reconstructed electron density of a mouse heart treated with the evaporation of solvent method. (b) Small cubic sub-volume with a defined edge length. (c) Two-dimensional slice of the sub-volume. Scale bar: 50 μm (d) 3d Fourier space of this volume. By comparing the eigenvalues of the PCA, a degree of anisotropy in each sub-volume can be defined. (e) 2d slice of the FT. Sketch of two eigenvectors. Scale bar: 0.1 $\frac{1}{\text{px}}$ (f) Possible shapes of the ellipsoid spanned by the results from the PCA. (g) Peaks in Fourier space along the direction of the main component correspond to dominant fiber thicknesses.

degree of orientation and varies within the tissue. By sliding a sub-volume though the whole reconstructed volume, the local orientation and degree of alignment of the entire heart can be determined. Further, we stress that Fourier-based analysis in sub-volumes processed in a sliding window approach can go well beyond simple analysis of anisotropy, since all structural information is contained in the phase and amplitude data. Fig. 2g illustrates a simple example, showing a characteristic peak in the intensity distribution $I(q)$ after a radial average. The peak at $q_{max} = 0.071 \frac{1}{\text{px}}$ corresponds to a real space periodicity of 74.2 μm, attributed to distances between adjacent cardiomyocyte chains.

The results of the analysis including the 3d vector field of alignment, degree of orientation, helical and radial angle, as well as the fiber thickness are shown in Fig. 3. The anisotropy is color-coded in the 3d visualization of the vector field in Fig. 3a. For an improved display, the vector field is cut in the same way as the reconstructed electron density in Fig. 2a. Based on the direction of the structures and its anisotropy, a stream-tracing algorithm implemented in paraview was performed. It results in a complex mesh of traces which is shown in Fig. 3b. The traces are winding around the ventricle in a helical manner but there are also paths crossing with a high helical angle. In order to give a more ordered impression of the structure, two orthogonal slices of the heart are shown in Fig. 3c and Fig. 3d. On the left, the reconstructed electron density of the 2d slices is shown. The structure of the muscle tissue can clearly be derived from the background. Furthermore, the surrounding tube, with a higher electron density can

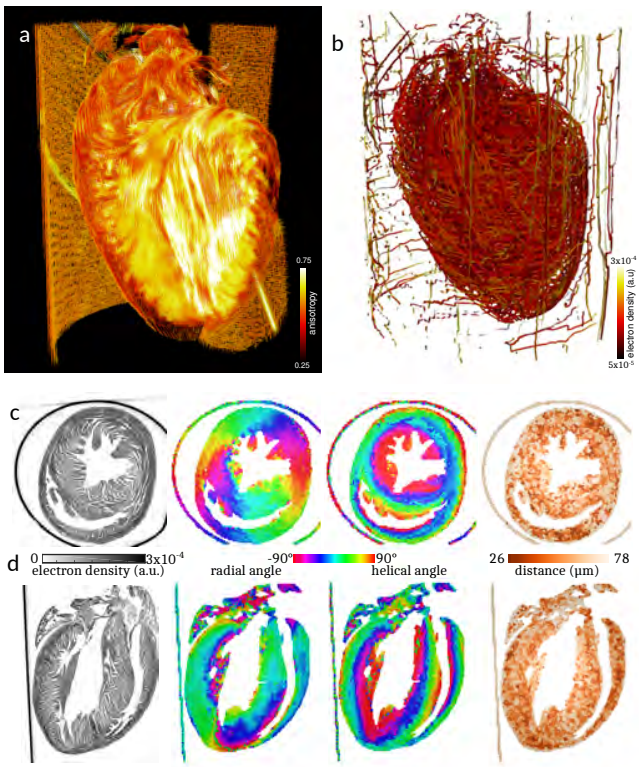


Fig. 3. Results of the FT-based algorithm. (a) Virtual cut of the three dimensional vector field from the FT analysis color-coded with the anisotropy. (b) Streamlines calculated from the vector field (c) lateral and (d) transversal slice of the reconstructed electron density, radial angle, helical angle and calculated fiber distance.

be identified. The middle left shows the radial angle obtained from the analysis. It indicates that the cardiomyocytes are arranged in a loop surrounding the ventricle. In the middle right, the slices are color coded by the the helical angle. The value of the helical angle decreases from epicardium to endocardium. It is approximately 20° and decreases to almost -90° near the lumen. Thus, there is a borderline in the middle of the myocardium around the left ventricle where the helical angle is close to zero. Consequently, the structures are oriented in the radial plane. The combination of radial and helical orientation indicates an arrangement of cardiomyocytes in a closed loop in the radial plane. This supports the old concept of the *Triebwerkzeug* by Krehl [9], according to which a strong muscle ring is surrounding the lumen to assist the contraction of the heart. In addition, the thickness of the structures was determined by the procedure explained above, see the color-coded in slices displayed on the right. It indicates a smaller inter-fiber distance in the apex of the heart and in the wall of the right ventricle. Note that the reconstructed distance highly depends on the preparation and dehydration of the tissue.

IV. Multi-scale imaging of the heart

In order to support the laboratory results, synchrotron experiments with higher resolution were performed at the GINIX of the P10 endstation at the PETRAIII storage ring (DESY, Hamburg). The study of the cardiac structure on multiple length scales from the level of the whole organ down to the (sub-cellular) sarcomeric level requires a very particular experimental approach involving different optical schemes for the illumination with synchrotron radiation. In this work three different configuration of the endstation were used to image the heart with different magnifications. For scans of the whole organ the configuation shown in Fig. 4a was used. To create an illumination area in the size of a whole heart, the undulator beam was broadened using the focus of a pair of KB-mirrors as as a secondary source with a much higher divergence and hence the beam size can is easily be adjusted by the defocus position. The high-spacial frequency contribution emerging from the micro-structure of the mirrors are filtered by placing a pinhole with a diameter of 3 μm in the focus of the KB-mirrors. Since the reflectivity of the mirrors is about 95%, this scheme almost preserves the full flux.

For ROI scans with higher resolution, the second configuration sketched in Fig. 4 was used. To this end, the KB-mirror is moved out by a few millimeters to use the primary beam with a beam size (FWHM) of about 5.8 mm $\times$ 0.77 mm.

For imaging of the cardiac tissue with sub-cellular resolution, a biopsy of the heart tissue near the apex was taken from the previously scanned heart and investigated by holographic tomography using the cone beam geometry with WG illumination (see Fig. 4c). In this geometry, a waveguide is positioned in the focus of the KB mirrors. X-rays are coupled into the waveguide channel which acts as a spatial and coherence filter with a lateral confinement of ≤ 100 nm. The small secondary source size at the exit of the waveguide and the high coherence of the WG illumination enables high geometric magnification $M = \frac{x_{02}}{x_{01}}$ which can be scaled by the source sample distance.

For all projections, recorded at the three different configurations of the endstation, phase retrieval was performed after the empty beam correction. To this end, we first inspected data by linearized phase retrieval according to the contrast transfer function (CTF) algorithm. To this, the approximation of a homogeneous object was used, specifying the phase distribution in the object plane $\phi(\mathbf{r}_\perp)$ by

$$\phi(\mathbf{r}_\perp) = F^{-1}\left[\frac{(\sin(\chi) + \frac{\beta}{\delta}\cos(\chi)) \cdot F\left[I^{(exp)}(\mathbf{r}_\perp) - 1\right]}{2(\sin(\chi) + \frac{\beta}{\delta}\cos(\chi))^2 + \alpha(\mathbf{k}_\perp)}\right],$$

with natural units $\chi = \frac{\lambda x k_x^2}{4\pi}$ for the (squared) spatial frequencies, measured intensities in the detector plane $I^{(exp)}(\mathbf{r}_\perp)$, $\frac{\beta}{\delta}$ ratio of imaginary and real part of the refractive index n, and a frequency-dependent regularization parameter $\alpha(\mathbf{k}_\perp)$. The results of the CTF then served as an initial guess for subsequent iterative phase retrieval based on a nonlinear Tikhonov approach, which does not require the assumption of a weakly varying phase. Further, additional object constraints such as range constraints for the phase shift can be included.The method can be regarded as a nonlinear generalization of CTF phase retrieval. The minimization of the Tikhonov-functional is implemented by a gradient-descent method with adaptive step sizes.

The results of the tomographic reconstructions for the paraffin-embedded heart based on the three different modalities of the beamline setup is shown in Figure 4(d-f). In Fig. 4d,

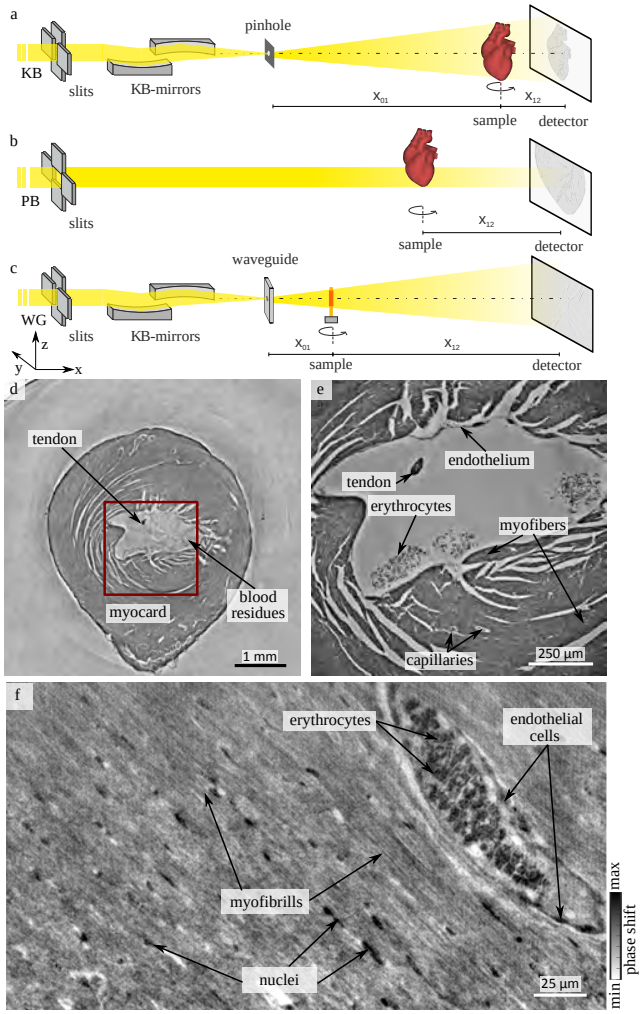


Fig. 4. (a-c)Experimental configurations of the GINIX endstation. (a) KB configuration for overview scans with an FOV of several mm^2. (b) Configuration with parallel beam illumination for ROI scan with an FOV of about 1 mm^2. (c) Configuration with waveguide illumination for high resolution scans of tissue biopsies with an FOV of a few hundred μm. (d) Slice of the tomographic reconstruction acquired by KB-illumination. The red box marks a ROI of a scan with the PB illumination which is shown in (e). (f) In order to image sub-cellular features such as myofibrils, a biopsy of the heart was scanned at high resolution with the WG optic.

a slice near the apex of the heart from the overview scan is shown. Besides the myocardium and some ruptures along the muscle fibers, the single cell layers of endo- and epicardium as well as some blood residues and a tendon within the left ventricle were identified. The red box indicates the area in which a ROI scan was recorded with PB illumination. The corresponding slice is shown in Fig. 4e. In this scan, more details of the tissue structure were resolved. For instance, single vessels and also the single cell layer of epithel cells in the epicardium are visible. The substructure of the cardiomyocytes starts to appear in the reconstruction. The typical elongated shape of the nuclei is also clearly visible as black dots within the myocardium. Moreover, the myofiber orientation of the cardiomyocyte chains becomes visible. Figure 4f shows high resolution data acquired with the WG configuration. This geometry allows identification of sub-cellular structures such as nuclei, erythrocytes and myofibrills. In the given slice, also a blood vessel with the surrounding endothelial cells is visible. Most notably, the structures of the cardiac tissue (from intercellular gaps to the myofibril orientation) are oriented in the same direction for all levels of magnification. This indicates that the orientations reconstructed from the laboratory data with moderate resolution match the molecular organization of the cardiac structure as inferred from the high resolution synchrotron radiation scans.

V. Conclusion

In summary, we have successfully imaged the structure of whole mice hearts by propagation-based phase-contrast using a laboratory μ-CT setup. Due to optimization in sample preparation, image acquisition and suited phase retrieval algorithms, we could extract the myofibril network of small animal hearts. Compared to other techniques as diffusion tensor MRI or grating based x-ray phase-contrast CT [10] which are well suited for low resolution measurements at lower dose, the propagation-based phase-contrast approach allows a scalable magnification at higher resolution data acquisitions. While the laboratory setup allows resolutions below 10 μm in this work, the cardiac orientation was corroborated by a multi-scale imaging approach using synchrotron radiation with the possibility to analyze structures with a resolution down of a few hundred nm.

References

[1] P.-S. Jouk, A. Mourad, V. Milisic, G. Michalowicz, A. Raoult, D. Caillerie, and Y. Usson, "Analysis of the fiber architecture of the heart by quantitative polarized light microscopy. accuracy, limitations and contribution to the study of the fiber architecture of the ventricles during fetal and neonatal life," *European Journal of Cardio-Thoracic Surgery*, vol. 31, no. 5, pp. 915–921, may 2007.

[2] F. Torrent-Guasp, M. J. Kocica, A. F. Corno, M. Komeda, F. Carreras-Costa, A. Flotats, J. Cosin-Aguillar, and H. Wen, "Towards new understanding of the heart structure and function," *European Journal of Cardio-Thoracic Surgery*, vol. 27, no. 2, pp. 191–201, feb 2005.

[3] R. H. Anderson, J. Yanni, M. R. Boyett, N. J. Chandler, and H. Dobrzynski, "The anatomy of the cardiac conduction system," *Clinical Anatomy*, vol. 22, no. 1, pp. 99–113, jan 2009.

[4] B. Bijnens, M. Cikes, C. Butakoff, M. Sitges, and F. Crispi, "Myocardial motion and deformation: What does it tell us and how does it relate to function?" *Fetal Diagnosis and Therapy*, vol. 32, no. 1-2, pp. 5–16, 2012.

[5] C. Mekkaoui, P. Porayette, M. P. Jackowski, W. J. Kostis, G. Dai, S. Sanders, and D. E. Sosnovik, "Diffusion MRI tractography of the developing human fetal heart," *PLoS ONE*, vol. 8, no. 8, p. e72795, aug 2013.

[6] M. Reichardt, M. Töpperwien, A. Khan, F. Alves, and T. Salditt, "Fiber orientation in a whole mouse heart reconstructed by laboratory phase-contrast micro-ct," *Proc. SPIE 11113*, sep 2019.

[7] Y. D. Witte, M. Boone, J. Vlassenbroeck, M. Dierick, and L. V. Hoorebeke, "Bronnikov-aided correction for x-ray computed tomography," *J. Opt. Soc. Am. A*, vol. 26, no. 4, pp. 890–894, 2009.

[8] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible X-ray tomography using the ASTRA toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25 129–25 147, 2016.

[9] L. I. Krehl, *Beiträge zur kenntnis der füllung und entleerung des herzens*. Hirzel, 1891.

[10] H. Hetterich, S. Fill, J. Herzen, M. Willner, I. Zanette, T. Weitkamp, A. Rack, U. Schüller, M. Sadeghi, R. Brandl, S. Adam-Neumair, M. Reiser, F. Pfeiffer, F. Bamberg, and T. Saam, "Grating-based x-ray phase-contrast tomography of atherosclerotic plaque at high photon energies," *Zeitschrift für Medizinische Physik*, vol. 23, no. 3, pp. 194 – 203, 2013.

Tensor Tomography of Dark Field Scatter Using Single-Exposure Moiré Fringe Analysis of X-ray Bi-prism Interferometry

Grant T. Gullberg, Weijie Tao, Sally J. Kim, Yongjin Sung, Qiu Huang, Youngho Seo, Michael Fuller

Abstract– **Technology advances in X-ray phase-contrast interferometry have provided methodology for obtaining simultaneously excellent soft-tissue contrast of phase, attenuation, and small angle scatter. In this work we present new hardware concepts for an X-ray interferometer wherein the phase gratings are replaced with an array of Fresnel bi-prisms and Moiré fringe analysis is used to replace "phase stepping" popular with grating-based setups. Our goal is to develop imaging systems based on bi-prism interferometry with improved polychromatic performance that produce high-contrast fringes with spatially incoherent X-ray illumination. In our investigation we develop a method for reconstructing a vector field of scattering directions that can be used to characterize a three-dimensional tensor field using this technology. Results of phantom simulations using wave optics representation of X-ray dark field projections compares reconstructions of asymmetric scattering fields from "phase stepping" projections with those of single Moiré projections. This will eventually be used in modified bi-prism-based differential phase-contrast imaging to obtain information of tissue properties through mathematical reconstruction of tensor tomographic data from single-exposure Moiré fringe analysis.**

I. Introduction

To take advantage of the full potential of X-ray CT, there is a need for imaging systems that measure the full extent of the physics (attenuation, phase, small angle scatter) involved in the interaction of X-rays with soft tissue. This can be accomplished by developing imaging systems that can provide highly resolved X-rays with spatially-modulated intensity. Fine modulation is feasible even with a conventional X-ray tube by using either a Talbot–Lau interferometer with diffraction gratings [1], or an interferometer with refractive bi-prisms [2], [3] in which there are no optics in the X-ray beam between the sample and the detector [4]. Interferometry-based X-ray imaging has shown to provide in addition to attenuation properties of tissues, common in present day X-ray CT, the potential to provide excellent soft-tissue contrast of phase and small angle scatter information.

In our investigation (Fig. 1) [4], we propose a novel X-ray interferometer wherein an array of Fresnel bi-prisms produces interference fringes with X-rays from a source grating (G_0). The system does not have a phase stepping grating (G_2) common in most X-ray phase contrast imaging systems, but rotates an analysing grating (G_2) to produce a Moiré pattern with a period three times the resolution of the detector [5]-[7]. The source grating G_0 forms multiple mutually-incoherent sources of X-ray illumination. X-rays from a single source refracting through each element of a bi-prism overlap as if proceeding from two slightly separated virtual sources of coherent rays. Through selected geometries (bi-prism period, component spacing, etc.), the bi-prism array interferometer provides spatially modulated intensity across a wide field as each center fringe thus produced falls in a resonant position at the detector. An advantage is found in improved fringe visibility with broad X-ray spectra in that an X-ray bi-prism material has widely varied refracting power relative to wavelength and thus the separation of the virtual sources also varies with X-ray wavelength. When using a detector/scintillator with small hexagonal elements, which acts as a G_2 analyser grating, the setup allows for rotating G_0 and the bi-prism array around the longitudinal axis or instead the sample itself to capture different angular views needed for analysing the X-ray scattering resulting from the orientation of a sample's internal microstructure.

This work was supported in part by the National Institutes of Health under Grant Nos. R43 EB027535, R01 EB026332, and R01 HL135490.

Grant T. Gullberg is with the Department of Radiology and Biomedical Imaging, University of California San Francisco, 94143-0946 USA (e-mail: gtgullberg@lbl.gov).

Weijie Tao is with the Department of Radiology and Biomedical Imaging, University of California San Francisco, 94143-0946 USA and School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China (e-mail: weijie.tao@ucsf.edu)

Sally J. Kim is with the Department of Radiology and Biomedical Imaging, University of California San Francisco, 94143-0946 USA (e-mail: jiwhokim@gmail.com).

Yongjin Sung is with Department of Biomedical and Mechanical Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI, USA (e-mail: ysung4@uwm.edu)

Qiu Huang is with School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China and Department of Nuclear Medicine, Rui Jin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China (e-mail: qiuhuang@sjtu.edu.cn)

Youngho Seo is with the Department of Radiology and Biomedical Imaging, University of California San Francisco, 94143-0946 USA (e-mail: youngho.seo@ucsf.edu).

Michael Fuller is with TF Instruments, Salinas, CA USA (e-mail: michael.fuller@tfinstruments.com).

Our simulation methodology used to investigate our imaging system is based upon our work published in [8]-[11] using wave optics representation of X-rays [12], [13] to produce dark field projections [14]-[16]. To investigate algorithms, projections are simulated with conventional G_1, G_2 gratings instead of a bi-prism for X-ray phase contrast imaging of an asymmetric scattering phantom. A rigorous forward model for phase imaging of a 3-D object illuminated by an X-ray beam is based on a full-wave approach [8] valid under the first Rytov approximation [9]. Therefore, our model accurately calculates X-ray attenuation, phase, and small angle scattering (dark field) images formed with sources of arbitrary shape, and is incorporated into X-ray phase contrast tomography as a forward model so that we can evaluate our reconstruction of vector and tensor scattering fields.

For our imaging system based on bi-prism interferometry, we propose to develop algorithms for tomographic reconstructions from projections of phase contrast data to retrieve images of absorption, differential phase, and dark field small-angle scatter. In this particular work, we focus on the reconstruction of small angle scatter. The simulated projections of X-ray wave optics are reconstructed by modeling dark field projections of a finite set of fixed scattering directions at each voxel in space.

II. MODEL OF SCATTER

Following the work in [17]-[20], we develop here analytical expressions for the dark field projections of a finite set of fixed scattering directions. We assume that an X-ray beam proceeds from a spatially incoherent planar source and illuminates a thin bi-prism or grating emitting an irradiance distribution projected onto an object of interest. If we assume a fixed finite number of scattering directions $\epsilon_k = \zeta_k(x)\hat{\epsilon}_k \in \mathbb{R}^3$ and use expressions in [17] and [18], dark field projections of the image have the form

$$d_j = exp\left[-\int_{L_j} \Sigma_k \langle |\hat{l}_j \times \hat{\epsilon}_k| \zeta_k(x)\hat{\epsilon}_k, t_j\rangle^2 \, dx\right] \quad , \qquad (1)$$

where $\hat{l}_j \in \mathbb{R}^3$ is the direction of the incoming beam, L_j is the line along this direction, and $t_j = |t_j|\hat{t}_j \in \mathbb{R}^3$ is the sensitivity direction parallel to the detector surface. The $\hat{}$ indicates a unit vector. One can show that this reduces to

$$d_j = exp\left[-\sum_k \nu_{kj} \int_{L_j} \eta_k(x) dx\right] \quad ,$$

where $\eta_k(x) = \zeta_k(x)^2$ and $\nu_{kj} = |t_j|^2 |\hat{l}_j \times \hat{\epsilon}_k|^2 \langle \hat{\epsilon}_k, \hat{t}_j \rangle^2$. The $\eta_k(x)$ at the position x are the square of the coefficients of the vector scattering directions $\hat{\epsilon}_k$.

Let's parameterize the X-ray direction so that we replace j with i, j, θ_q, ϕ_r, such that the capital letters I, J, Q, R are the dimensions for each coordinate. Defining the detector elements with coordinates (i, j) and the projection angles of the sample as θ_q, ϕ_r as shown in Fig. 2, we have dark field measurements

$$D(i, j, \theta_q, \phi_r) = exp\left[-\sum_k \nu_{kijqr} \int_{L_{ijqr}} \eta_k(x) dx\right] \quad . \qquad (2)$$

We can form the reconstruction problem as the solution to a large system of linear equations

$$m = (m_{ijqr}) = \left(-ln\left(D(i, j, \theta_q, \phi_r)\right)\right) = H\eta$$

where $H = (D_1 A, D_2 A, \ldots, D_K A)$ is a J×IK matrix and

$$D_k = \begin{pmatrix} \nu_{k1111} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nu_{kIJQR} \end{pmatrix}$$

is a IJQR×IJQR diagonal matrix of weighting coefficients ν_{kijqr}, IJQR is the number of projection samples, NMP is the number of voxels in the 3-D array, K is the number of scattering directions $\hat{\epsilon}_k$, and

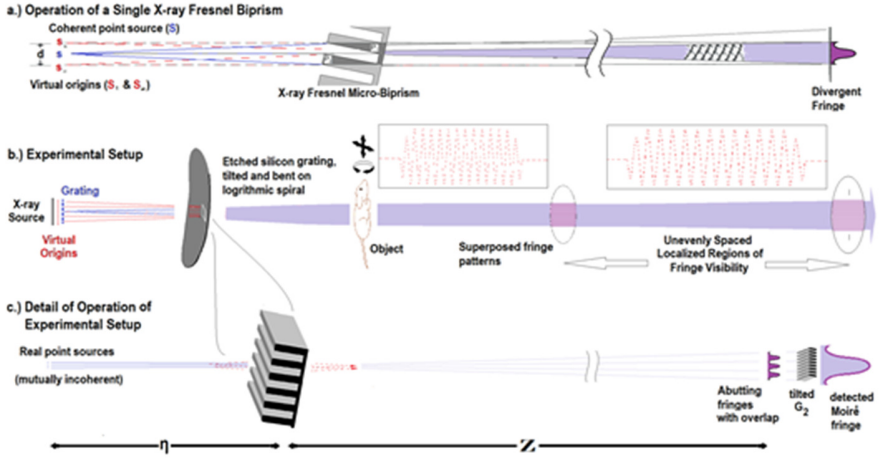
$$\eta^T = (\eta_{11}, \eta_{12}, \ldots, \eta_{1NMP}, \eta_{21}, \eta_{22}, \ldots, \eta_{2NMP}, \ldots \ldots, \eta_{K1}, \eta_{K2}, \ldots, \eta_{KNMP})$$

is a NMPK×1 matrix of unknown coefficients to be determined. A is the system matrix of the tomographic projections formed by the integral in Eq. (2).

Writing the matrix formulation of the system of equations explicitly in terms of the unknown coefficients η_{ij}, we have

$$m = H\eta = \overbrace{(D_1 A, D_2 A, \ldots D_K A)}^{IJQR \times KNMP} \overbrace{\begin{pmatrix} \eta_{11} \\ \vdots \\ \eta_{1NMP} \\ \eta_{21} \\ \eta_{22} \\ \vdots \\ \eta_{2NMP} \\ \vdots \\ \eta_{K1} \\ \eta_{K2} \\ \vdots \\ \eta_{KNMP} \end{pmatrix}}^{KNMP \times 1} \quad . \qquad (3)$$

Fig. 1. Schematic diagram of X-ray optical approach, showing (a.) operation of a Fresnel x-ray bi-prism, (b.) experimental setup with multiple X-ray origins each producing fringes in resonant position, and (c.) detail of production of Moiré fringes via a tilted grating G_2 at the detector.



III. Simulating Dark Field Projections

Dark field projections were simulated using an integrated wave optics framework [12] to model X-ray-matter interaction and free-space propagation. A forward model for phase imaging of a 3-D object illuminated by an X-ray beam is based on a full-wave approach [8] of the scattered field under the first Rytov approximation [9]. Our model accurately calculates X-ray phase images formed with sources of arbitrary shape, and objects of preferential scattering directions.

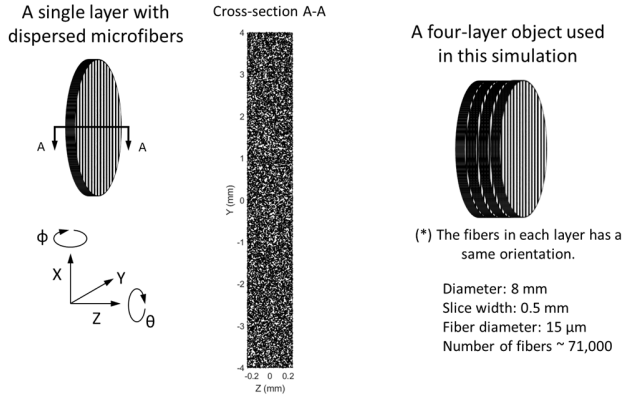


Fig. 2. Phantom used in simulations. The phantom consisted of four layers of parallel carbon microfibers to provide preferential scatter in the direction of the fibers. The angles ϕ and θ show the rotation directions of the phantom.

X-ray projections were simulated for an asymmetric scattering phantom (Fig. 2). The object consisted of four layers of carbon fibers. Each layer consisted of an array of 71,000 fibers of 15μm diameter. Each layer width was 0.5mm and the diameter of the disk was 8mm. The solid carbon fibers in each layer had the same orientation providing preferential scatter along the fibers. In these preliminary simulations, G_1 was not a bi-prism but a grating same as G_2 with transmitting slits of 10.2 mm and period of 5×10^{-6} meters.

Phase Stepping. The projection of the irradiance distribution onto the detector surface was approximated by fitting eight phase stepping projections to the Fourier expansion [20]

$$I(i,j,\theta_q,\phi_r,x_g) \approx a_0(i,j,\theta_q,\phi_r) + a_1(i,j,\theta_q,\phi_r)cos\left[\frac{2\pi(i-(I+1)/2)}{x_p}x_s - \Phi\,(i,j,\theta_q,\phi_r)\right], \quad (4)$$

where (i,j) are coordinates of the detector pixel ($0 \le i \le I$); x_s is the spatial sampling; x_p is the period in x; θ_q, ϕ_r is the rotation angle of the sample around the optical axis; and a_0, a_1, and Φ are the mean, amplitude, and phase of the sinusoidal curve, respectively. Defining the visibility of the scatter and reference signal as

$$V^s(i,j,\theta_q,\phi_r) = a_1^s(i,j,\theta_q,\phi_r)/a_0^s(i,j,\theta_q,\phi_r) \ ,$$

the dark field scatter signal is the following ratio of normalized visibilities (V^r is reference visibility)

$$m(i,j,\theta_q,\phi_r) = -\log\left[D(i,j,\theta_q,\phi_r)\right] = -\log\left[\frac{V^s(i,j,\theta_q,\phi_r)}{V^r(i,j)}\right] .$$

An example of one of the dark field projections D is shown in Fig. 3. One sees preferential scatter oriented at approximate 45° from the vertical axis. There also appears to be some isometric scatter throughout the projection.

Moiré Fringe Analysis. A Moire´ pattern was superimposed on the detector irradiance distribution by rotating the G_2 grating (Fig. 1) by about 3.5833º. The detector pixel size was 4×10^{-5} meters and the period of the Moiré pattern was approximately 8×10^{-5} meters, twice that of the detector pixel size. For phase stepping, we used an image size of 16384×256, but for the Moiré pattern, we used 16384×16384. Each Moire´ projection was processed using Fourier analysis [7], so that the small angle scatter information encoded by the G_1 grating could be extracted. Let $|\mathcal{F}_n^{-1}|$ be the absolute value of the inverse Fourier transform (FT) around an area of the n[th]-order harmonic. The visibility was calculated by taking the ratio of the absolute values of the inverse FT of the area around the first-order harmonic and of the area around the zero-order harmonic $V = |\mathcal{F}_1^{-1}|/|\mathcal{F}_0^{-1}|$. Thus, the dark field image:

$$m = -\log(D) = -\log\left(\frac{V_{obj}}{V_{ref}}\right) .$$

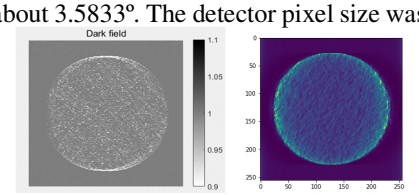


Fig. 3. **A.** Example of a dark field projection (θ = 45°, ϕ=0°). Preferential scattering is shown rotated counterclockwise. **B.** Reconstruction of absorption image using conjugate gradient algorithm for projection angles θ and ϕ in text.

Figure 4 compares projections obtained by phase stepping and that obtained by Moiré fringe analysis.

IV. Reconstructing Dark Field Projections

The tomographic data were simulated assuming an X-ray energy of 17.5 keV. A total of 546 parallel projection images were formed for phantom rotation angles θ from 0° to 90° at 18° steps (6 angles), and ϕ from -90° to 90° at 2° steps (91 angles). Eight phase steps were used to form projections for reconstruction that were compared with reconstructions of projections formed from a single exposure Moiré fringe analysis. The size of each projection image was 16384 × 256, which was downsampled to 256 × 256. Note that the image was oversampled along the y-scan direction to capture the small-angle scattering.

The reconstructions were performed using 25 iterations of the MLEM algorithm to maximize the likelihood function:

$$f(\eta) = L(m|\eta) = \prod_i \frac{e^{-\bar{m}_i}(\bar{m}_i)^{m_i}}{m_i!} \ , \quad (5)$$

for measurements $\bar{m} = H\eta$. The model η included $K = 7$ vectors as the basis for each voxel. Every voxel was spanned by the 7 vectors: [1,0,0], [0,1,0], [0,0,1], $\left[\sqrt{1/3},\sqrt{1/3},\sqrt{1/3}\right]$, $\left[-\sqrt{1/3},\sqrt{1/3},\sqrt{1/3}\right]$, $\left[\sqrt{1/3},-\sqrt{1/3},\sqrt{1/3}\right]$, $\left[-\sqrt{1/3},-\sqrt{1/3},\sqrt{1/3}\right]$.

The reconstruction of the vector coefficients for each voxel is shown in Fig. 5. The image intensity in each voxel correspond to the vector amplitude. The three images across each column are the transaxial (x-y plane), coronal (x-z plane) and sagittal section (y-z plane) through the central axis of the phantom. The images going down the columns correspond to the amplitude of the 7 vectors in the sequence above starting with [1,0,0], at the top and $\left[-\sqrt{1/3},-\sqrt{1/3},\sqrt{1/3}\right]$ at the bottom of each column.

V. Discussion

Here we present results of the reconstruction of coefficients of a vector basis from simulated projections of X-ray small angle scatter. At every image voxel a fixed set of scatter vector directions is assumed for which the coefficients are estimated from the measured projections m. Maximizing the likelihood in Eq. (5), provides estimates of these coefficients from which a tensorial representation can be obtained by fitting the weights of the fixed set of vectors to ellipsoids [17], [18]. In the present work the estimates of the coefficients were obtained by MLEM algorithm assuming Poisson noise; however, the results show some limited angle artifacts. If angular sampling is limited as in the present case with sampling only 6 angles in θ, one potential solution would be to implement a compressed sensing CT reconstruction method [21]-[22]. For more complex samples, it may also be necessary to measure projections around the y-axis.

We feel that bi-prism-based interferometry has the potential to revolutionize phase-contrast and dark-field X-ray CT. The possible benefits of this approach as compared to conventional phase contrast imaging using phase and analyzer gratings are 1) Greater optic-to-detector distance allows placement of the sample after the bi-prism array; 2) Divergent nature of the bi-prism's fringes allows detection in a single exposure of a Moiré pattern; 3) Setup is well suited to commercially available multi-spot X-ray sources; 4) Better use of a polychromatic X-ray spectrum by the naturally varied offset (wavelength dependent) of the bi-prism's virtual sources; and 5) Easier manufacturing, straightforward alignment, and improved throughput.

References

[1] F. Pfeiffer, O. Bunk, C. David, M. Bech, G. L. Duc, A. Bravin, and P. Cloetens, "High resolution brain tumor visualization using three-dimensional x-ray phase contrast tomography," *Phys. Med. Biol*., vol. 52, pp. 6923-6930, 2007.

[2] J. M. Ablett, K. Evans-Lutterodt, and A. Stein, "Hard x-rays Fresnel prisms: Properties and applications," in Design and Microfabrication of Novel X-ray Optics II, A. A. Snigirev and D. C. Mancini, Eds. Proceedings of SPIE, vol. 5539, August 2, 2004.

[3] A. Doblas, G. Saavedra, M. Martinez-Corral, J. C. Barreiro, E. Sanchez-Ortiga, and A. Llavador, "Axial resonance of periodic patterns by using a Fresnel biprism," *J. Opt. Soc. Am. A*, vol. 30, pp. 140-148, 2013.

[4] G. T. Gullberg, M. Fuller, U. Shrestha, and Y. Seo, "Tensor tomography of dark field scatter using x-ray interferometry with bi-prisms," 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC), Oct. 21 – 28, 2017, Atlanta, Georgia.

[5] E. E. Bennett, R. Kopace, A. F. Stein, and H. Wen, "A grating-based single-shot x-ray phase contrast and diffraction method for *in vivo* imaging," *Med. Phys.*, vol. 37, pp. 6047-6054, 2010.

[6] N. Bevins, J. Zambelli, K. Li, and Z. Qi, "Multicontrast x-ray computed tomography imaging using Talbot-Lau interferometry without phase stepping," *Med. Phys.*, vol. 39, pp. 424-428, 2012.

[7] M. Seifert, M. Gallersdörfer, V. Ludwig, M. Schuster, F. Horn, G. Pelzer, J. Rieger, T. Michel and G. Anton, "Improved reconstruction technique for Moiré imaging using an X-ray phase-contrast Talbot–Lau interferometer," *J. Imaging*, vol. *4*, 62, pp. 68-81, 2018.

[8] Y. Sung, C. J. R. Sheppard, G. Barbastathis, M Ando, and R. Gupta, "Full-wave approach for x-ray phase imaging," *Opt. Express*, vol. 21, pp. 266–277, 2013.

[9] Y. Sung and G. Barbastathis, "Rytov approximation for x-ray phase imaging," *Opt. Express*, vol. 21, pp. 2674–2682, 2013.

[10] Y. Sung, W. P. Segars, A. Pan, M. Ando, C. J. R. Sheppard and R. Gupta, "Realistic wave-optics simulation of X-ray phase-contrast imaging at a human scale," *Sci. Rep.,* vol. 5:12011, 2015.

[11] E. R. Shanblatt, Y. Sung, R. Gupta, B. J. Nelson, S. Leng, W. S. Graves, and C. H. McCollough, "Forward model for propagation-based x-ray phase contrast imaging in parallel- and cone-beam geometry," *Optics Express*, vol. 27(4), 4504, 2019.

[12] D. M. Paganin, Coherent X-ray Optics, Oxford University Press, New York, NY, 2013.

[13] J. W. Goodman, Introduction to Fourier Optics (Third Edition), Roberts & Company, Englewood, Colorado, 2005.

[14] M J Kitchen, D Paganin, R A Lewis, N Yagi, K Uesugi and S T Mudie, "On the origin of speckle in x-ray phase contrast images of lung tissue," *Phys. Med. Biol*., vol. 49, pp. 4335–4348, 2004.

[15] W. Yashiro, Y. Terui, K. Kawabata, and A. Momose, "On the origin of visibility contrast in x-ray Talbot interferometry," *Optics Express*, vol. 18, pp. 16890-16901, 2010.

[16] M. Bech, O. Bunk, T. Donath, R Feidenhans'l, C. David and F. Pfeiffer, "Quantitative x-ray dark-field computed tomography," *Phys. Med. Biol*., vol. 55, pp. 5529–5539, 2010.

[17] A. Malecki, G. Potdevin, T. Biernath, E. Eggl, K. Willer, T. Lasser, J. Maisenbacher, J. Gibmeier, A. Wanner, and F. Pfeiffer, "X-ray tensor tomography," *EPL* (Europhysics Letters), vol. 105(3), 38002, 2014.

[18] J. Vogel, F. Schaff, A. Fehringer, C. Jud, M. Wieczorek, F. Pfeiffer, and T. Lasser, "Constrained X-ray tensor tomography," *Opt Express*, vol. 23, pp. 15134-51, 2015.

[19] Y. Sharma, M. Wieczorek, F. Schaff, S. Seyyedi, F. Prade, F. Pfeiffer, and T. Lasser, "Six dimensional X-ray tensor tomography with a compact laboratory setup," *Appl. Phys. Lett.*, vol. 109, 134102, 2016.

[20] T. H. Jensen, M. Bech, O. Bunk, T. Donath, C. David, I. R. Feidenhans, and F. Pfeiffer, "Directional x-ray dark-field, imaging," *Phys. Med. Biol.*, vol. 55, pp. 3317-3323, 2010.

[21] P. Bao, W. Xia, K. Yang, W. Chen, M. Chen, Y. Xi, S. Niu, J. Zhou, H Zhang, H. Sun, Z. Wang, and Y. Zhang, "Convolutional sparse coding for compressed sensing CT reconstruction," *IEEE Trans. Med. Imag*, vol. 38, pp. 2607-2619, 2019.

[22] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–4807, 2008.

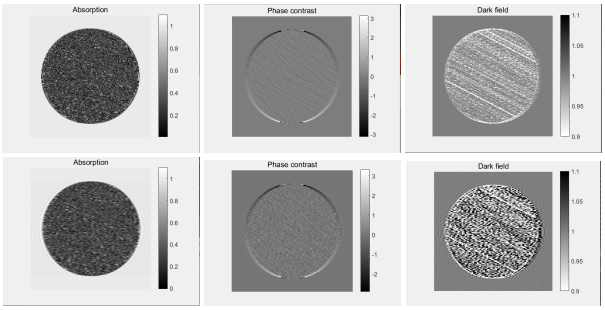


Fig. 4. Example of absorption, phase contrast, and dark field projections obtained by phase stepping (upper) and Moiré fringe analysis (lower). As expected, increased noise is observed in the Moire´ dark field image.

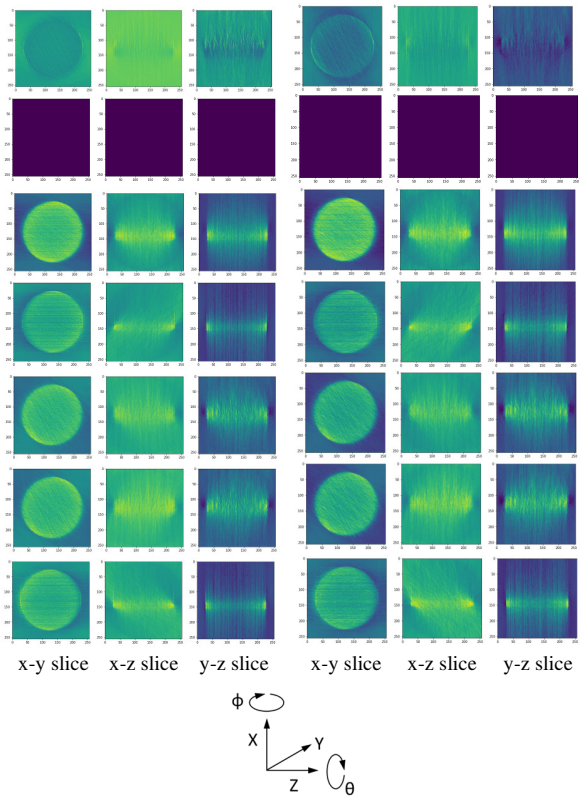
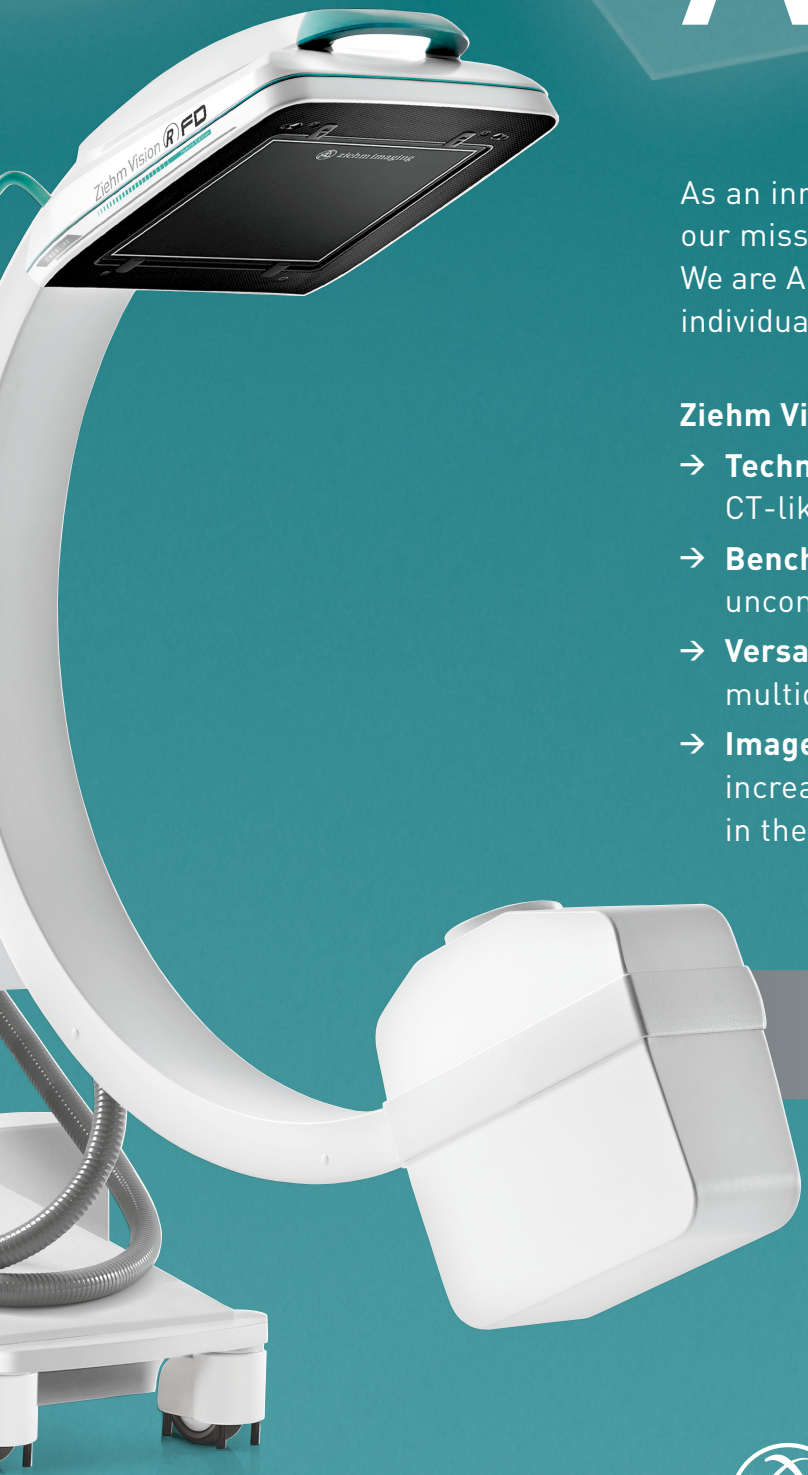


Fig. 5. Reconstructed vector coefficients from projections obtained: from phase stepping (Left); from Moiré fringe analysis (Right). The three images for each at the top are the coefficients for the vector $[1,0,0]$, in the x-y slice, x-z slice, and y-z slice through the center of the phantom, respectively. The series of images from top to bottom are the coefficients for the vectors $[1,0,0]$, $[0,1,0]$, $[0,0,1]$, $\left[\sqrt{1/3},\sqrt{1/3},\sqrt{1/3}\right]$, $\left[-\sqrt{1/3},\sqrt{1/3},\sqrt{1/3}\right]$, $\left[\sqrt{1/3},-\sqrt{1/3},\sqrt{1/3}\right]$, $\left[-\sqrt{1/3},-\sqrt{1/3},\sqrt{1/3}\right]$, respectively. Note: the second row of images are zero because these are coefficients for the vector pointing along the optical axis.

ALWAYS AHEAD



As an innovation leader, we are committed to our mission of setting new technology standards. We are ALWAYS AHEAD in helping you meet your individual goals.

Ziehm Vision RFD 3D

- → **Technological leadership** with CT-like image quality
- → **Benchmark dose settings** and uncompromised image quality
- → **Versatile use** for 2D, 3D and multidisciplinary applications
- → **Image-guided navigation** for increased control and confidence in the OR

Visit us: **www.ziehm.com**

ziehm imaging

Friday, August 7

Observer

Novel CT Technoligies

Developing signal detectability as an image quality metric for use with non-linear image reconstruction

Emil Y. Sidky[1], Weimin Zhou[2], Greg Ongie[3], Juan P. Cruz-Bastida[1], Ingrid S. Reiser[1], Mark A. Anastasio[4] and Xiaochuan Pan[1]

***Abstract*—The overall objective of this research effort is to develop useful image quality metrics that can help to determine parameters of non-linear tomographic image reconstruction algorithms along with parameters of the scan configuration. To illustrate the need for new image quality metrics, images are shown for a breast CT simulation reconstructed by use of total variation (TV) constrained, least-squares minimization. The resulting image properties are difficult to characterize objectively with standard image quality (IQ) metrics. Accordingly, in this work, we aim to develop useful quantitative IQ metrics for this situation. The proposed metrics center on detection of small signals. Specifically, we study the impact of non-linear image reconstruction on signal detectability. We compute the non-linear and linear ideal observer for signal detection in a signal known exactly/background known exactly (SKE/BKE) paradigm. For the results presented here, we focus on a single scan condition and investigate the impact of signal amplitude and region-of-interest for the observer models.**

I. Introduction

Gradient sparsity regularization involves iterative image reconstruction with either a total variation (TV) penalty or constraint. This form of regularization has shown to be effective for low-dose X-ray CT [1] when used properly. As with any other form of regularization, over-smoothing causes artifacts. In the case of gradient sparsity regularization with TV, over-smoothing causes images to have a patchy appearance. Due to the non-linearity of this type of image reconstruction algorithm it is difficult to translate the subjective assessment of patchiness to an objective quantitative measure. We hypothesize that a useful characterization of non-linear image reconstruction can be derived from properties of imaging small signals. The idea being that patchiness may be equated with the degradation of small signals. In particular, we investigate in detail signal detectability in a SKE/BKE paradigm. The aim of developing these metrics is to assist in objectively tuning parameters such as the TV regularization strength to a setting that agrees with subjective visualization.

The setting for this work is a simulation for dedicated breast computed tomography (CT), which is a low-dose imaging modality. We focus on image characterization at a single scan configuration and a single setting of scan parameters. The properties of small signal imaging are investigated in detail. Multiple realizations of noisy data are generated for a uniform background phantom and the same phantom with a signal added. Signal-present realizations are generated for multiple signals at different amplitudes. From reconstruction of these data realizations, statistical properties of the resulting images can be obtained. We compute the ideal linear observer for signal detection, and the ideal non-linear observer for signal detection and compare them against the detectability of the signal in the data, or sinogram, domain. The degradation of the signal detectability is examined as possible useful measure of image quality.

II. Methods

For CT image reconstruction by gradient sparsity regularization, we use total-variation constrained least-squares (TVLSQ) optimization

$$f^\star = \arg\min_f \frac{1}{2}\|Xf - g\|_2^2 \text{ such that } \|\nabla f\|_1 \leq \gamma,$$

where f is the image; X is the X-ray transform matrix; g is the projection data; ∇ is the discrete image gradient operator; $\|\nabla f\|_1$ is the image TV; and γ is the TV constraint value. Regularization decreases with increasing γ, and in this set-up we can use the TV value, γ_0, of the test phantom as a lower limit for the range of regularization. In Sec. III, a noiseless reconstruction is shown and compared with noisy reconstructions that result from use of different γ for realistic breast CT scanning conditions.

[1]The University of Chicago, Department of Radiology MC-2026, 5841 S. Maryland Avenue, Chicago IL, 60637.

[2] Washington University in St. Louis, Department of Electrical and Systems Engineering, St. Louis MO, 63130.

[3]The University of Chicago, Department of Statistics, 5747 S. Ellis Avenue, Chicago IL, 60637.

[4]University of Illinois at Urbana-Champaign, Department of Bioengineering, Urbana IL, 61801.

In order to provide an objective measure of image quality, we are investigating the use of signal detection theory [2]. In the context of linear image reconstruction, we have developed a detection efficiency metric that computes the ratio of ideal observer signal-to-noise-ratios (SNRs) in the reconstructed image domain and the projection data domain [3]. This metric can be computed accurately and is independent of signal amplitude for linear image reconstruction. The TVLSQ image reconstruction model, however, is non-linear and the same methodology cannot be used. The non-linearity of TVLSQ leads to images that cannot be described by a Gaussian model even if the data noise model is Gaussian. Furthermore, the mean of image noise realizations will not be the noiseless reconstructed image as is the case for linear algorithms such as filtered back-projection.

For the detection task considered, we use a uniform disk of attenuation corresponding to adipose tissue. A small two-pixel wide Gaussian signal is placed at the center of the phantom, and the SNR for detection measures the ability to distinguish between signal-absent and signal-present in the presence of noise. A signal amplitude in the object domain is selected so that its SNR for detection by the ideal observer in the data domain is not perfect, corresponding to a two-alternative forced choice (2-AFC) performance between 80 and 90 percent correct or, equivalently, an area-under-the-curve (AUC) of 0.8 to 0.9 for the corresponding receiver operating characteristic (ROC) curve. In Sec. III the mean of noise realizations of the signal-present images is shown to illustrate distortion of this borderline detectable signal due to non-linearity of TVLSQ. For computing the observer models, 4000 data realizations are generated and reconstructed by TVLSQ for both signal-present and signal-absent cases.

It has been shown by Zhou et al. [4] that the ideal linear observer (LO) and ideal non-linear observer (NLO) can be well approximated by machine learning using neural networks. The LO computes the maximum separability in the signal-absent and signal-present images using only linear operations on the images; the NLO has no such constraint and it aims for maximum separability of the two classes of images.

To implement the LO, a linear single-layer neural network (SLNN) was trained by use of the supervised learning method [4]. To approximate the NLO, convolutional neural networks (CNNs) were also trained by the supervised learning method described in Ref. [4]. In this study, the architecture of CNNs comprised three convolutional (CONV) layers, a max-pooling layer, and a fully connected (FC) layer. Each CONV layer comprised 32 filters with 3x3 spatial support and was followed by a LeakyReLU function. The train-validation-test scheme was employed: the SLNNs and CNNs were trained on 3000 pairs of training images, the trained neural networks were specified based on a set of 500 pairs of validation images, and the signal detection performances were evaluated on a set of 500 pairs of testing images. The signal detection performances were quantified by the AUC values, which were estimated by use of the "proper" binormal model [5]. For each of the observer models a region-of-interest (ROI) is selected so that it is centered on the signal. Accordingly, we refer to the ideal linear observer as LO-ROI, and the ideal non-linear observer as NLO-ROI.

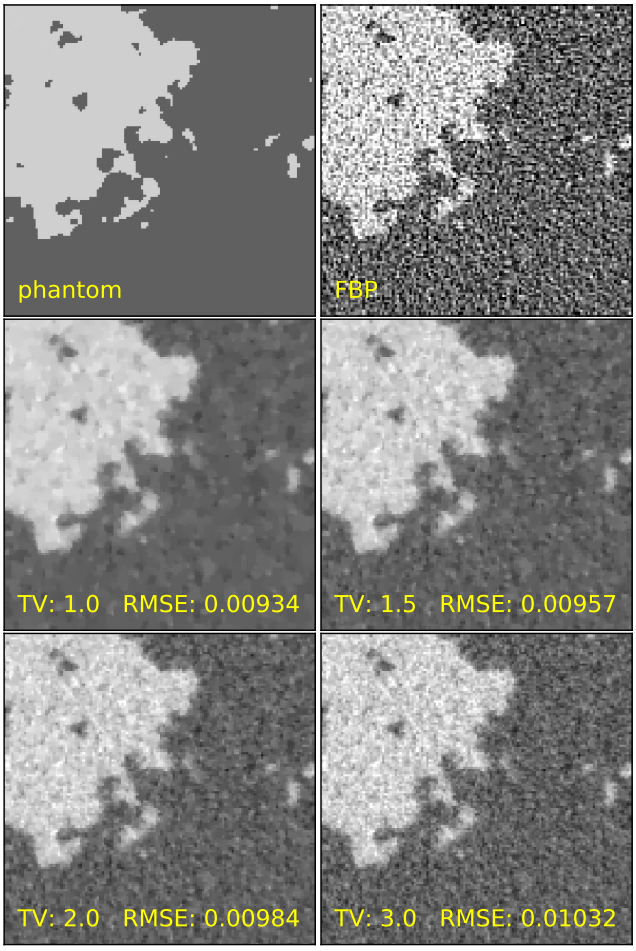


Fig. 1. Shown are 128x128 ROIs of reconstructed images of the breast CT simulation. In the top row is shown noiseless reconstruction by TVLSQ with $\gamma = \gamma_0$ and unregularized FBP, in order to gain a sense of the noise level. The remaining for panels show TVLSQ images for different values of γ. The regularization value shown in each panel is the TV constraint value divided by the TV of the phantom γ/γ_0. Also displayed in the bottom four panels is the image RMSE relative to the test phantom. The images are shown in a gray scale window [0.17,0.24] cm^{-1}.

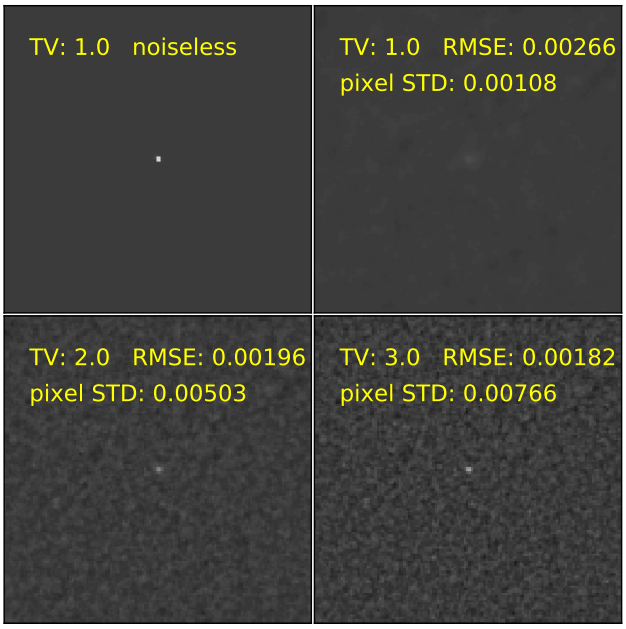


Fig. 2. Upper left panel shows a 128x128 central ROI of noiseless reconstruction of the signal by use of TVLSQ $\gamma = \gamma_0$. The other panels show the mean over 100 realizations of TVLSQ images for different values of γ/γ_0. The displayed quantities are RMSE between the mean of reconstructions and the noiseless reconstruction of the signal present image and the mean pixel standard deviation estimated from the 100 realizations. The gray scale for all panels is [0.1875,0.21] cm^{-1}.

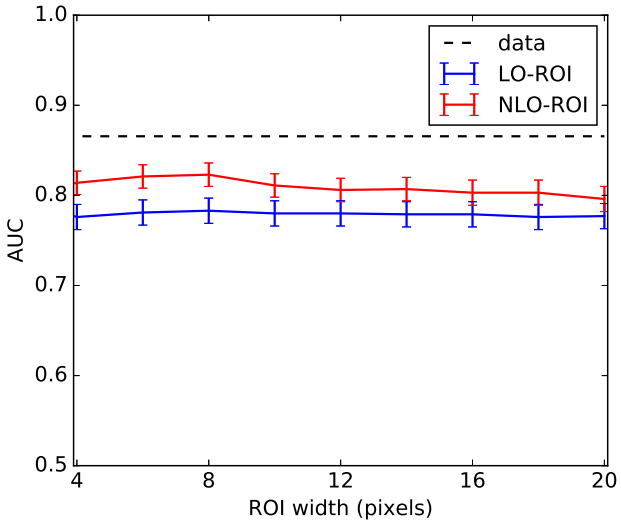


Fig. 3. Dependence of the LO-ROI and NLO-ROI observers as a function of ROI size. For reference the data domain AUC is indicated.

III. Results

The CT system parameters determining X for the present 2D breast CT simulation involves an acquisition of 128 projections on to a linear detector with 512 bins. The modeled incident intensity is 140,000 photons per pixel per view. The breast model contains glandular and adipose tissue and the breast diameter is 16 cm. Noiseless and noisy reconstruction of the test phantom is shown in Fig. 1 and for reference an unregularized FBP reconstruction from an equal-dose scan, 512 projections with 35,000 photons per pixel per view, is shown.

As might be expected, setting the TV constraint to the TV of the phantom $\gamma = \gamma_0$ yields the image with the lowest RMSE. Examining the subject appearance, however, it is not clear that $\gamma = \gamma_0$ is the most desirable setting due to the patchy noise artifacts. Loosening the TV constraint alleviates the patchiness and yields images that still show noise reduction relative to the FBP image.

To qauntify the noise texture of the TVLSQ reconstructions, we investigate small signal detection by generating 4000 signal-present and signal-absent images. We show mean reconstructed images of a signal-present image in Fig. 2, where the signal amplitude is set to a data domain SNR of 1.56 for the same noise conditions of Fig. 1. This SNR corresponds to an ideal observer 2-AFC performance of 86% correct in the data domain. From these images it is seen that the low γ settings distort the signal, decreasing the amplitude, and as γ is loosened the mean signal approaches that of the noiseless reconstruction. From the mean images, it is clear that the lowest $\gamma = \gamma_0$ setting distorts the signal significantly, potentially indicating a link between patchiness and small signal detectability.

For the LO-ROI and NLO-ROI model observers, we focus on the $\gamma = \gamma_0$ setting of the image TV, which corresponds to the most regularized setting shown in Figs. 1 and 2. As a first study, the impact of the ROI size is shown in Fig. 3. The image realizations are centrally cropped to reduce the rate at which overfitting happens. The neural networks for the observer models are trained on ROIs that have dimensions ranging from 4x4 to 20x20. The size of the ROIs influence the network topology and hence the number of weights to be trained in the neural networks, which affects the training behavior. Specifically, at small ROI sizes there is a slight upward trend in the AUC up to the dimensions 8x8. Increasing the size of the ROIs further causes model overfitting and their appears a slight decrease in AUC values. We settle on the 6x6 ROI size for the ROI observer models.

Apparent from these results is that there is a loss in the efficiency of the detection task from the data domain to the reconstructed image domain when using TVLSQ with $\gamma = \gamma_0$. This loss is more pronounced for LO-ROI than it is for NLO-ROI. That the NLO-ROI shows an efficiency loss is significant because this observer model represents the ceiling in performance for any observer. Thus at this setting of the regularization for TVLSQ there is an irretrievable loss of information pertaining to the ability to distinguish signal-present and signal-

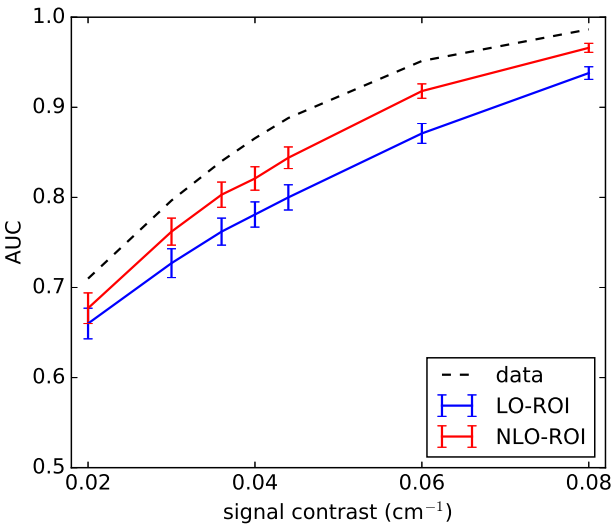


Fig. 4. Dependence of the LO-ROI and NLO-ROI observers as a function of the signal contrast in the object domain. For reference the data domain AUC is also plotted.

absent cases. We note also that this setting leads to the subjectively patchy appearance of the reconstructed image.

The gap between LO-ROI and NLO-ROI detection task performance results directly from the non-linearity of the image reconstruction algorithm. If the algorithm were linear, the Gaussian noise model used in the data domain simulation would transform to a Gaussian noise model with a transformed covariance in the image domain. Because the signal is small, the image domain covariance would be the same for signal-present and signal-absent cases; thus the ideal linear observer would be the ideal observer for linear image reconstruction.

We also investigate the model observer AUCs as a function of signal contrast, and the results are shown in Fig. 4. Again, TVLSQ is used for image reconstruction at the regularization setting $\gamma = \gamma_0$. The data domain ideal observer shows a monotonically increasing performance with signal contrast as expected. Both the LO-ROI and NLO-ROI image domain model observers show a decrease in AUC relative to the data domain AUC. The gap in observer performance seems to be greatest around a signal amplitude of 0.04, but the relative performances are quite similar across the range of signal amplitudes. From a practical point of view this result is useful for sweeping system or image reconstruction parameters. Because the signal amplitude does not play a large role in the relative performances of the shown observer models, it is not necessary to also perform a parameter search over signal contrast. We also note that the LO-ROI has a similar trend as the NLO-ROI trend. This is potentially significant because computation of the LO-ROI is simpler, requiring fewer image realization than the NLO-ROI.

IV. Summary

The use of signal detection theory with a weak signal may lead to a useful image quality metric for determining regularization strength in TVLSQ and non-linear image reconstruction in general. RMSE-based metrics alone appear to be insufficient as it suggests over-regularization in the noisy breast CT simulation. On the other hand, the image domain model observers for a detection task show a clear drop in performance relative to the data domain observer for the same image reconstruction parameter setting that yielded the best RMSE. As a result, these model observers may provide an objective numerical indication of when an image is over-regularized in non-linear image reconstruction.

At the meeting we will show results for various parameter sweeps such as regularization strength. We will also perform fixed-dose system parameter sweeps where the number of projections is varied but the total patient dose is held fixed. As is noted in the results sections the simpler LO-ROI model, which utilizes fewer parameters, may show similar trends as the more complex NLO-ROI model. We will also investigate approximate LO models using channelization in order to check if the trends persist and we can further improve the efficiency of observer computation by reducing the number model parameters and hence the number of necessary image realizations.

V. Acknowledgment

This work is supported in part by NIH Grant Nos. R01-EB026282, R01-EB023968, R01-EB020604, and R01-EB028652, and The University of Chicago Women's Board. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] E. Y. Sidky, Y. Duchin, X. Pan, and C. Ullberg, "A constrained, total-variation minimization algorithm for low-intensity x-ray CT," *Med. Phys.*, vol. 38, pp. S117–S125, 2011.

[2] H. H. Barrett and K. J. Myers, *Foundations of image science*, John Wiley & Sons, 2013.

[3] A. A. Sanchez, E. Y. Sidky, and X. Pan, "Task-based optimization of dedicated breast CT via hotelling observer metrics," *Med. Phys.*, vol. 41, 2014, art. no. 101917.

[4] W. Zhou, H. Li, and M. Anastasio, "Approximating the ideal observer and hotelling observer for binary signal detection tasks by use of supervised learning methods," *IEEE Trans. Med. Imaging*, vol. 38, pp. 2456–2468, 2019.

[5] L. L. Pesce and C. E. Metz, "Reliable and computationally efficient maximum-likelihood estimation of "proper" binormal ROC curves," *Acad. Radiol.*, vol. 14, pp. 814–829, 2007.

Objective assessment of clinical image quality in thorax CT: nodule detectability in a 3D printed anthropomorphic phantom by a 3D model observer.

Irene Hernandez-Giron, Johan Michiel den Harder, Geert J. Streekstra, and Wouter J. H. Veldkamp

***Abstract*—There is a chasm between the results obtained with geometric phantoms combined with traditional image quality metrics, and the performance of diagnostic tasks carried out by radiologists on CT patient images. We propose a CT image quality assessment framework combining diagnostic task-based metrics and anthropomorphic phantoms as a representation of the patients' anatomy and disease presence or stage.**

Our goals were: (1) to create an anthropomorphic thorax phantom with two realistic 3D printed lung vessel distribution inserts, one of them containing nodules (25 cork spheres, 6mm diameter); and (2) to apply a 3D-model observer to assess the detectability of nodules as a function of dose and reconstruction parameters.

Several phantom scans (120 scans) were performed (Canon Aquilion ONE Genesis, ultra-low dose thorax protocol), reconstructed with filtered-back projection (FBP) and AIDR3De. The detectability (d') of the nodules was assessed by a 3D-Channelized Hotelling model observer (DDOG channels), analyzing volumes of interest containing either individual lesions or equivalent regions in the lesion-free lung. The detectability of the lesions increased with dose for both reconstructions and was higher with iterative reconstruction (AIDR3De) compared to FBP for all dose levels. This approach has the potential to be used in CT protocol optimization, providing a task-based objective method to assess image quality closer to the actual practice of radiologists.

***Index Terms*— Computed tomography, iterative reconstruction, image quality, phantom, 3D printing, model observer.**

I. Introduction

COMPUTED tomography (CT) has a strong presence and usage in Radiology departments worldwide. The broad range of clinical CT applications and the availability of these devices, has turned CT into the biggest contributor to the collective radiation dose among medical imaging modalities [1]. The optimization of the imaging CT protocols is a critical and complex task, due to the extensive set of acquisition and reconstruction parameters (which in turn can usually vary in a range of values) with impact on dose and image quality. Each protocol should be adapted to the clinical indication, anatomical region and patient characteristics to produce at the lowest possible dose, images with sufficient quality to render a diagnosis. The evaluation of this "sufficient" diagnostic image quality level is not straightforward from the medical physics point of view. This is frequently based on traditional metrics for image quality such as noise, contrast or spatial resolution, among others, applied to CT images of geometric phantoms. These, contain, in a uniform background, distributions of test objects, and lack in representing the anatomy, attenuation or texture of patient structures or tissues [2]. This makes it challenging to translate the obtained results with such a framework to the clinical scenario, where radiologists score patient images, and determine the required image quality level for a given diagnostic task.

Task-based image quality assessment, rests on the idea of defining image quality taking into account the relevant diagnostic objective (task) of the image, for instance lung nodule detection [3,4]. The "observer" performing this task can be a human (radiologist), a Computer Aided Detection (CAD) algorithm or a so-called model observer. A figure of merit (FOM) that is adequate for the selected task and imaging modality (detection or discrimination indexes, for example) is defined. Model observers have been successfully applied to simple detection and discrimination tasks, such as low contrast detectability in CT or microcalcification detection in mammography in simple phantoms [5-7].

There is a need for new image quality phantoms in CT which incorporate the influence of anatomical backgrounds and simulate the patient shape and attenuation. Some anatomically realistic phantoms are available commercially but their use is not widely spread due to their high cost [8]. The use of 3D printing for customized image quality phantoms in medical imaging has grown in the past few years as an affordable alternative [9-11].

This work was supported by the NWO Foundation, with the project CLUES (Clinical Image Quality Assessment, project nr 13592, http://clues-iq.blogspot.com/p/main-page.html) and the project "Through the eyes of AI" (Veni grant 2020, project nr 17378).

I. Hernandez-Giron is with the Radiology Department at the Leiden University Medical Center (LUMC), Leiden, The Netherlands (I.Hernandez_Giron@lumc.nl)

J. M. den Harder is with the Amsterdam Academic Medical Center (AMC), Amsterdam, The Netherlands (c.denharder@amc.uva.nl)

G. J. Streekstra is with the Amsterdam Academic Medical Center (AMC), Amsterdam, The Netherlands (g.j.streekstra@amc.uva.nl)

W. J. H. Veldkamp is with the Radiology Department at the LUMC (W.J.H.Veldkamp@lumc.nl)

The aim of this study was to develop a framework for the assessment of clinical image quality in thorax CT, for the task of lung nodule detectability, combining model observers and 3D printed anthropomorphic phantoms. The effect of dose and reconstruction method on lesion detectability was analyzed with a channelized Hotelling model observer in CT images of a thorax phantom, containing two 3D-printed lung vessel inserts and lesions.

II. MATERIALS AND METHODS

A. Phantom design and 3D printing

Software was developed in MATLAB® (MathWorks®, Natick, MA) to simulate the distribution and sizes of the lung vessels inside an elliptically shaped structure with a given height. A seed point was placed centrally on one of the inner sides of the ellipse. The created algorithm placed two massive cylinders of 8.5 mm diameter at the seed point to represent the parent vessels. The end point of each vessel was used as a starting point for either one or two new vessels. Inspired on the work by Weibel et al, the algorithm was designed as follows [12-14]: at a bifurcation, the diameter of the vessel was reduced by a constant factor and so did the vessel length, taking care to avoid that bifurcations follow each other too closely. Thus, a 3D distribution of the vessels was created up to a preset maximum number of vessels, resulting in vessel diameters down to 0.25 mm. Binary output files ('1' for vessels and '0' for surrounding air) were produced in DICOM format and later transformed into Standard Tessellation Language (STL) files using two freeware packages: 3D slicer and Meshlab (**Fig. 1**). A mirrored version of the original lung vessel phantom file was also generated.

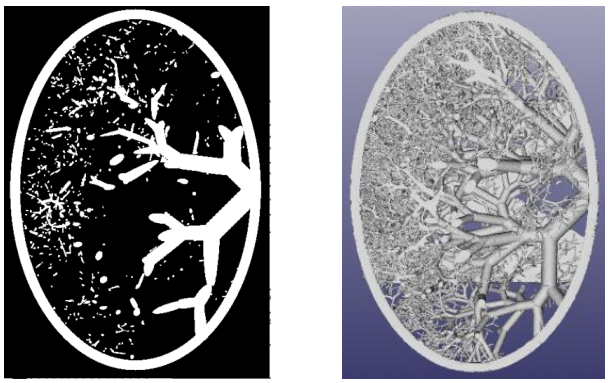


Figure 1. Matlab binary image with the vessel distribution in the lung phantom (left) together with the generated STL file for the 3D printer visualized using 3D slicer (right).

The STL files (two lung vessel distributions, one to mimic the left lung, other the right one; 15×10.5×2.5 cm) were printed in a ProJet 3D printer (multi-jet modeling (MJM) technique in ultrahigh definition mode (UHD, 32 µm resolution). The printing material for the vessels was Visijet EX200. As a surrogate for the patient's thorax in terms of attenuation and shape, a PMMA elliptical holder was designed and manufactured (30×20×2.5cm), containing two elliptical gaps for the 3D printed lung vessel phantoms and a small insert of Teflon as a surrogate of the spine [14]. As surrogates of lung nodules, 25 cork spheres (6 mm diameter) were attached with double-sided tape to different positions on one of the lung inserts (**Fig. 2**). Thus, in the phantom configuration, one of the lung inserts (which were exact mirrored copies), contained lesions and the other was lesion free (Fig. 2A).

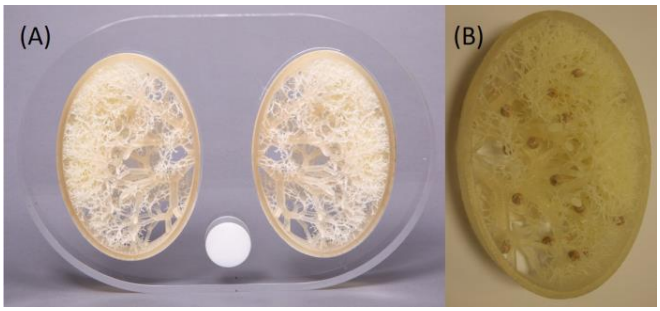


Figure 2. Photo of the thorax holder (cast on PMMA and with a small Teflon insert), containing both 3D printed lung vessel phantoms, both without lesions (2A). Photo of the vessel insert with the cork spheres (6 mm diameter) attached as surrogates for lung nodules (2B).

B. CT scans

The thorax phantom, with the two lung inserts and lesions, was scanned 120 times for different dose levels, in a Aquilion ONE Genesis CT system, selecting a ultra-low dose thorax protocol [15]: 120 kV, 40x0.5mm collimation, pitch=1.475, 0.275 s, varying mA (10-20-40-80-150 mA; CTDIvol=0.1-0.3-0.5-1.1-2 mGy), 1mm slice thickness and interval, FC08 kernel, filtered back-projection (FBP) and AIDR3De, respectively. **Figure 3** shows one of the slices of the CT scans with some of the lesions highlighted in green.

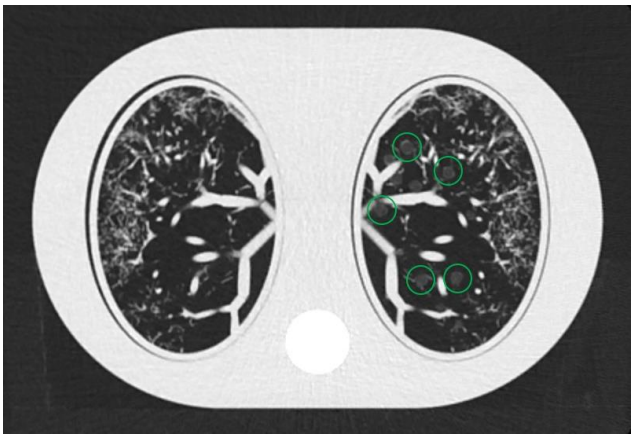


Figure 3. Cropped CT image of the thorax phantom and the two lung vessel inserts, with lesions on the right insert (some of them highlighted with green circles) and, without lesions, on the left ($CTDI_{vol}$= 2 mGy, AIDR3De).

C. Sample subtraction

The median of 10 high dose phantom scans (600 mAs) with the same protocol, acquired on the same session was calculated. This was used as a template for a semi-automated method for the detection of the locations of the nodules, which determined the central slice and central position of each nodule [16]. Then, the mirrored central locations on the other lung phantom were automatically derived. Based on those coordinates, volumes of interest (VOIs, 24x24 px x 9 slices; 18.77x18.77 mm^2 x 9 slices) were automatically subtracted, containing the lesions and surrounding vessel structures, and also the equivalent lesion-free counterpart volumes.

Figure 4 shows the central slice for 5 nodule VOIs and the equivalent lesion-free volumes (after mirroring) for one dose level and both reconstruction methods. From each of the 120 scans, 50 VOIs were subtracted (25 nodule present, 25 nodule absent), which led to a total of 3000 VOIs per image class (per dose and reconstruction) to be analyzed with the model observer.

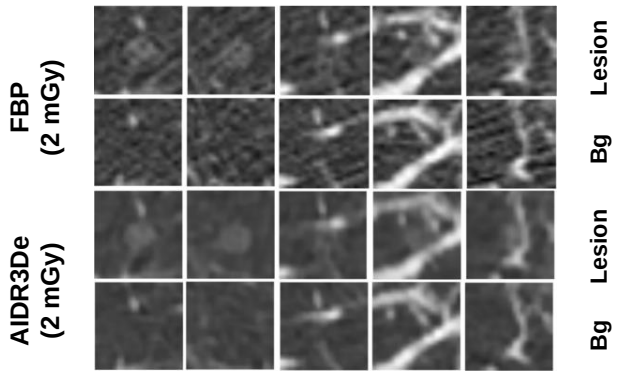


Figure 4. Examples with lesion present and lesion absent (anatomical background, Bg). The latter were flipped or mirrored (XY) direction, as shown, before being analyzed by the 3D model observer.

D. 3D model observer

A multi-slice model observer proposed by Platisa et al was adapted for this study, which computes the 2D responses of a channelized Hotelling model observer in a range of slices, combining them to get a global response or detectability [17]. Ba et al, successfully applied such a model observer to detect low contrast spheres (of a similar size as the nodules used in this study), in an abdominal phantom with a uniform background [18]. According to Platisa et al, this multi-slice model observer emulates how human observers evaluate a stack of images in two phases: first the observer buffers the scores slice by slice during scrolling, then for the z-direction, those scores are processed as a test-statistic for the full stack of images [17]. This value is in turn used to classify the stack between lesion present or absent categories. An adapted multi-slice model observer with 10 difference of Gaussian (DDOG) channels [18] was applied to the two classes of images (lesion present/lesion absent) and one template was estimated per nodule (based on the central slice per lesion) and applied for all the slices, avoiding when possible the influence of nearby vessel structures.

Test statistics (such as the mean (<T>) and the standard deviation (σ) of the two classes of images (lesion present (1), lesion absent (2)) after being transformed applying the model observer template are calculated and combined to obtain a detectability index or d' as follows:

$$d' = \frac{\langle T \rangle_1 - \langle T \rangle_2}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2}}$$

The higher the detectability index, implies that the statistical distributions of both classes of images are more separated and thus easier to differentiate.

For training-testing, detectability (d') was estimated for groups of 20 scans (training set different than testing set) without repetition, and as error, 2σ of the 6 values for d' obtained for these image subgroups was calculated. The individual d' for each lesion was also calculated to estimate the influence of the surrounding background or the cork composition into detectability.

III. RESULTS

The attenuation of the material used as a surrogate for the lung vessels was (102±22 HU), for the thorax PMMA holder (118±10 HU) and for the nodule surrogates (cork, measured at the central slices) it was (-845±20 HU). The values for the phantom materials are slightly higher as those found in patients, but in the order of other thorax phantoms [10-11-14]. The effect of dose and reconstruction on some of the nodules visibility in the phantom can be seen in **Figures 4 and 5**.

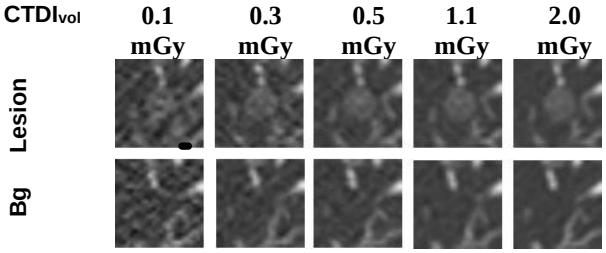


Figure 5. CT image crops, showing the central slice for one of the lesions and their equivalent lesion absent (Bg) slices in the other lung phantom for all the dose levels and filtered back projection reconstruction.

Visually, image quality (in terms of the rendering of the lung nodules) is improved with dose (increasing mAs) and also with iterative reconstruction compared to FBP (**Figs. 4 and 5**).

The results of the model observer, comparing both reconstruction methods for one dose level are shown in **Fig. 6**, per lesion. It can be appreciated that the model observer obtained higher d' values for AIDR3De compared to FBP for all lesions. It also has to be noted the variability in d' between lesions, that can be due to their composition (natural cork), as they were not all exactly the same and could have some small differences in porosity.

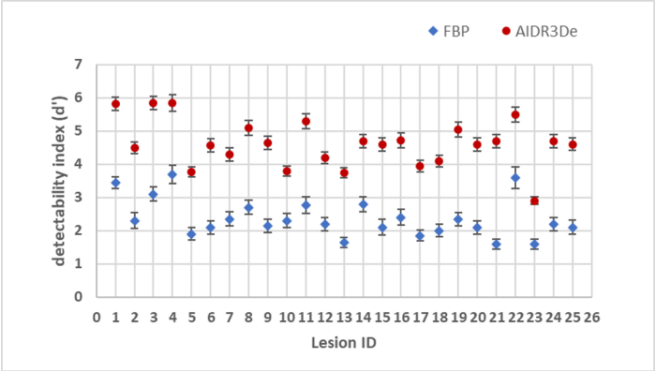


Fig. 6. Detectability index (d') vs dose for FBP and AIDR3De reconstructions (CTDIvol=0.3 mGy), for each lesion individually. Each d' represents the average of 6 runs, using each 1 object × 20 acquisitions for training and 1 object × 20 acquisitions for testing per reconstruction. The error bars correspond to 2σ.

The global results of the DDOG multi-slice model observer comparing both reconstruction methods and all dose levels are summarized in **Fig. 7**.

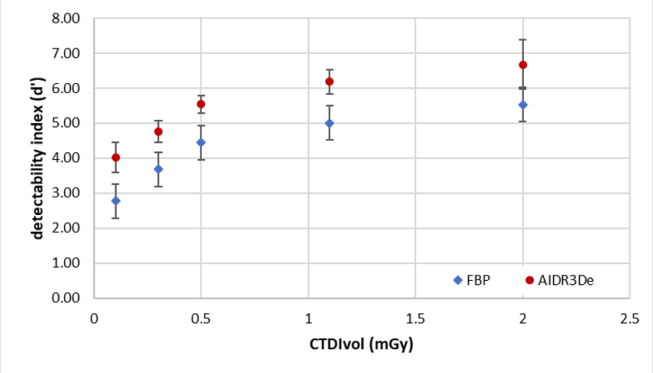


Fig. 7. Detectability index (d') vs dose for FBP and AIDR3De reconstructions, using 25 objects × 20 acquisitions for training and 25 objects × 20 acquisitions for testing per dose level and reconstruction, repeated 6 times. The error bars correspond to 2σ of the d' obtained per dose level.

In terms of the effect of dose on d', for FBP, the relative increase of d' in the considered dose range (CTDIvol = 0.1–0.3–0.5–1.1–2.0 mGy) was (~33%–21%–13%–10%), respectively. Whereas for iterative reconstruction (AIDR3De) it was (~19%–16%–11%–8%). Comparing both reconstructions, with AIDR3De, d' was ~45%–30%–25% –24% – 21% higher than with FBP.

IV. DISCUSSION

We presented an anthropomorphic lung vessel phantom with attachable lesions to be used in combination with objective task-based image quality metrics. With the proposed multi-slice model observer, differences in image quality were determined with an improvement in nodule detectability as dose increased, with a bigger effect for FBP compared to iterative reconstruction. For all the nodules, the detectability was higher with iterative reconstruction (AIDR3De) compared to FBP.

There is still room for improvement in the phantom, as the attenuation of the material used for the lung vessels is slightly higher (but in the order) of human tissue. In the future, the phantom will be combined with parenchyma-like structures, as the lack of this tissue makes the lesions more conspicuous than in patients [19]. This limitation is shared by all available commercial anthropomorphic phantoms and also other 3D printed lung vessel phantoms [8, 10,11,14].

The proposed multi-slice model observer was able to detect nodules in CT images of a thorax phantom with 3D printed lung vessel inserts in a semi-automated manner. In the near future the results of the model observer will be compared to the performance of human observers, leading to the optimization of the channels type and fine-tuning.

V. CONCLUSIONS

A framework based on the combination of 3D printed lung vessel phantoms containing lesions and a model observer has been successfully applied to analyze the detectability of lung nodules in clinical CT protocols. This approach has the potential to be used in CT protocol optimization, giving a task-based objective method to assess image quality closer to the actual practice of radiologists.

REFERENCES

[1] Tack D, Kalra MK, Genevois PA, editors. Radiation dose from multidetector CT, 2nd ed. Springer-Verlag; 2012.
[2] Vaishnav JY, Jung WC, Popescu L, Zeng R, Myers KJ. Objective assessment of image quality and dose reduction in CT iterative reconstruction. Med Phys 2014:41,071904.
[3] Verdun FR, Racine D, Ott JG, Tapiovaara MJ, Toroi P, Bochud FO, et al. Image quality in CT: From physical measurements to model observers. Physica Medica 31 (2015):823-843.
[4] Barrett HH, Myers KJ, Hoeschen C, Kupinski MA, Little MP. Task-based measures of image quality and their relation to radiation dose and patient risk. Phys Med Biol 2015;60(2):R1-R75.
[5] Hernandez-Giron I, Calzado A, Geleijns J, Joemai RMS, Veldkamp WJH. Comparison between human and model observer performance in low-contrast detection tasks in CT images: application to images reconstructed with filtered back projection and iterative algorithms. Br J Radiol 2014;87:20140014.
[6] Ba A, Abbey CK, Baek J, Han M, Bouwman RW, Balta C, et al. Inter-laboratory comparison of channelized hoteling observer computation. Med Phys2018; 45(7):3019-3030.
[7] Balta C, Bouwman RW, Sechopoulos I, Broeders MJM, Karssemeijer N, van Engen RW, Veldkamp WJH. A model observer study using acquired mammographic images of an anthropomorphic breast phantom. Med Phys 2018;45(2):655-665.
[8] https://www.kyotokagaku.com/products/detail03/ph-1.html
[9] Filippou V, Tsoumpas C. Recent advances on the development of phantoms using 3D printing for imaging with CT, MRI, PET, SPECT, and ultrasound. Med Phys 2018;45(9):e740-e760
[10] Solomon J, Bochud F, Samei E. Design of anthropomorphic textured phantoms for CT performance evaluation. Proc SPIE Medical imaging 2014. Vol 9033:90331U.
[11] Solomon J, Samei E. Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE. Med Phys 2014; 41(9):091908-1.
[12] Weibel ER. What makes a good lung? Swiss Med Wkl 2009;139 (27–28):375–386
[13] Weibel ER, Gomez DM. Architecture of the human lung. Use of quantitative methods establishes fundamental relations between size and number of lung structures. Science 1962;137(3530):577-85.
[14] Hernandez-Giron I, den Harder JM, Streekstra GJ, Geleijns J, Veldkamp WJH. Development of a 3D printed anthropomorphic lung phantom for iamge quality assessment in CT. Med Phys 2019;57:47-57.
[15] Kroft LJM, van der Velden L, Hernandez-Giron I, Roelofs JJH, de Roos A, Geleijns J. Added value of ultra-low-dose Computed Tomography, dose equivalent to chest X-ray radiography, for diagnosing chest pathology. Journal of Thoracic Imaging 2019;34(3):179-186.
[16] Veldkamp WJH, van der Vaart J, Streekstra GJ, Hernandez-Giron I. An automated method to assess clinical image quality using FROC analysis for nodule detection in simulated lung phantom images. Presented at: European Conference of Radiology (ECR), Vienna 2017 (C-3036). http://dx.doi.org/10.1594/ecr2017/C-3036.
[17] Platisa L, Goossens B, Vansteenkiste E, Park S, Gallas BD, Badano A, Philips W. Channelized Hotelling observers for the assessment of volumetric imaging data sets. J Opt Soc Am A Opt Image Sci Vis 2011;28(6):1145-63.
[18] Ba A, Eckstein MP, Racine D, Ott JG, Verdun F, Kobbe-Schmidt S, Bochud FO. Anthropomorphic model observer performance in three-dimensional detection task for low-contrast computed tomography. Journal of medical imaging 2016;3(1):011009
[19] van Hooff S, Tarrahi I, Veldkamp WJH, den Harder JM, Streestra GJ, Geleijns J, Hernandez-Giron I. Innovative methodology to mimic lung parenchyma based on Voronoi models: application to 3D-printed anthropomorphic CT image quality phantoms. Presented at RSNA (2019).

Training Strategies of a Convolutional Neural Network-based Anthropomorphic Model Observer for Signal-known-exactly and Background-known-statistically Detection Tasks

Byeongjoon Kim, Minah Han, and Jongduk Baek

***Abstract*—We implement anthropomorphic model observer using a convolutional neural network (CNN) for signal-known-exactly and background-known-statistically detection tasks on breast cone beam computed tomography images. Because labeling large amounts of training data by human observer is very time-consuming and expensive, we propose training strategies (data labeling method and early stopping rule) for CNN-based model observer. For data labeling, we use nonprewhitening observer with an eye-filter as a surrogate for human observer. CNN with proposed training strategies closely predicts human observer performance while reducing the cost of a human observer study.**

***Index Terms*—Anthropomorphic model observer, Convolutional neural network, Nonprewhitening observer with an eye-filter.**

I. Introduction

In X-ray computed tomography (CT) imaging, optimizing the imaging parameters is essential to improve the diagnostic performance. Task-based assessment has been considered as a thorough image quality evaluation method because it directly quantifies diagnostic accuaracy (e.g., lesion detectability) of given images [1]–[3]. To maximize the detection performance, various imaging parameters should be examined for the given tasks, and evaluated by the human observer. However, human observer study is time-consuming and expensive.

To overcome this limitation, anthropomorphic model observers have been developed (e.g., nonprewhitening observer with an eye-filter (NPWE) and channelized Hotelling observer (CHO) [4]–[6]) by mimicking the frequency selective sensitivity of the human visual system using frequency filters [7], [8] and reflecting human inefficiency using internal noise [9]. They have shown good correlation with human observer performance [4]–[6], but the proper filter design and internal noise should be determined for each imaging task (e.g., signal shape and background noise structure) [10].

In recent studies, a convolutional neural network (CNN) has been used to implement model observers [11], [12]. The main advantage of using CNNs is that we do not have to manually find task-specific design of filter or internal noise model because CNNs can be trained in an end-to-end fashion. CNNs also have strong representational power to learn nonlinear functions, which makes it possible to develop a single model that handles multiple tasks. To implement CNN-based anthropomorphic model observers, it is necessary to reflect the human inefficiency in CNNs. The most popular approach is to manually label images with a decision or confidence level of a human observer for signal presence or absence. However, since CNNs require a large amount of data for network training, this approach would be exhaustive and time-consuming.

In this work, we propose new training strategies that can replace human-labeled data. There are two key components in our training strategies: data labeling and early stopping. We obtain percent correct (P_c) values of human observers on a subset of training images and automatically label the rest of images using the conventional model observer. To better match the performance of CNN-based anthropomorphic model observer to that of human observer, we introduced the early stopping technique. Observer study was conducted on signal-known-exactly (SKE) and background-known-statistically (BKS) four alternative forced choice (4AFC) detection tasks in cone beam CT (CBCT) images with breast anatomical backgrounds. The performance agreement between the human and model observers was measured using the difference in and correlation between the P_c values of the human and model observers.

B. Kim, M. Han, and J. Baek are with the School of Integrated Technology and Yonsei Institute of Convergence Technology, Yonsei University, 162-1, Songdo-dong, Yeonsu-gu, Incheon, South Korea (e-mail: bjkim2006@naver.com, hma90@naver.com, jongdukbaek@yonsei.ac.kr).

II. Methods

A. Image Generation

For the breast background volume, we generated 512×512×512 voxels of Gaussian random noise volume and filtered the volume using $1/f^{3/2}$ kernel. Note that we used breast anatomy structure in mammograms which has been characterized by $P(f) = K/f^3$ [13], where K is a scaling factor. The central sphere with 128 voxel diameter was extracted to prevent the wrap-around effect. We set the upper 50% (lower 50%) voxel values to the attenuation coefficient of glandular (adipose) tissue to generate a 50% volume glandular fraction (VGF). The attenuation coefficients of glandular and adipose tissues are 0.233 and 0.194 cm^{-1} at 50 keV, respectively [14].

For modeling breast mass, we inserted 2 mm diameter sphere near the center of the generated breast background volume. We set the voxel values of the sphere to the attenuation coefficient of breast mass (0.238 cm^{-1} at 50 keV [14]).

We used Siddon's algorithm [15] to acquire 200 CBCT projection data, and then, added log-normalized uniform Poission noise with 6914 photons/detector pixel [16]. Simulation parameters are summarized in Table I. We reconstructed the projection data using the FDK algorithm [17], where Hanning, Shepp-Logan, and Ram-Lak weighted ramp filters were used as the reconstruction filters. We filtered projection data with and without Fourier interpolation and performed voxel-driven back-projection with linear interpolation. For observer study, we used the central transverse (i.e., x-y plane) and longitudinal (i.e., x-z plane) images and extracted central 129×129 pixels. Consequently, 12 types of background noise structures were generated (Table II) and corresponding signal-present images are shown in Fig. 1. We generated 4000 pairs of signal-present and signal-absent images for each noise structure.

TABLE I. CBCT simulation parameter values

Parameter	Value
Source to iso-center distance	460 mm
Source to detector distance	880 mm
Detector pixel size	0.388×0.388 mm^2
Number of projection data	200
Reconstructed voxel size	0.1014×0.1014×0.1014 mm^3

TABLE II. Background noise structures

Noise	Reconstruction filter	Interpolation method	Image plane
N1	Hanning	Linear	Transverse
N2	Shepp-Logan		
N3	Ram-Lak		
N4	Hanning	Fourier	
N5	Shepp-Logan		
N6	Ram-Lak		
N7	Hanning	Linear	Longitudinal
N8	Shepp-Logan		
N9	Ram-Lak		
N10	Hanning	Fourier	
N11	Shepp-Logan		
N12	Ram-Lak		

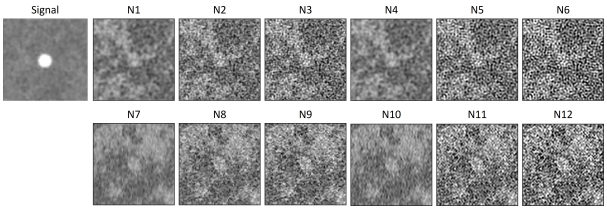


Fig. 1. Example signal-present images with different noise structures.

B. Detection Task

We conducted SKE/BKS 4AFC detection tasks. Hypotheses for signal-absent (H_0) and signal-present (H_1) images are given by

$$H_0 : \mathbf{g} = \mathbf{f_b} + \mathbf{f_n} \tag{1}$$

$$H_1 : \mathbf{g} = \mathbf{f_s} + \mathbf{f_n} \tag{2}$$

where $\mathbf{g}$ is the observed image, $\mathbf{f_b}$ is the breast background, $\mathbf{f_s}$ is the breast background with the signal, and $\mathbf{f_n}$ is the reconstructed CT noise.

C. Human Observer Study

Seven human observers performed 12 detection tasks with different background noise structures (Table II). For each detection task, training session consisted of 30 trials and testing session consisted of 100 trials. In each trial, four images were displayed and observers were asked to choose the signal-present image among 1 signal-present and 3 signal-absent images. Human observer performance was quantified by

$$P_c = \frac{1}{N_t}\sum_{j=1}^{N_t} o_j \tag{3}$$

where N_t is the number of trials in testing session and o_j is score of $j-th$ trial, 1 for correct answer and 0 for incorrect answer. We estimate the variance of P_c by bootstrapping scores 1,000 times.

D. Data labeling using the anthropomorphic model observer

To label the data for network training, we propose new labeling strategies using the NPWE model for the generated signal-present and signal-absent images. The eye-filter function is defined by

$$E(f) = f^{1.3}exp(-cf^2) \tag{4}$$

where c is an eye-filter parameter. The template of NPWE is defined by

$$\mathbf{w} = F^{-1}\big\{\mathbf{E}\cdot F\big\{\mathbf{\Delta g}\big\}\big\} \tag{5}$$

where $F\big\{\cdot\big\}$ and $F^{-1}\big\{\cdot\big\}$ indicate DFT and inverse DFT operators, respectively, and $\Delta\mathbf{g} = E[\mathbf{g} \mid H_1] - E[\mathbf{g} \mid H_0]$. We estimated the template $\mathbf{w}$ using 400 image pairs. Decision variable of test image is computed by

$$t = \mathbf{w}^t F^{-1}\big\{\mathbf{E}\cdot F\big\{\mathbf{g}\big\}\big\} + \epsilon. \tag{6}$$

ϵ is the internal noise term sampled from normal distribution of $N(0,a)$, where a is an internal noise level.

NPWE does not predict the exact human observer performance for all noise structures with the same value of a or c. Therefore, we adjusted those values for each noise structure for data labeling. When the value of a was adjusted with $c = 2$, we denoted the model observer by NPWE4i. When the value of c was adjusted with $a = 0$, we denoted the model observer by NPWEf. Note that $c = 2$ yielded the peak value of eye-filter at 4 cyc/deg for 40 cm viewing distance, which is the most sensitive frequency of the human visual system [6], [7].

We used two new labeling strategies:

(1) We adjusted the value of a to minimize the difference in P_c between NPWE4i and human observer for each background noise structure, and then, assigned decision variable of NPWE4i as label value.
(2) We adjusted the value of c to minimize the difference in P_c between NPWEf and human observer for each background noise structure, and then, assigned decision variable of NPWEf as label value.

TABLE III. Values of a and c used for labeling the data.

Noise structures		N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12
NPWE4i	$a \times 10^2$	0.115	0.15	0.165	0.09	0.14	0.185	0.08	0.1	0.13	0.045	0.15	0.2
NPWEf	c	0.3	0.52	0.54	0.61	0.7	0.63	0.25	0.45	0.45	0.54	0.56	0.6

E. CNN-based Anthropomorphic Model Observer

We implemented 3-layer CNN that had two convolutional layers and one fully connected layers. The convolutional layers had 13×13 filters with a stride of 2, and the number of filters was two at the first layer and doubled if the feature map size was halved to preserve the time complexity per layer. The size of filters was chosen to match the receptive field seen at the last convolution layer to the signal size, 37×37. After each convolution layers, the feature maps were activated with rectified linear unit. We used mean squared error in decision variables as loss function.

$$L = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{t}^{(i)} - t^{(i)}\right)^2, \tag{7}$$

where N is the number of mini-batch and $\hat{t}^{(i)}$ and $t^{(i)}$ are the label and the output of model, respectively, on the i-th image.

F. CNN Training and Testing

Each dataset of the same noise structure were split into training, validation, and testing sets with the ratio of 2:1:1. Then, the mean value of each image was subtracted from each image. CNN was initialized with the Glorot Uniform method and optimized by the Adam algorithm [18] with learning rate of 10^{-3}, β_1 of 0.9, and β_2 of 0.999. The maximum number of training epochs was 300 (18,900 iterations).

Model observer performance was computed using (3) with 10,000 trials in each noise structure. To evaluate the agreement between model and human observers, we measured the root mean squared error (RMSE) and correlation coefficient in P_c.

CNNs trained with NPWE4i decision variable and NPWEf decision variable are denoted as CNN-NPWE4i and CNN-NPWEf, respectively.

G. Stopping Criteria

As a form of regularization, one can terminate CNN training early when a stopping criterion is met. We propose to employ the early stopping rule that directly monitored the RMSE in P_c on the validation set. The RMSE is computed as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{P}_c^{(k)} - P_c^{(k)})^2}, \tag{8}$$

where K is the number of different noise structures in the training set, and $P_c^{(k)}$ and $\hat{P}_c^{(k)}$ are P_c values of human and CNN observers, respectively, on the k-th noise structure. Here, we used the same $P_c^{(k)}$ values used to label the data. For simplicity, we implemented this method by measuring RMSE in P_c after every epoch and saving the model only if it made a new record. For comparison, other early stopping methods based on the validation loss and training loss, respectively, were used.

III. Results

Fig. 2 compares P_c of the human and model observers with different labeling methods and early stopping rules. Note that for labeling the training set of CNN, we used different a and c values for each noise structure, as summarized in Table III. The RMSE and correlation coefficient in P_c between human and model observers are summarized in Table IV.

With the conventional early stopping rules, CNN-NPWE4i and CNN-NPWEf showed similar performance trends with human observers. This indicates that the proposed labeling methods function as internal noise. However, their performance were not matched with that of human observers for all tasks.

With the proposed early stopping rule, the CNNs produced similar P_c to human observers, indicating that proposed early stopping rule is more effective to predict human observer performance.

Fig. 3 shows learning curves of CNN-based anthropomorphic model observer. It is observed that validation loss and validation RMSE in P_c starts to increase after a certain point, indicating overfitting of CNN. It is also observed that the validation loss, training loss, and validation RMSE in P_c show different convergence behavior. For example, the lower validation or training loss does not always produce the lower validation RMSE in P_c.

IV. Conclusion

In this work, we proposed new training strategies for CNN-based anthropomorphic model observers on SKE/BKS detection tasks in CBCT images with breast anatomical backgrounds. CNN with the proposed training strategies closely predicted human observer performance while reducing the cost of the human observer study. The proposed training strategies can potentially be used for implementing one CNN model that generalizes to various detection tasks. The limitation of this work is that it has not fully investigated the generalization errors and CNN architectures, which will be addressed in our future research.

Acknowledgment

This research was supported by National Research Foundation of Korea (2019R1A2C2084936, 2018M3A9H6081483).

References

[1] H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," *J. Opt. Soc. Am. A.*, vol. 7, no. 7, pp. 1266–1278, 1990.

[2] P. Timberg, M. Båth, I. Andersson, S. Mattsson, A. Tingberg, and M. Ruschin, "Visibility of microcalcification clusters and masses in breast tomosynthesis image volumes and digital mammography: a 4afc human observer study," *Medical physics*, vol. 39, no. 5, pp. 2431–2437, 2012.

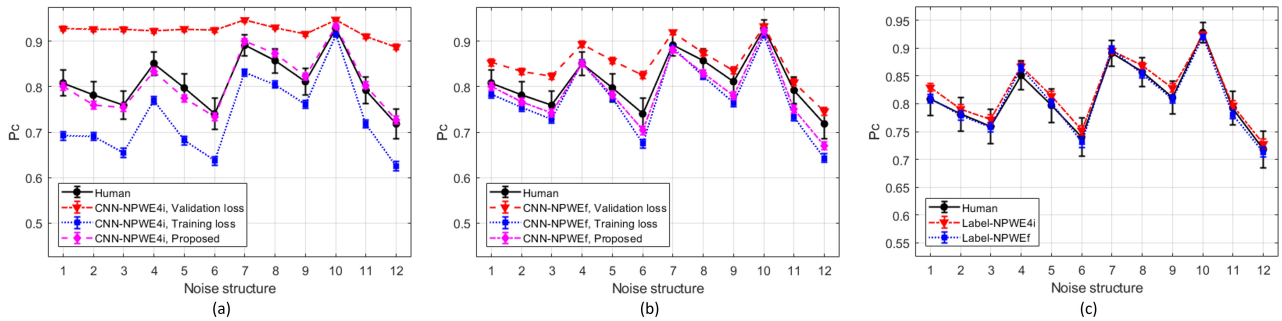


Fig. 2. Comparison of human observer and CNN-based model observer P_c with 95% confidence intervals for different stopping rules. (a) CNN-NPWE4i, (b) CNN-NPWEf, and (c) NPWE4i and NPWEf labels of the testing sets.

TABLE IV. RMSE and correlation coefficient in P_c between human observer and CNN-based model observer for different stopping rules and data labeling methods.

Stopping rule	Validation loss		Training loss		Validation RMSE in P_c	
Label	NPWE4i	NPWEf	NPWE4i	NPWEf	NPWE4i	NPWEf
RMSE	0.1230	0.0450	0.0850	0.0406	0.0130	0.0257
Pearson's corr.	0.7692	0.9319	0.9710	0.9801	0.9800	0.9868

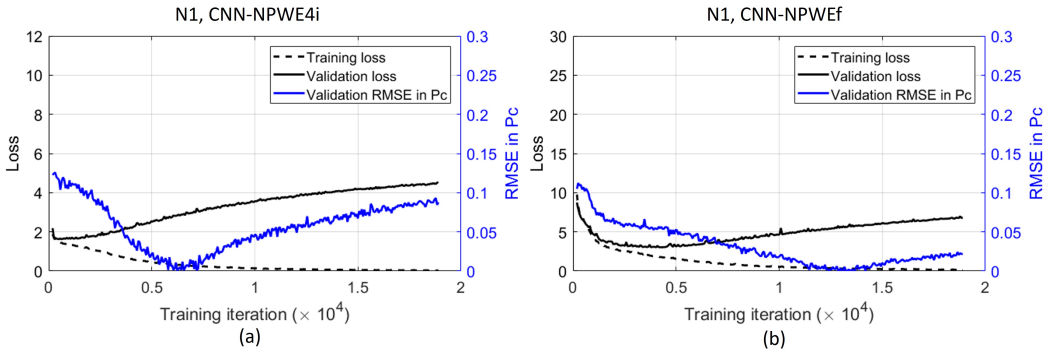


Fig. 3. Validation loss, training loss, and validation RMSE in P_c during CNN training with noise structure N1 for different labeling methods: (a) CNN-NPWE4i and (b) CNN-NPWEf.

[3] L. Chen, J. M. Boone, C. Abbey, J. Hargreaves, C. Bateni, K. K. Lindfors, K. Yang, A. Nosratieh, A. Hernandez, and P. Gazi, "Simulated lesion, human observer performance comparison between thin-section dedicated breast ct images versus computed thick-section simulated projection images of the breast," *Physics in Medicine & Biology*, vol. 60, no. 8, p. 3347, 2015.

[4] M. P. Eckstein, J. L. Bartroff, C. K. Abbey, J. S. Whiting, and F. O. Bochud, "Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal known exactly detection tasks," *Opt. Express*, vol. 11, no. 5, pp. 460–475, 2003.

[5] M. Han, C. Lee, S. Park, and J. Baek, "Investigation on slice direction dependent detectability of volumetric cone beam CT images," *Opt. Express*, vol. 24, no. 4, pp. 3749–3764, 2016.

[6] M. Han, B. Kim, and J. Baek, "Human and model observer performance for lesion detection in breast cone beam CT images with the FDK reconstruction," *PloS one*, vol. 13, no. 3, p. e0194408, 2018.

[7] A. Burgess, "Statistically defined backgrounds: performance of a modified nonprewhitening observer model," *J. Opt. Soc. Am. A.*, vol. 11, no. 4, pp. 1237–1242, 1994.

[8] C. K. Abbey and H. H. Barrett, "Human-and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *J. Opt. Soc. Am. A.*, vol. 18, no. 3, pp. 473–488, 2001.

[9] Y. Zhang, B. T. Pham, and M. P. Eckstein, "Evaluation of internal noise methods for Hotelling observer models," *Med. Phys.*, vol. 34, no. 8, pp. 3312–3322, 2007.

[10] C. Castella, M. P. Eckstein, C. K. Abbey, K. Kinkel, F. R. Verdun, R. S. Saunders, E. Samei, and F. O. Bochud, "Mass detection on mammograms: influence of signal shape uncertainty on human and model observers," *J. Opt. Soc. Am. A.*, vol. 26, no. 2, pp. 425–436, 2009.

[11] F. Massanes and J. G. Brankov, "Evaluation of CNN as anthropomorphic model observer," in *Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment*, vol. 10136. International Society for Optics and Photonics, 2017, p. 101360Q.

[12] F. K. Kopp, M. Catalano, D. Pfeiffer, A. A. Fingerle, E. J. Rummeny, and P. B. Noël, "CNN as model observer in a liver lesion detection task for x-ray computed tomography: A phantom study," *Med. Phys.*, vol. 45, no. 10, pp. 4439–4447, 2018.

[13] A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," *Med. Phys.*, vol. 28, no. 4, pp. 419–437, 2001.

[14] P. C. Johns and M. J. Yaffe, "X-ray characterisation of normal and neoplastic breast tissues," *Phys. Med. Biol.*, vol. 32, no. 6, pp. 675–695, 1987.

[15] R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Med. Phys.*, vol. 12, no. 2, pp. 252–255, 1985.

[16] J. M. Boone, A. L. C. Kwan, J. A. Seibert, N. Shah, K. K. Lindfors, and T. R. Nelson, "Technique factors and their relationship to radiation dose in pendant geometry breast CT," *Med. Phys.*, vol. 32, no. 12, pp. 3767–3776, 2005.

[17] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am.*, vol. 1, no. 6, pp. 612–619, June 1984.

[18] P. K. D and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

Grating-based Spectral CT using Small Angle X-ray Beam Deflections

J. Webster Stayman[1], Matthew Tivnan[1], Grace J. Gang[1], Wenying Wang[1], Nadav Shapira[2], Peter B. Noël[2]
1. Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD
2. University of Pennsylvania, Department of Radiology, Philadelphia, PA

Interest in spectral CT for diagnostics and therapy evaluation has been growing. Acquisitions of data from distinct energy spectra provide, among other advantages, quantitative density estimations for multiple materials. We introduce a novel spectral CT concept that includes a fine-pitch grating for prefiltration of the x-ray beam. The attenuation behavior of this grating changes significantly if x-rays are slightly angled in relation to the grating structures. To apply such an angle (i.e. switch between the different filtrations) we propose a fast, controllable, and precise solution by moving the focal spot of the x-ray tube. In this work, we performed preliminary evaluations with a grating prototype on a CT test bench. Our results include x-ray spectrometer measurements that reveal diverse and controllable spectral shaping between 4° and 6° for a specific grating design. Additional experiments with a contrast agent phantom illustrated the capability to decompose clinically relevant iodine concentrations (5, 10, 20, and 50mg/mL) – demonstrating the feasibility of the grating-based approach. Ongoing and future studies will investigate the potential of this novel concept as a relatively simple upgrade to standard energy-integrating CT.

***Index Terms*—Spectral CT, dual energy CT, material decomposition, quantitative CT, adaptive imaging**

I. Introduction

SPECTRAL CT is finding increased use in both research and clinical settings. Obtaining projection data across two or more spectral channels enables material decomposition and quantitative material density estimation. This drives multiple clinical diagnostic tests including quantitation of exogenous contrast agents, virtual non-contrast images, and endogenous material quantification (e.g. bone mineral density estimation [1]). Similarly, spectral CT can be used to reduce energy-dependent effects like beam hardening. There are multiple strategies for obtaining multi-channel spectral data. For example, x-ray detectors with energy discrimination including multi-layer [2] and photon counting detectors [3], [4] have permitted data acquisition with multiple spectral bins. On the source side, kV-switching [5] and multiple x-ray tubes [6] can provide two (or more) channels by changing the peak energy of the x-ray beam. Other techniques include various beam filtration methods where the x-ray spectrum is shaped using, for example, k-edge filters. Such strategies include split filters[7], rotating filters[8], and moving tiled filters[9]. All of these techniques have various advantages and disadvantages. In particular, many options involve new and expensive hardware, motion systems that can be complicated to integrate within the CT gantry, or limited control over the spectral response. In this work, we present a new source-side filtration method that allows dynamic spectral beam shaping without moving the filter.

II. Methods

A. Overview of Grating-based Filtration

The essential idea explored in this work is to use a periodic grating (much like an anti-scatter grid) to filter the x-ray beam wherein the grating is composed of, for example, a transition metal with a k-edge in the diagnostic range. Consider a regular grating like that shown in Fig. 1 with no material in-between the metal grating structure. Additionally, consider that the grating period is on the order of the focal spot size or smaller, so that the grating structure itself cannot be resolved (again, like a grid). When x-rays traverse the filter at 0°, a portion of the beam is attenuated by a relatively long length of metal (effectively blocking the beam). The remaining portion of the beam is essentially unfiltered through the air. Thus, the net effect is to not change the beam spectrum very much at all, but there is a decrease in the overall fluence by a proportion based on the relative blocker and gap widths. Next, consider the case when the incident beam is angled slightly. More of the beam is filtered by the metal, and the thickness through the metal is decreased. Thus, the overall beam spectrum has a smaller contribution of unfiltered x-rays and the x-ray filtration is less extreme, yielding a weighted average of the unfiltered beam and the k-edge filter. Similar weighted averaging of spectra occurs for increased angles wherein the beam traverses two or more periods. Thus, by changing the angulation of the beam, one can control the x-ray spectrum. While this illustration relies on parallel x-rays, one could also design a focused grating to achieve the same effect with a diverging beam.

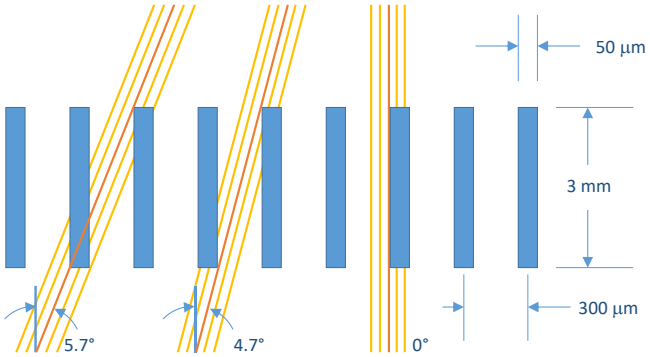
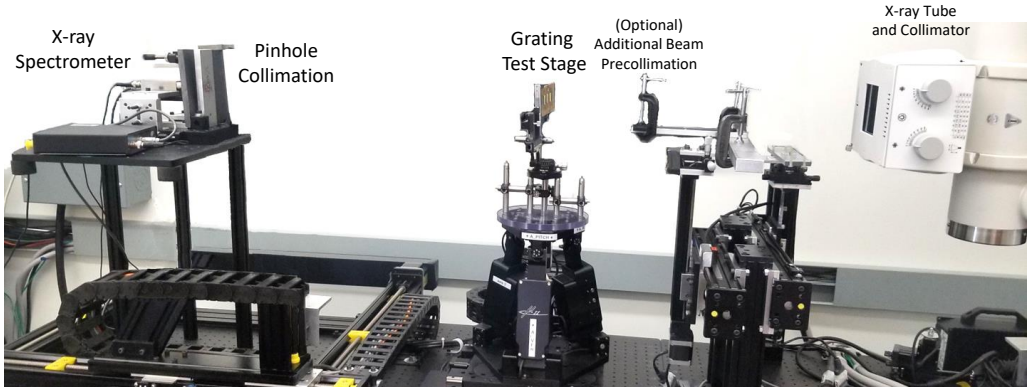


Fig. 1. Illustration of the spectral grating filter concept showing lamellae in blue. As the incident x-ray beam angle changes the amount of beam filtration can be controlled. Grating specifications for the particular grating used in this work are specified. Similarly, two critical angles are shown: 1) The angle at which all rays pass through exactly one lamella of the grating – 4.7°; and 2) the angle where all x-rays pass through exactly two lamellae – 5.7°.

This work was supported in part by NIH grant R21EB026849.



Note that this kind of spectral shaping does require angulation of the x-ray beam. While this could be achieved through source or grating motion, the grating can be designed to be sensitive to small angular changes. With proper design of the grating to achieve a significant spectral change over a small enough angle, x-ray motion can be achieved through electromagnetic deflection of the focal spot (of the same kind used in modern CT systems to achieve improved sampling [10], [11]).

In the following subsections we illustrate the design and fabrication of a grating prototype for spectral shaping, present x-ray spectrometer measurements through this grating as a function of x-ray angle, and integrate the grating prototype into CT test-bench and conduct a simple material decomposition of clinically relevant iodine concentrations.

B. *Grating Design*

There are various options for grating design. In this work, we adopt a fabrication approach based on the creation of a notched form in which thin rectangles of metal foil are placed to form the grating. The notched form is created from deep etched acrylic using a laser cutter/etcher. This manufacturing method permits the use of a wide range of filtration materials with no interstitial material. It is limited by the resolution of the laser cutter and depth control, and the need for metal used for the grating to be self-supporting.

The design chosen for this work is illustrated in Fig. 1 with 50 μm wide tungsten lamellae 3 mm thick. The center-to-center spacing of the metal foil is 300 μm yielding a 250 μm air-gap. The acrylic form is deep etched to about 300 μm. The face of the fabricated grating is 2.5 cm wide x 1.5 cm tall. (The results of this fabrication can be seen in Fig. 3A.) This design results in two "critical points" at 4.7° and 5.7° - namely the first angle where all x-rays pass through one grating lamella; and where all x-rays pass through two grating lamellae (as illustrated in Fig. 1). We expect that the greatest amount of spectral shaping occurs between these two critical points based on the thickness of tungsten traversed.

C. *X-ray Spectrometer Measurements*

While we envision that spectral grating will ideally be implemented using focal spot deflection, we investigate the grating structure using a computer-controlled rotation stage. In a first set of experiments, we use an Amptek XR-100T CdTe x-ray spectrometer (Bedford, MA) to characterize the filtered x-ray beam. Due to count rate limitations, the spectrometer is collimated using a Fluke 10 μm pinhole (part number 07-633) placed in a 13 mm thick lead sheet. The experimental setup is shown in Figure 2. All measurements were taken at 100 kVp and 0.8 mAs/pulse, and averaged over 300 pulses. The pinhole was carefully aligned by maximizing count rate. Spectra were obtained over 0°-7° with 0.25° increments.

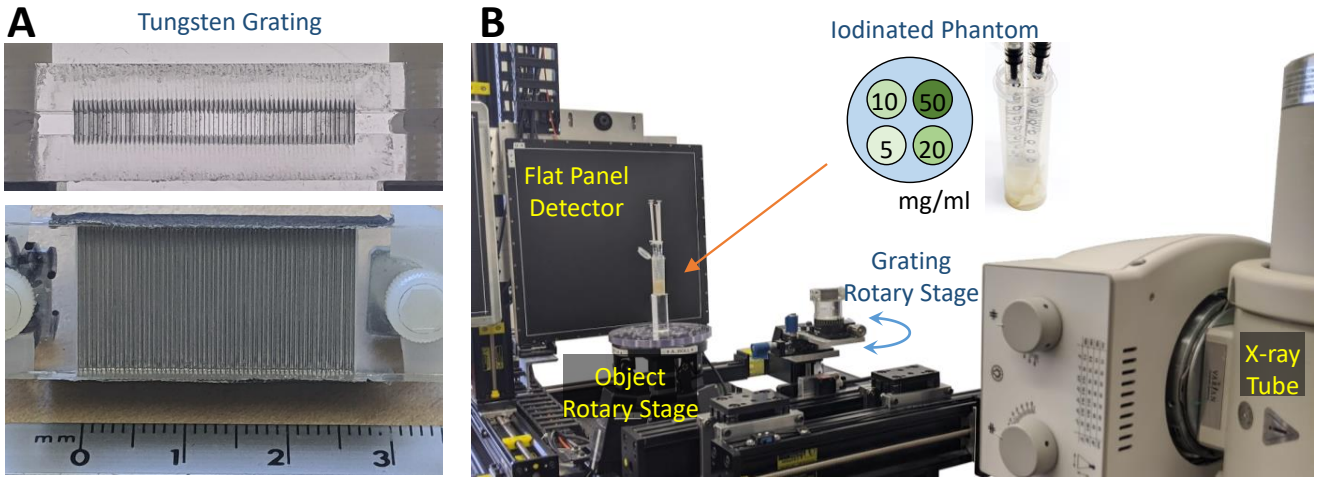


Fig. 3. A) Illustration of the manufactured grating showing the top view and the surface perpendicular to the x-ray path. B) The grating-based spectral CT test-bench including a small phantom with four concentrations of iodine (5, 10, 20, and 50 mg/mL of iodine - Omnipaque).

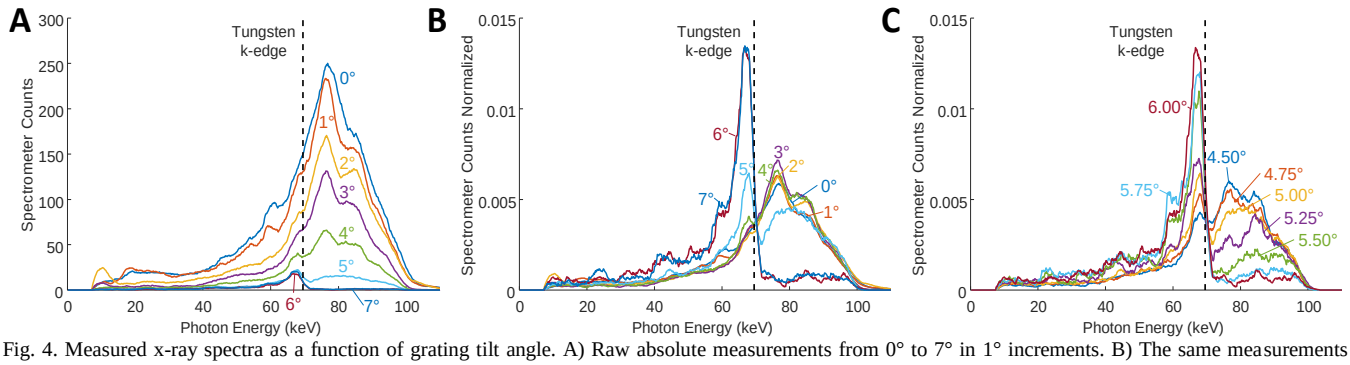


Fig. 4. Measured x-ray spectra as a function of grating tilt angle. A) Raw absolute measurements from 0° to 7° in 1° increments. B) The same measurements normalized by the total number of photon counts better illustrating the change in spectral shape from 0° to 7°. C) A finer sampling of spectra using 0.25° increments over 4.5° to 6.0° to illustrate additional intermediate spectra. Note the prominent change in spectral shape about 69.5 keV – the k-edge of tungsten.

D. Grating-based Spectral CT Experiment

A spectral CT experiment was conducted using the test-bench shown in Fig. 3B. The test-bench consists of an object rotation stage, a Varex 4343CB (Salt Lake City, UT) flat-panel detector, and a Varex Rad-94 Rad/Fluoro tube. A small 25 mm diameter iodine-water phantom was prepared with four different concentrations of Omnipaque – 5, 10, 20, and 50 mg/mL of iodine. The phantom is intentionally small because the spectral grating was constructed with parallel lamellae – a relatively small phantom limits differential filtration due to diverging x-rays by constraining the projection data to a smaller angular range. Similarly, a relatively long baseline geometry was used to limit the effects of divergent x-rays. Specifically, the source-to-detector distance was 1600 mm, the source-to-axis distance was 1100 mm, and source-to-grating distance was 640 mm. Projection data were obtained using the 0.8 mm focal spot, 100 kVp, and 2 mAs/frame with 360 frames evenly spaced over 360°. Data were obtained at grating angles of 4° and 6°. These two angles were chosen based on the results of the x-ray spectrum estimates.

Since some banding artifacts appeared in projection data with simple gain correction (likely because the grating period was not small enough compared to the focal spot size, and because of focal spot instability and small shifts), we applied two projection-domain correction strategies: 1) a focal spot shift estimation and application to the gain scan [12]; and 2) a radial filtering operation to suppress residual bands. Data were reconstructed using filtered-backprojection and additional image-domain ring-artifact suppression routines were applied. Both 4° and 6° grating projection data were processed. Iodine and water maps were formed using a simple empirically selected image-domain weighted sum of the two spectral channels.

III. Results

A. X-ray Spectra as a Function of Grating Angle

The results of the x-ray spectrometry experiment are shown in Fig 4. Fig. 4A shows the raw photon count measurements for grating angulations from 0° to 7°. While some beam shaping is evident, the more noticeable feature is the overall reduction in fluence with the changing angle. To highlight the spectral shaping, we show normalized spectra in Fig. 4B. In this case, the spectra are divided by the total number of counts. Note the obvious shaping of the beam and the dramatic reduction of higher energy photons (above the k-edge of tungsten at 69.5 keV). We see the most significant change in spectral shape between 4° and 6° where the angulation crosses the previously identified "critical points." To better illustrate the transition spectra in this range, a third plot in Fig. 4C shows a finer sampling of spectra between 4.5° and 6°. Note that 0.25° increments over this range result in significant changes to the x-ray spectrum.

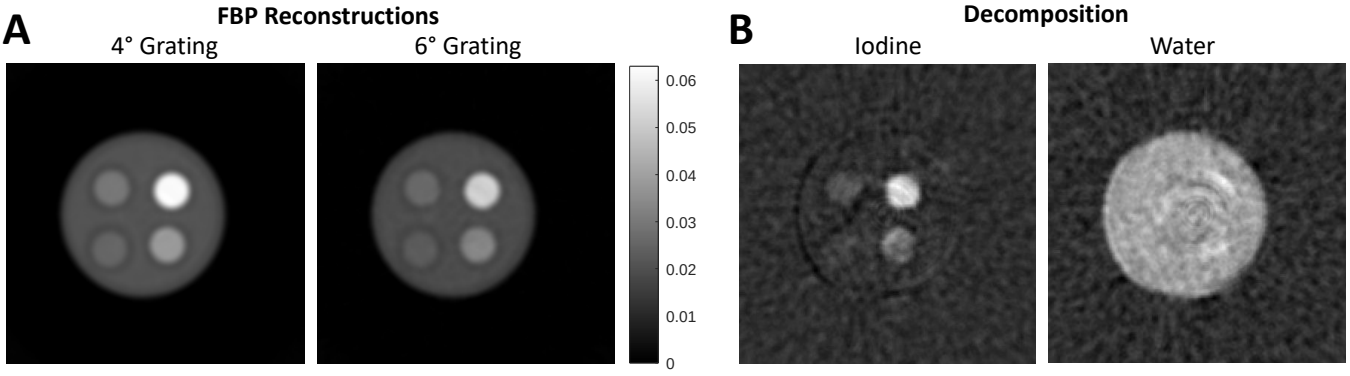
B. Grating-based Spectral CT

Based on the spectral estimation results from the previous subsection we chose to collect CT projection data at 4° and 6° rotations of the grating. Reconstructions from these two datasets are shown in Fig. 5A. Note the contrast differences (particularly noticeable in the highest concentration of iodine) between the inserts and background due to the differing incident x-ray spectrum. Image-domain decomposition was applied using a weighted sum of the two reconstructions. Approximate iodine and water basis channels are shown in Fig. 5B. We see a good representation of the four concentration levels of iodine in the iodine basis channel, and a relatively uniform water basis channel for both the inserts and the water background.

IV. Discussion

The presented results reflect a proof-of-principle demonstration of a novel spectral CT concept that is a relatively simple upgrade to standard energy-integrating CT. Our concept entails a fine-pitch grating that controls the spectral shaping through slight rotations. We demonstrated in a preliminary experiment that spectral CT imaging with a grating prototype is feasible and that material decomposition of iodine in clinical relative concentration is possible. For a detailed analysis of sensitivity and robustness, studies with increased field-of-view, enabled focal-spot movements, and optimized grating design are required. To this end, our results illustrate that gratings with the required specifications can be manufactured without need of highly expensive processes. For future studies, we will investigate different fabrication processes, including lithography, mechanical folding, and 3D laser cutting, with the goal to drive grating designs with high aspect ratios to translate this concept from bench to clinical CT systems.

For implementation into a clinical CT gantry, we propose to move the focal spot of the x-ray tube through electromagnetic steering, to switch between the filtration scenarios (i.e. applying the necessary angle). Focal spot movement is a precise (sub-



pixel) and commonly used technique in most modern x-ray tubes and is rapidly performed from projection to projection image [11]. Utilizing focal spot movements will allow: (i) the generation of spectral data without mechanical motions (which are challenging to coordinate within a fast-rotating gantry), and (ii) extension of our concept to multi-energy CT. For example, one can combine two gratings – each manufactured from a different k-edge material (e.g., tantalum and gold) in series. With orthogonally oriented lamella, two-dimensional angulations achieved via electromagnetically deflection of the focal spot in two directions would allow an even more diverse continuum of emitted x-ray spectra. This extension to multi-energy CT will drive the clinical introduction and translation of novel diagnostic applications like dual-contrast protocols in oncology or cardiovascular imaging [13]–[18].

Future studies will investigate comparisons between this and other spectral CT methods. For example, kV-switching can achieve different spectra without mechanical motion; however, the effected spectra may have decreased separation versus the grating based approach. Moreover, combinations of spectral methods – e.g. kV-switching and grating – may be combined for increased sensitivity and material decomposition performance.

V. Conclusion

The translation and introduction of spectral CT into the clinical arena is ongoing and becoming increasingly popular. Our concept of spectral CT imaging with gratings adds an attractive and relatively simple alternative solution to existing implementation, which may enable multi-energy CT imaging in the future.

VI. Acknowledgment

We thank Roland Proksa and Kevin Brown for fruitful discussions.

References

[1] K. Mei *et al.*, "Bone mineral density measurements in vertebral specimens and phantoms using dual-layer spectral computed tomography," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, 2017.

[2] R. Carmi, G. Naveh, and A. Altman, "Material Separation with Dual-Layer CT," in *IEEE Nuclear Science Symposium Conference Record*, 2005, vol. 4, pp. 1876–1878.

[3] J.-P. Schlomka *et al.*, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys. Med. Biol.*, vol. 53, no. 15, pp. 4031–4047, Aug. 2008.

[4] E. Roessl and R. Proksa, "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," *Phys. Med. Biol.*, vol. 52, no. 15, pp. 4679–4696, Aug. 2007.

[5] D. Xu *et al.*, "Dual energy CT via fast kVp switching spectrum estimation," in *SPIE Medical Imaging*, 2009, p. 72583T.

[6] L. Yu, A. N. Primak, X. Liu, and C. H. McCollough, "Image quality optimization and evaluation of linearly mixed images in dual-source, dual-energy CT," *Med. Phys.*, vol. 36, no. 3, pp. 1019–1024, 2009.

[7] B. Rutt and A. Fenster, "Split-filter computed tomography: a simple technique for dual energy scanning.," *J. Comput. Assist. Tomogr.*, vol. 4, no. 4, pp. 501–9, Aug. 1980.

[8] R. Hauck, F. Gray, and R. Brandt, "Spinning filter for x-ray," US Patent Number: 4,399,550.

[9] M. Tivnan, W. Wang, S. Tilley, J. H. Siewerdsen, and J. W. Stayman, "Optimized spatial-spectral CT for multi-material decomposition," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2019, p. 21.

[10] T. G. Flohr, K. Stierstorfer, S. Ulzheimer, H. Bruder, A. N. Primak, and C. H. McCollough, "Image reconstruction and image quality evaluation for a 64-slice CT scanner with z-flying focal spot," *Med. Phys.*, vol. 32, no. 8, pp. 2536–2547, Jul. 2005.

[11] R. Behling, *Modern diagnostic x-ray sources : technology, manufacturing, reliability*. CRC Press, 2015.

[12] G. J. Gang *et al.*, "Dynamic fluence field modulation in computed tomography using multiple aperture devices," *Phys. Med. Biol.*, vol. 64, no. 10, May 2019.

[13] D. P. Cormode *et al.*, "Multicolor spectral photon-counting computed tomography: in vivo dual contrast imaging with a high count rate scanner," *Sci. Rep.*, vol. 7, no. 1, p. 4784, Dec. 2017.

[14] J. Dangelmaier *et al.*, "Experimental feasibility of spectral photon-counting computed tomography with two contrast agents for the detection of endoleaks following endovascular aortic repair," *Eur. Radiol.*, pp. 1–8, 2018.

[15] D. Muenzel *et al.*, "Spectral Photon-counting CT: Initial Experience with Dual–Contrast Agent K-Edge Colonography," *Radiology*, vol. 283, no. 3, pp. 723–728, 2017.

[16] S. Si-Mohamed *et al.*, "Spectral Photon-Counting Computed Tomography (SPCCT): in-vivo single-acquisition multi-phase liver imaging with a dual contrast agent protocol," *Sci. Rep.*, vol. 9, no. 1, p. 8458, Dec. 2019.

[17] R. Symons *et al.*, "Dual-contrast agent photon-counting computed tomography of the heart: initial experience," *Int. J. Cardiovasc. Imaging*, vol. 33, no. 8, pp. 1253–1261, Aug. 2017.

[18] S. Tao, K. Rajendran, C. H. McCollough, and S. Leng, "Feasibility of multi-contrast imaging on dual-source photon counting detector (PCD) CT: An initial phantom study," *Med. Phys.*, vol. 46, no. 9, pp. 4105–4115, Sep. 2019.

C-Arm Non-Circular Orbits: Geometric Calibration, Image Quality, and Avoidance of Metal Artifacts

P. Wu, N. Sheth, T. Wang, A. Sisniega, A. Uneri, R. Han, R. Vijayan, P. Vagdargi,
B. Kreher, H. Kunze, G. Kleinszig, S. Vogt, and J. H. Siewerdsen

***Abstract*—Metal artifacts present a frequent challenge to cone-beam CT (CBCT) in image-guided surgery, obscuring visualization of metal instruments and adjacent anatomy. Recent advances in mobile C-arm systems have enabled 3D imaging capacity with non-circular orbits. We extend a previously proposed metal artifacts avoidance (MAA) method to reduce the influence of metal artifacts by prospectively defining a non-circular orbit that avoids metal-induced biases in projection domain.**

Accurate geometric calibration is an important challenge to accurate 3D image reconstruction for such orbits. We investigate the performance of interpolation-based calibration from a library of co-registered circular orbits to estimate the projection matrix for any non-circular orbit. We apply the method to non-circular scans acquired for MAA, which involves: (i) coarse 3D localization of metal objects via two scout views (without strong priors on the patient or implants) using an end-to-end neural network; (ii) calculation of the metal-induced x-ray spectral shift for all possible views; and (iii) identification of the non-circular orbit that minimizes the variations in spectral shift.

Non-circular orbits with interpolation-based geometric calibration yielded reasonably accurate 3D image reconstruction with a slight loss of spatial resolution (MTF reduced from ~0.25 to ~0.18 at 1 mm^{-1}). Non-circular scans demonstrated expected benefits to 3D sampling – e.g., a ~40% reduction in cone-beam artifact in a Defrise disk phantom. The end-to-end neural network accurately localized metal implants with just two scout views even in complex anatomical scenes, improving Dice coefficient by ~42% compared to a more conventional cascade of separately trained U-nets. In a spine phantom with pedicle screw instrumentation, non-circular orbits identified by the MAA method reduced the magnitude of metal "blooming" artifacts (apparent width of the screw shaft) in CBCT reconstructions by ~70%.

The proposed imaging and calibration methods present a fairly practical means to improve image quality in mobile C-arm CBCT by identifying non-circular scan protocols that improve sampling and reduce metal-induced biases in the projection data.

***Index Terms*—Cone-beam CT, non-circular orbit, mobile C-arm, geometric calibration, metal artifact, image-guided surgery**

I. Introduction

Intraoperative cone-beam CT (CBCT) is increasingly used for guidance and validation in placement of surgical instruments [1], [2], including spinal pedicle screws, neurosurgical electrodes etc. Such highly attenuating metal objects within the field-of-view (FOV) can cause metal artifacts that degrade image quality and confound visualization of adjacent anatomical structures. Many algorithms have been proposed for metal artifacts reduction, including classic projection / image domain postprocessing ("MAR") methods [3] and polyenergetic reconstruction methods.[4] However, these methods contend with artifacts *after* biases are already incurred in the projection data. A recent method (called metal artifact *avoidance*, MAA) [5] proposes to acquire projection data with minimal metal-induced bias, allowing MAR or polyenergetic reconstruction to work even better – or in some cases yielding "raw" CBCT images with artifacts reduced to a level sufficient for the imaging task.

The research was supported by NIH R01-EB-017226 and academic-industry partnership with Siemens Healthineers.
P. Wu, N. Sheth, T. Wang, A. Sisniega, A. Uneri, R. Han, R. Vijayan, P. Vagdargi, and J. H. Siewerdsen are with Johns Hopkins University, USA.
B. Kreher, H. Kunze, G. Kleinszig, and S. Vogt are with Siemens Healthineers, Germany.

Increasingly, interventional CBCT systems are capable of computer-controlled motion of the x-ray source and detector for general, non-circular orbits that can e.g. improve sampling characteristics [6], and maximize the spatial-frequency-dependent signal and noise transfer characteristics with respect to the imaging task [7]. Such orbits also provide a means to reduce the influence of metal instrumentation (i.e., reduce / avoid metal artifacts) by collecting data in a manner that minimizes the factors of metal-induced bias (e.g. beam hardening). Recently, such capability for complex source-detector orbits is available on mobile C-arms with motorized control of multiple axes of gantry motion. Comparted to fixed-room, robotic C-arms, such mobile systems feature a relatively small footprint, lower cost, and flexibility in workflow, bringing the potential for non-circular orbits to surgical guidance.

In this work, we report the capacity for 3D imaging with non-circular orbits on a mobile C-arm with motorized tilt and angulation of the gantry. Specifically, we address issues of geometric calibration for general (not pre-defined) non-circular orbits and evaluate 3D imaging performance characteristics of non-circular orbits compared to conventional circular orbits. We use such capability to extend the MAA method [5] to non-circular orbits that minimize metal-induced bias. The method is shown to operate well with just two low-dose scout views (without other prior information of patient anatomy / metal objects) and is compatible with MAR and polyenergetic reconstruction methods that can further improve image quality.

II. Methods

A. Non-Circular Orbits on a Mobile C-Arm System

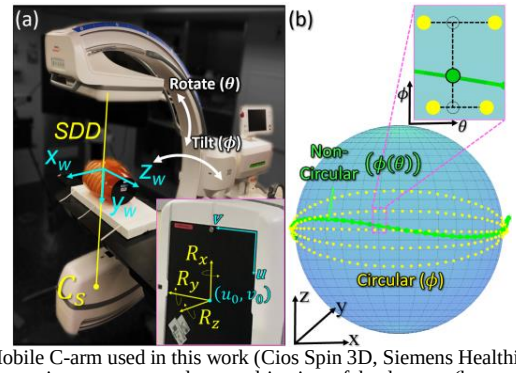


Fig. 1. Mobile C-arm used in this work (Cios Spin 3D, Siemens Healthineers). (a) C-arm geometric parameters and zoomed-in view of the detector (bottom right). (b) Source-detector orbits illustrated on a sphere. Circular orbits are in yellow (various settings of gantry tilt, ϕ). An example non-circular orbit is in green (ϕ changing with θ).The zoomed inset illustrates estimation of geometric parameterization for one view (green dot) determined by interpolation of four nearest vertices in a library of geometric calibrations of circular scans (yellow dots).

The mobile C-arm used in this work (Cios Spin 3D, Siemens Healthineers, Forcheim, Germany) is shown in Fig. 1(a). The C-arm has motorized control of rotation angle (θ: 0° to 196°) and gantry tilt (ϕ: -30° to 30°), permitting non-circular orbits to be executed by variation of θ and ϕ during the scan. Tilted circular orbits (constant ϕ) can be calibrated using established methods – e.g., the BB phantom of Cho et al.[8] but non-circular orbits present a challenge to pre-calibration, given the large variety of feasible (θ, ϕ) combinations. For non-circular scan trajectories defined by methods like "task-driven" imaging [7] and MAA, the

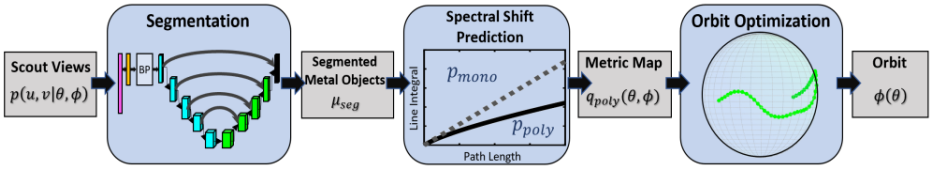


Fig. 2. Flowchart of the MAA algorithm. Two scout views are used for coarse 3D localization of metal objects (μ_{seg}) using an end-to-end CNN. Biases arising from x-ray spectral shift are predicted as a function of gantry rotation angle and tilt. A non-circular orbit that best avoids metal-induced biases is determined by finding a curve in $q_{poly}(\theta,\phi)$ with minimal standard deviation in spectral shift.

orbit is designed on the fly and thus cannot be pre-calibrated. We therefore developed an interpolation-based approach to address this issue, detailed as follows.

First, we built a library of geometric calibrations for circular obits with tilt ranging from ϕ = -30° to +30° at 5° intervals using a cylindrical BB phantom placed at the same tilt angle as the C-arm gantry [8]. Affine transformation of the library registers all calibrations into a common coordinate system:

$$\mathbf{A}_{\phi=0} = \mathbf{A}_{\phi=\phi^*} \times T_{tilt} \times T_w \quad (1)$$

$$T_{tilt} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\phi^*) & -\sin(\phi^*) & 0 \\ 0 & \sin(\phi^*) & \cos(\phi^*) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad T_w = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & r_z & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

where $\mathbf{A}_{\phi=\phi^*}$ is the projection matrix at tilt angle ϕ^*, and T_{tilt} is a transformation through tilt angle ϕ^* (read from the C-arm motor encoder). The transformation T_w accounts for small discrepancies in displacement (t) and tilt (r_z) between the BB phantom and the gantry in each member of the library, determined by 3D-3D registration. The co-registered calibrations were decomposed into parameters: SDD (source-detector distance), C_s (source location), u_0, v_0 (piercing point), and R (detector rotation matrix) as shown in Fig. 1(a). The system geometry for a general non-circular orbit can then be estimated by interpolating the geometric parameters from the calibration library – viz., the four nearest views from calibrated tilted circular orbits as illustrated in Fig. 1(b). Parameters SDD, u_0 and v_0 were estimated by scattered linear interpolation, and C_s, R by spherical linear interpolation (slerp) [9]. The slerp is non-communitive and was performed first in θ and then in ϕ, giving reduced interpolation error, since the θ direction is more finely sampled in the calibration library.

B. Source-Detector Orbits for Metal Artifact Avoidance (MAA)

A flowchart for the MAA method is shown in Fig. 2, with main stages marked in blue squares, summarized as follows:

1) 3D Localization of Metal Objects from Two Scout Views

Given two scout views, the MAA method determines a binary 3D reconstruction (coarse segmentation, μ_{seg}) describing the 3D location of metal objects in the FOV. Unlike in MAR, a coarse segmentation was found to be sufficient for MAA, requiring only an approximation of the shape, size, and orientation of metal objects.

Coarse segmentation used a novel, end-to-end convolutional neural network (CNN) shown in Fig. 3, which learns features simultaneously in projection and image domains. Specifically, projection domain feature extraction layers are connected to image domain 3D segmentation layers through a backprojection layer. Intuition underlying such a network is that image domain segmentation is improved by fusing shape information from the projection domain (before being smeared by backprojection), and projection domain feature extraction is guided by the segmentation loss backpropagated from the image domain.

The feature extraction layers contain two 3×3 convolutions, each followed by a rectified linear unit (ReLU) with batch normalization, and in the end a 2×2 max pooling layer. The 3D segmentation layers were implemented as a three-layer 3D U-Net (slightly modified from [10], detailed in Fig. 3) with general Dice loss function. The derivative of the Dice loss function with respect to the input of the backprojection layer is the forward projection operation (determined with C-arm geometry and Siddon ray tracing [11], not taken as learnable parameter of the network), enabling joint learning in projection and image domains. The network was "He normal" initialized and trained using the Adam optimizer with an initial learning rate of $5\text{x}10^{-3}$ for 30 epochs.

For simplicity and to avoid the requirement for vendor-specific metal instrument models, the network was trained with only simulated data, which was generated by overlaying DRRs of clinical CT images with a random number of forward-projected metal objects: ellipsoids ranging in size and eccentricity and coarsely segmented spine screws obtained from CT images. The effects of data truncation, partial volume effect, noise (quantum and electronic), scatter, and beam hardening were included in DRR generation. Data augmentation included variation in the location, size, orientation, aspect ratio, and attenuation of simulated metal objects in each DRR, resulting in a total of ~8,400 (8,000 training + 400 validation) images.

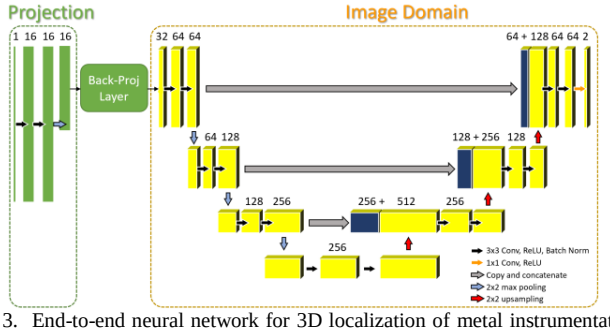


Fig. 3. End-to-end neural network for 3D localization of metal instrumentation. Projection domain feature extraction layers are connected to the image domain segmentation layers through a backprojection layer (determined from C-arm geometry), enabling joint learning in projection and image domains.

2) Prediction of Metal-Induced Bias in Projection Data

The second step of the MAA method is to estimate the metal-induced bias (e.g., summation of x-ray spectral shift) in the projection domain for all rotation (θ) and tilt (ϕ) angles [5]:

$$q_{poly}(\theta,\phi) = \sum_{u,v}\left(p_{mono}(u,v|\theta,\phi) - p_{poly}(u,v|\theta,\phi)\right) \quad (3)$$

$$p_{mono}(u,v|\theta,\phi) = c\left(\mathbf{A}\mu_{seg}\right) \quad (4)$$

$$p_{poly}(u,v|\theta,\phi) = \sum_{i=1}^{N}\alpha_i p_{mono}^{i}(u,v|\theta,\phi) \quad (5)$$

where q_{poly} is the spectral shift predicted for each θ and ϕ, $p_{mono}(u,v|\theta,\phi)$ is the monoenergetic line integral, $p_{poly}(u,v|\theta,\phi)$ is the corresponding polyenergetic line integral affected by beam-hardening, $\mathbf{A}$ is the projection matrix, c is an empirically estimated scalar representing the attenuation coefficient of the metal object, and α_i are precalculated polynomial mapping coefficients determined by the spectrum and metal material. The MAA method is relatively insensitive to both the spectral model and the metal material (including c), since the spectral model is only used to predict desirable orbits (not to perform an actual beam-hardening correction).

3) Orbit Optimization & Reconstruction

The third step of the MAA method is to identity the non-circular orbit that minimizes projection data biases (and thereby avoid metal artifacts in the 3D image reconstruction). Rather than simply

summing the spectral shift at each θ, we chose an objective that describes the *inconsistency* in spectral shift between projection views, since such inconsistencies are the underlying source of streaks commonly identified as metal artifacts. To reduce dimensionality and encourage a smooth non-circular orbit, we modeled $\phi(\theta)$ as a superposition of cubic b-splines:

$$\hat{f} = \arg\min_{f} = Q_{poly}(f) \quad (6)$$

$$Q_{poly}(f) = \sigma\left[q_{poly}(\theta, \sum_{i=0}^{M} f_i B(\theta - \theta_i))\right] \quad (7)$$

where Q_{poly} is a metric of spectral shift *variation* determined by the standard deviation operator (σ), B is the cubic b-spline (M = 10 control points), and f_i is the control parameter for spline knot control points i. The non-convex optimization in Eqs. (6-7) was solved with the covariance matrix adaptation evolution strategy (CMA-ES)[12] solver operating in near real-time (~0.5 s runtime). The penalized weighted least squares (PWLS) method [13] was used for image reconstruction for both circular and non-circular orbits.

C. Experimental Study

1) Geometric Calibration of Non-Circular Orbits

The reproducibility of geometric parameters was evaluated for an example non-circular orbit (ϕ linearly increasing from -20° to +20° while θ linearly increasing from 0° to 196°) by repeating the geometric calibration (not interpolation) 4 times over an 8-hour period of normal use. Three scenarios were evaluated: (i) a conventional "Circular Protocol" with a pre-calibrated circular orbit; (ii) a "Non-Circular Protocol" assuming a pre-calibrated non-circular orbit; and (iii) a "General Non-Circular" scan for which the projection matrix is determined by the interpolation-based method described above.

Basic image quality characteristics were evaluated in terms of spatial resolution (modulation transfer function, MTF) and 3D sampling characteristics (cone-beam artifacts) in a head phantom containing a variety of custom inserts. The axial plane MTF was determined from the edge-spread function measured from a high-contrast (~300 HU) cylindrical rod insert. The magnitude of cone-beam artifacts was measured in terms of the full-width-at-half-maximum (FWHM) of the superior-inferior edges of flat disks ("Defrise phantom") inserted within the cranium. All scans involved 400 projections over a 196° arc in θ at 110 kV, 1.65 mAs / view, with a 30 s scan time.

2) Metal Artifact Avoidance (MAA) with Non-Circular Orbits

The performance of the CNN-based metal object localization step was evaluated in terms of Dice coefficient in the validation dataset and in the testing dataset (scans of a cadaver instrumented with six pairs of pedicle screws (Nuvasive, San Diego USA)). We compared the performance of the proposed end-to-end method with two other methods: (i) conventional image domain U-Net segmentation on direct backprojection of scout views (referred to as the "single U-Net");[5] and (ii) projection domain U-Net segmentation of metal objects, followed by backprojection and image domain U-Net segmentation (referred to as the "dual U-Net"). Note that the two U-Nets in (ii) were trained separately (not end-to-end).

The performance of the MAA method was evaluated in a phantom study involving an anthropomorphic chest phantom containing a natural human skeleton in tissue-equivalent plastic and 8 spinal pedicle screws (DePuy-Synthes, Raynham USA; ranging 30-60 mm in length). The screws were placed with out-of-plane angle ranging from -20° to +30° (positive on one side of the spine, negative on the other). Metal artifact magnitude was assessed in terms of "blooming" about the shaft of the screw (FWHM of the screw minus its true width) for conventional circular orbit scans and the optimal non-circular orbit defined by MAA.

III. Results

A. Geometric Calibration of Non-Circular Orbits

Figure 4 shows the degree of reproducibility in system geometry for the non-circular orbit described in §C.1. A reduction in reproducibility was observed: for example, the standard deviation in piercing point location (u_0, v_0) over repeat trials was ~1.7 mm (vs. ~0.8 mm for a conventional circular orbit). Note that the mobile C-arm was not electromechanically tuned beyond its standard clinical deployment, which does *not* support non-circular orbits in standard use. As shown below, however, despite the reduced mechanical reproducibility, the interpolation-based calibration provided a reasonable estimate of system geometry supporting 3D image reconstruction.

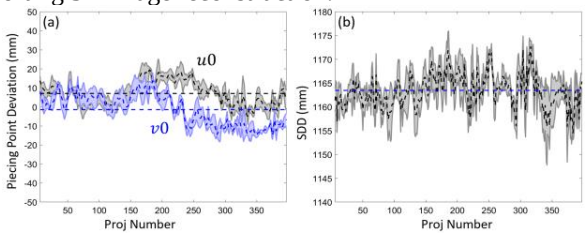


Fig. 4. Reproducibility of geometric parameters for the non-circular orbit described in §C.1. (a) Location of the piecing point (u_0, v_0) and (b) SDD in 4 repeat trials over an 8-hour interval of routine use.

Figure 5 illustrates how errors in geometric calibration relate to reduction in spatial resolution for the three scenarios defined in §C.1. Compared to the standard "Circular Protocol," the calibrated "Non-Circular Protocol" showed minor reduction in MTF (~3.5% reduction in the spatial frequency at which MTF = 0.10, f_{10}) due to the decreased reproducibility in calibration parameters. The "General Non-Circular" scenario showed further reduction in resolution (~12% reduction in f_{10}) due to two effects: smoothing of geometric parameters in the interpolation; and ignoring differences in gantry momentum for a continuous non-circular orbit compared to the discrete library of circular scans. Despite the reduction in MTF, images obtained with the "General Non-Circular" scenario appear visually comparable to the others, as in Fig. 5(b-c). Therefore, the interpolation-based method appears to provide a reasonable estimate of system geometry for non-circular scans for which neither a pre-calibration nor a "self-calibration" (using previous scan and 3D-2D registration [14]) is available.

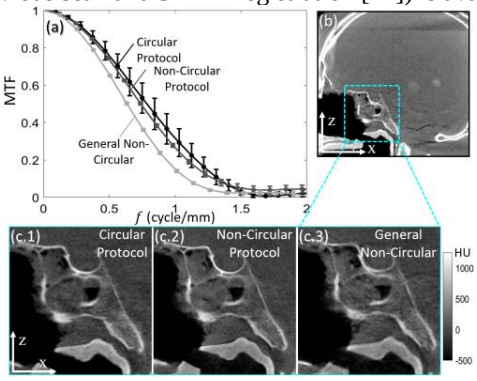


Fig. 5. Effect of geometric calibration accuracy on spatial resolution (MTF) and 3D image quality. (a) MTF for the three scenarios defined in §C.1. (b) Sagittal view of a head phantom with zoomed-in views of the skull base for the three scenarios.

Figure 6 shows the expected reduction in cone-beam artifacts in sagittal images of a head phantom containing stacks of flat disks (§C.1). The apparent thickness (FWHM) of the uppermost disk

was reduced from 9 mm for the standard "Circular Protocol" to its true thickness (~5.5 mm) for both of the Non-Circular scenarios.

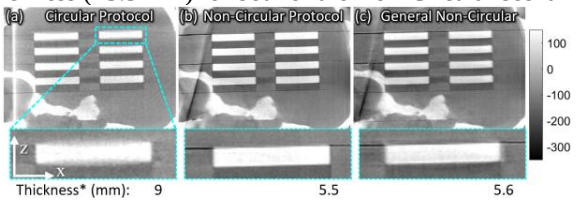


Fig. 6. Reduction of cone-beam artifacts for the non-circular orbits (§C.1) evident in sagittal images of a head phantom containing flat disk inserts.

B. MAA with Non-Circular Orbits

Figures 7 and 8 show the end-to-end network to outperform the single and dual U-Nets. The increase in Dice coefficient (especially for fewer scout views) confirms the effectiveness of end-to-end training: for just two scout views, the end-to-end method increased Dice by ~29% in the validation dataset, and by ~42% in the testing dataset compared to the dual U-Net. Figure 7(b) shows that ≥8 filters in the projection domain were needed for reasonable performance of the end-to-end network. Figure 8 shows that with only two scout views, the end-to-end network was able to capture the shape and direction of all six pairs of bilateral screws. While the Dice coefficient (0.7) would still be considered low for some applications (e.g., MAR) and was shown to improve with more scout views, it is more than sufficient for MAA orbit optimization, as shown below.

Figure 9 shows the $q_{poly}(\theta,\phi)$ metric map computed by the MAA method with just two scout views of the chest phantom and spine screws. Clearly, there is no circular orbit (horizontal line in (a)) that would substantially reduce metal artifact for all screws, as there is always at least one region of strong metal-induced bias in the (θ,ϕ) trajectory space (roughly corresponds to the out-of-plane angle of the screw). This issue is resolved by 2D minimization of Eqs. (6-7), resulting in the non-circular orbit marked by the cyan curve (a). The non-circular orbit avoids most of the low-fidelity views, steering a path in (θ,ϕ) that minimizes variation in spectral shift. The resulting image quality is shown in Figs. 9(c)-(f), showing strong reduction in blooming artifacts about the screw (~70% improvement in screw shaft blooming achieved with the optimal non-circular orbit).

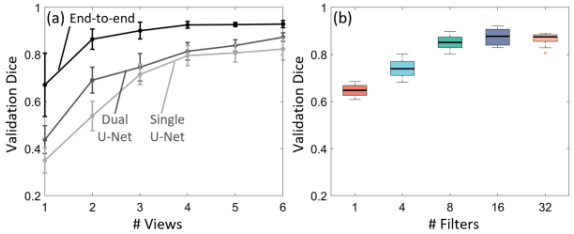


Fig. 7. Performance of CNN in 3D localization of metal instruments from sparse projection views. (a) Validation Dice for the three methods described in §C.2 evaluated as a function of the number of scout views. (b) Validation Dice for the end-to-end network with respect to the number of filters in the projection domain (input to the backprojection layer, as shown in Fig. 3).

REFERENCES

[1] J. H. Siewerdsen *et al.* Medical physics, vol. 32, no. 1, pp. 241–254, 2005.
[2] D. J. Lee *et al.* Journal of Neurosurgery, vol. 122, no. 3, pp. 526–531, Mar. 2015
[3] E. Meyer *et al.* Medical physics, vol. 37, no. 10, pp. 5482–5493, 2010.
[4] I. A. Elbakri *et al.* IEEE Trans. Med. Imaging, vol. 21, no. 2, pp. 89–99, 2002, doi: 10.1109/42.993128.
[5] P. Wu *et al.* SPIE Medical Imaging 2020: Physics of Medical Imaging, Houston, United States, 2020.
[6] G. J. Gang *et al.* Med Phys, vol. 45, no. 1, pp. 144–155, Jan. 2018, doi: 10.1002/mp.12659.
[7] S. Capostagno *et al.* J. Med. Imag., vol. 6, no. 02, p. 1, May 2019, doi: 10.1117/1.JMI.6.2.025004.
[8] Y. Cho *et al.* Med. Phys., vol. 32, no. 4, pp. 968–983, Mar. 2005, doi: 10.1118/1.1869652.
[9] E. B. Dam *et al.* Quaternions, interpolation and animation, vol. 2. Citeseer, 1998.
[10]Ö. Çiçek *et al.* MICCAI 2016, vol. 9901, pp. 424–432.
[11]R. L. Siddon, Medical Physics, vol. 12, no. 2, pp. 252–255, 1985, doi: 10.1118/1.595715.
[12]N. Hansen *et al.* in Parallel Problem Solving from Nature - PPSN VIII, vol. 3242, 2004, pp. 282–291.
[13]J. A. Fessler, IEEE transactions on medical imaging, vol. 13, no. 2, pp. 290–300, 1994.
[14]S. Ouadah *et al.* Phys. Med. Biol., vol. 61, no. 7, pp. 2613–2632, Apr. 2016.
[15]O. Oktay *el al*, arXiv preprint:1804.03999 (2018).

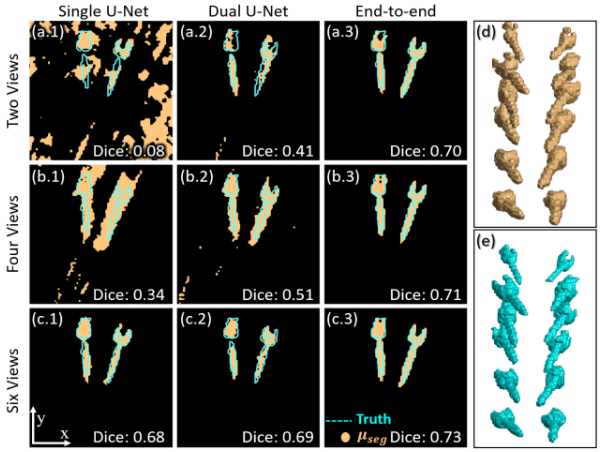


Fig. 8. Segmentation performance of three neural networks in the testing dataset (cadaver with 12 pedicle screws). (a)-(c) Example axial slice segmentation overlaid with ground truth (cyan). Segmentations are shown for varying number of scout views. (d) Isosurface of μ_{seg} computed from two views for the end-to-end method. (e) Isosurface of ground truth segmentation (downsampled the same as μ_{seg}).

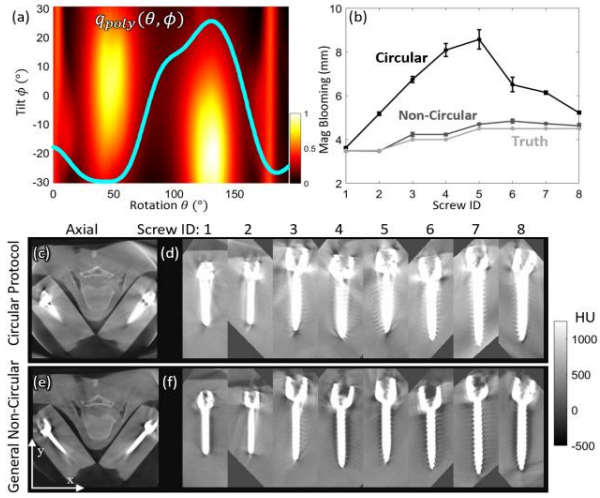


Fig. 9. The MAA method applied to a chest phantom implanted with eight pedicle screws. (a) The $q_{poly}(\theta,\phi)$ metric map overlaid with the optimal non-circular orbit (cyan). (b) Magnitude of blooming artifact (FWHM of the screw shaft) for each of the 8 screws. Axial images in (c and e) for circular and (MAA) non-circular scans show the improvement in visual image quality acquisition, illustrated further in (d and f) by zoomed quasi-axial slices in-plane with each screw.

IV. CONCLUSION

A method for geometric calibration of non-circular C-arm CBCT orbits was demonstrated, giving a practical means to estimate system geometry from a discrete library of circular scan calibrations. Although the geometric calibration method carried a measurable reduction in MTF, the effect on visual image quality was relatively minor. CBCT images acquired from non-circular orbits were shown to improve 3D sampling characteristics (e.g., reduction of cone-beam artifacts), as expected. Such capability enabled a method (called MAA) that identifies non-circular orbits with minimal metal-induced biases. A novel end-to-end neural network was shown to localize metal objects from just two scout views without strong prior information of the patient anatomy or metal instruments. Integration of the end-to-end network with the MAA method for non-circular orbits demonstrated strong reduction in metal artifacts in phantom and cadaver studies. Moreover, the method is compatible with established MAR and polyenergetic reconstruction algorithms to further reduce artifacts. Ongoing / future work includes adding the attention gate mechanism [15] into the end-to-end network and improved geometric calibration method for general non-circular orbits.

Multi-Contrast CT Imaging with a Prototype Spatial-Spectral Filter

Matthew Tivnan, Wenying Wang, J. Webster Stayman
Department of Biomedical Engineering
Johns Hopkins University
720 Rutland Ave., Baltimore MD, 21205
Email: web.stayman@jhmi.edu

***Abstract*—Spectral CT has great potential for a variety of clinical applications due to the improved material discrimination with respect to conventional CT. Many clinical and preclinical spectral CT systems have two spectral channels for dual-energy CT using strategies such as split-filtration, dual-layer detectors, or kVp-switching. However, there are emerging clinical imaging applications which would require three or more spectral sensitivity channels, for example, multiple exogenous contrast agents in a single scan. Spatial-spectral filters are a new spectral CT technology which use x-ray beam modulation to offer greater spectral diversity. The device consists of an array of k-edge filters which divide the x-ray beam into spectrally varied beamlets. This design allows for an arbitrary number of spectral channels; however, traditional two-step reconstruction-decomposition schemes are typically not effective because the measured data for any individual spectral channel is sparse in the projection domain. Instead, we use a one-step model-based material decomposition algorithm to iteratively estimate material density images directly from spectral CT data. In this work, we present a prototype spatial-spectral filter integrated with an x-ray CT test-bench. The filter is composed of an array of tin, erbium, tantalum, and lead filter tiles which spatially modulate the system spectral sensitivity pattern. After the system was characterized and modeled, we conducted a spectral CT scan of a multi-contrast-enhanced phantom containing water, iodine, and gadolinium solutions. We present the resulting spectral CT data as well as the material density images estimated by model-based material decomposition. The calibrated system model is in close agreement with the measured data, and the reconstructed material density images match the ground truth concentrations for the multi-contrast phantom. These preliminary results demonstrate the potential of spatial-spectral filters to enable multi-contrast imaging and other new clinical applications of spectral CT.**

I. Introduction

Spectral CT is a term used to describe any CT system which incorporates projection data acquisition with varied x-ray sensitivity spectra to allow for both spatial reconstruction as well as basis material density estimation. Dual-energy CT is an existing subcategory of spectral CT which is already being successfully utilized in clinical settings. A recent comprehensive review of the modern state of of the technology and its applications [3] enumerates important new clinical tasks made possible by dual-energy CT including automated bone removal in CT angiography [1], virtual monochromatic imaging [10], and virtual non-contrast imaging [2].

Greater diversity of spectral sensitivity has the potential to enable entirely new clinical imaging applications. For example, a multi-contrast protocol for simultaneous multi-phasic liver imaging with spectral CT is already under development [4]. With the goal of imaging multiple biologically safe contrast agents (such as iodine and gadolinium), future systems will need to move beyond dual-energy CT to designs which incorporate three of more sensitivity spectra into a single data acquisition.

One possible solution would be the use of direct photon-counting detectors which are capable of photon-energy discrimination. This technology shows great promise for spectral CT in the future but currently faces its own set of limitations and challenges [5]. We focus on a new technology which is compatible with current energy-integrating detectors called spatial-spectral filters which are shown in Figure 1. In previous work, we have modeled and optimized the filter design [8]. Using those simulation results as guidance for the device design, we have constructed a prototype spatial-spectral filter in an integrated x-ray CT test bench system.

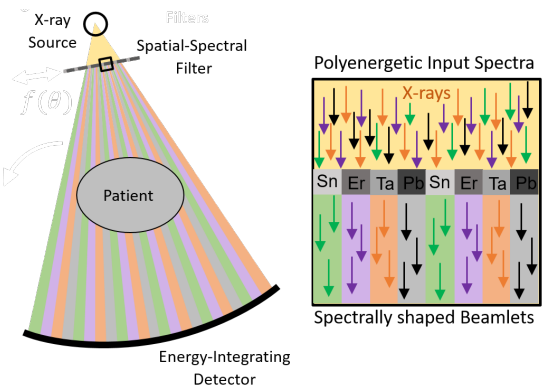


Fig. 1. Diagram of spatial-spectral filters for spectral CT. A tiled array of k-edge filters (e.g. tin, erbium, tantalum, lead) which modulate the spectra of the incident x-rays. The resulting projection data have varied spectral sensitivity.

In this work, we present the prototype spatial spectral filter as well as specialized system calibration and data processing procedures required to use the device for spectral CT. We investigate the efficacy of the spatial-spectral CT system using a multi-contrast phantom containing iodine and gadolinium solutions. We present the spectral CT projection data as well as the basis material density images reconstructed using a direct one-step model-based material decomposition algorithm.

II. Methods

A. Spatial-Spectral Filter Prototype

The prototype spatial-spectral filter consists of tin, erbium, tantalum, and lead filter tiles (ordered as such) which are 20 mm wide, 30 mm high and 0.250 mm thick with the exception of the tantalum tiles which are 0.127 mm thick. The relative position of the filter tiles is clamped in place by an aluminum frame as shown in Figure 2. The entire assembly is integrated into an x-ray CT test bench. It is mounted in front of the x-ray source onto a linear actuator which can translate the device laterally. This axis of motion allows for repositioning throughout the CT scan and corresponding view-dependent modulation of the system spectral sensitivity.

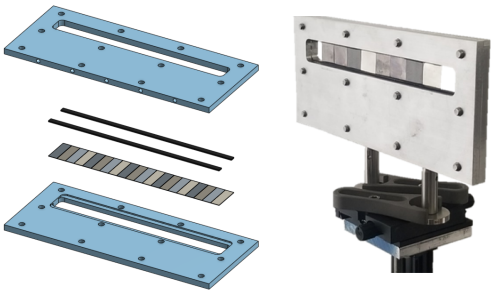


Fig. 2. Mechanical design of spatial-spectral filter prototype. The tin, erbium, tantalum, and lead filter tiles clamped in an aluminum frame.

B. Data Processing and Calibration

The spatial-spectral filter roughly divides the projection data into four spectral channels associated with the filter materials. Each of these channels is individually sparse in the projection domain, taking up approximately 25% of the CT dataset. With this incomplete CT data, it would be difficult to reconstruct images for each spectral channel individually which would be a pre-requisite for material decomposition in the image domain. Projection-domain decomposition is also challenging without any overlap between the spectral channels' acquisition geometries. Instead, we estimate material densities directly from the spectral CT data with a one-step model-based material decomposition algorithm [6] [7]. This algorithm involves a poly-energetic forward model for x-ray attenuation that properly considers the mass attenuation spectra for each of the basis materials, as well as the models for the system geometry and sensitivity spectra. The data log-likelihood term assumes a multivariate Gaussian noise model, and the regularization term is a quadratic smoothness penality with cross-material penalties as described in [9].

System spectral sensitivity is affected by the source fluence spectra, the source-side filtration, and the detector energy-dependent sensitivity. For model-based material decomposition, it is critical to have an accurate model for system spectral sensitivity. The spectral model was calibrated by fitting aluminum and tungsten source filtration to the measured transmissivity spatial-spectral filter.

Small gaps between the tiles result in an unfiltered beam and a corresponding increase in x-ray fluence. We have shown in previous work that reconstruction with filter tile gaps is possible by incorporating the gaps into the system spectral sensitivity model. However, in this work, we simply set the statistical weights to zero around the interface between filter tiles, thereby ignoring this data and avoiding the more complex calibration required to model the gaps.

C. Spatial-Spectral CT with a Multi-Contrast Phantom

With the clinical goal of multi-contrast imaging, we conducted a spectral CT experiment with the spatial-spectral filter x-ray test bench system in Figure 3. The subject of the scan is a multi-contrast phantom consisting of a 65 mm diameter phantom cylindrical water tank and 10 mm diameter vials containing various concentrations of iodine and gadolinium solutions distributed as shown in Figure 6. The source-to-axis distance was 910 mm and the source-to-detector distance was 1920 mm. The spatial-spectral filter was positioned 380 mm from the source and it was translated at a speed of 2.66 mm per view. The flat-panel detector has a caesium iodide scintillator thickness of 0.60 mm and a pixel spacing of 0.278 mm in both the row and column directions.

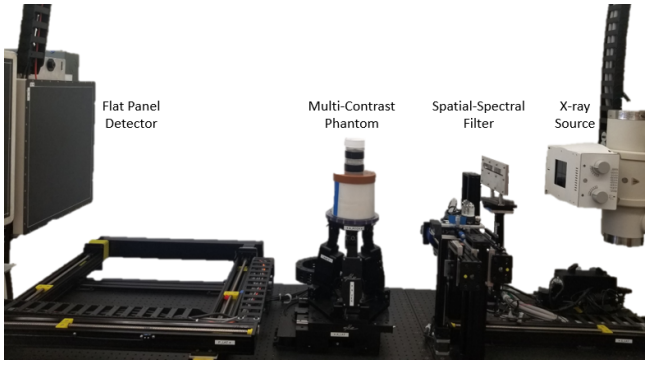


Fig. 3. X-ray test bench outfitted for spectral CT with a spatial-spectral filter

The x-ray technique was 120 kVp with one 40 ms pulse per frame at a current of 30 mA. Air and offset scans were collected with no filtration and no phantom in place. Then, a filter calibration scan was collected with the view-dependent filtration from the spatial-spectral filter, but no phantom in place. After standard air and offset corrections, the spectral calibration parameters were tuned such that the measured transmissivitiy of all four filter tile materials matched the model used for reconstruction for the known filter tile thicknesses. Projection data were acquired over 360° using 360 equally spaced view angles.

We ran 1000 iterations of the material decomposition algorithm. The voxelized representation of the object consisted of 88 voxels with 1 mm spacing. We used quadratic regularization to encourage smooth image results and we used a cross-material penalty to reduce cross-talk artifacts between basis material channels.

III. Results

A. Calibrated Spectral Sensitivity

The transmissivity of the spatial-spectral filter was 38.59% for tin, 20.82% for erbium, 30.85% for tantalum, and 23.97% for lead. Based on these levels, the spectral calibration output parameters are -0.0378 mm of aluminum filtration and 0.1436 mm of tungsten filtration relative to the unfiltered SPEKTR model. The output spectra are shown in Figure 4.

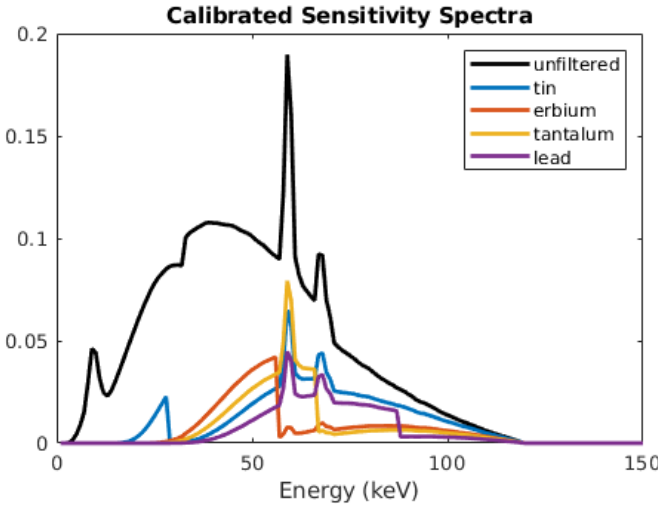


Fig. 4. Calibrated spectral sensitivity for each filter material.

In these spectra, we can see the k-edges at 29.20 keV for tin, 57.48 keV for erbium, 67.43 keV for tantalum, and 88.01 keV for lead. This type of k-edge filtration leads to diverse sensitivity spectra which help to separate the signals from water, iodine, and gadolinium, based on their respective mass attenuation spectra.

B. Multi-Contrast Imaging Experiment

The spectral CT scan of the multi-contrast phantom acquired with the spatial-spectral filter is shown in Figure 5. The linear motion of the filter throughout the scan results in the diagonal stripes. Note that any individual spectral channel corresponding to a certain filter material does not make up a complete CT dataset. We only achieve sufficient spatial and spectral sampling by considering all the data together. In total, 3% of the projection data were identified as part of the interface region between filter tiles and were disregarded in the statistical reconstruction. In Figure 5, those projections have been replaced via interpolation between neighboring views.

We can see the projection of the water tank as well as the lower-contrast projection signal of the iodine and gadolinium inserts. Based on the sensitivity spectra for a given projection, the same mass-length of a certain material will result in a different contrast. Only by considering this data together with a model for the system geometry and projection-dependent spectral sensitivity can we estimate the basis material densities.

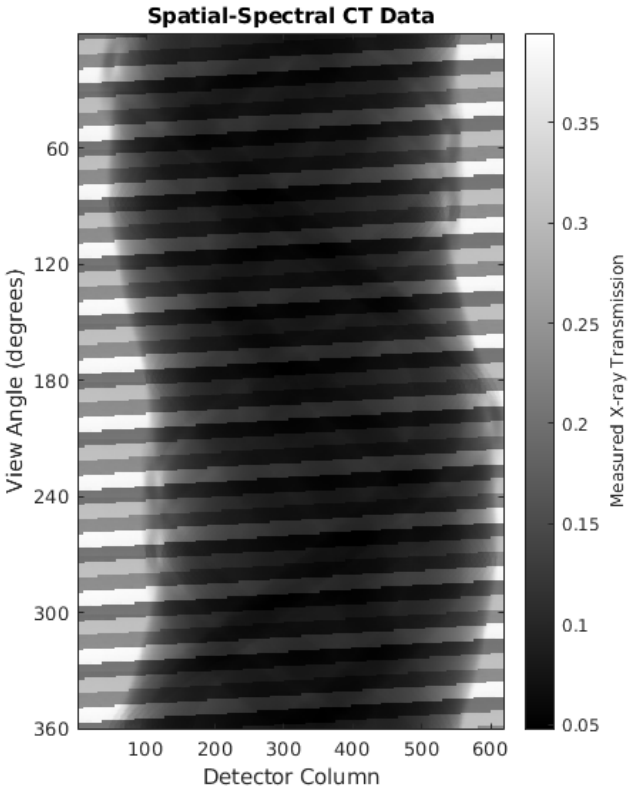


Fig. 5. Spatial-spectral CT data of a multi-contrast phantom scan.

C. Model-Based Material Decomposition

The output basis material density images from the model-based material decomposition are shown in Figure 6. The figure also shows overlays of ROI mean values which demonstrate a very close match to the ground truth concentrations for each material.

The water channel shows a relatively flat density with the exception of the plastic vials and the plastic rectangular fiducials on the outside of the tank, which are used to verify the orientation. The mean value of the flat response in the water tank is around 992 mg/mL which is close to the ground truth value of 1000 mg/mL. Some of the attenuation is falsely attributed to an elevated baseline in the gadolinium channel of around 0.78 mg/mL for regions of the phantom that should only contain water. This likely indicates some errors in the spectral calibration.

For vials containing a mixture of both gadolinium and iodine, the resulting decomposition appears to over-estimate the iodine and under-estimate the gadolinium. This is demonstrative of the strong anti-correlation between materials which poses a great challenge for material decomposition problems.Again, these errors could be a bias due to small errors in the modeled system sensitivity spectra. Finally, are some artifacts the material boundaries which could be caused by cross-talk from the quadratic regularization penalty or by bias from unmodeled spectral effects.

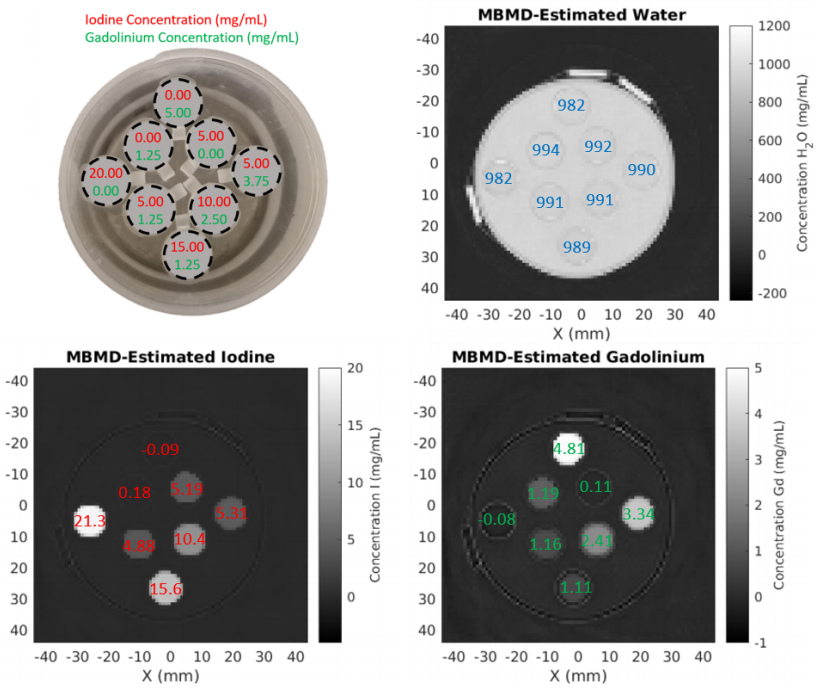


Fig. 6. Multi-contrast phantom with vials containing iodine and gadolinium solutions at various mixtures in water. Material density images estimated via model-based material decomposition.

IV. Conclusion

The hardware designs, data processing procedures, and experimental results presented here show that spatial-spectral filters show promise as a strategy for spectral CT with energy integrating detectors. While spatial-spectral filters yield sparse projection data for each spectral channel, these results shown that we can use one-step material decomposition algorithms to reconstruct basis material density images using the proper calibration and physical models. Morever, using these spatial-spectral filters we were able to conduct a multi-contrast imaging experiment using spectral-modulation of the x-ray beam without the need for energy-discriminating detectors.

Spectral CT shows promise for many clinical applications. As we develop new technologies which increase the number of spectral channels, we have the potential enable entirely new clinical studies. For example, three or more spectral channels enables three or more basis materials which is the fundamental requirement for multi-contrast imaging and quantification (one of the basis materials being water). Creative new clinical studies for multiple contrast agents are already under development for liver imaging, and it is reasonable to expect more new applications in the future.

In ongoing work, we seek to refine procedures to calibrate gaps and edges between filters for full data utilization. We will also aim to investigate combining spatial-spectral filters with other techniques like kVp-switching or energy-discriminating detectors to explore high-sensitivity spectral CT.

Acknowledgements

This work was supported, in part, by NIH grants R21EB026849.

References

[1] Boris Buerke, Gunnar Wittkamp, Harald Seifarth, Walter Heindel, and Stephan P Kloska. Dual-energy cta with bone removal for transcranial arteries: intraindividual comparison with standard cta without bone removal and tof-mra. *Academic radiology*, 16(11):1348–1355.

[2] Jiří Ferda, Milan Novák, Hynek Mírka, Jan Baxa, Eva Ferdová, Alena Bednářová, Thomas Flohr, Bernhard Schmidt, Ernst Klotz, and Boris Kreuzberg. The assessment of intracranial bleeding with virtual unenhanced imaging by means of dual-energy ct angiography. *European radiology*, 19(10):2518–2522, 2009.

[3] Cynthia H McCollough, Shuai Leng, Lifeng Yu, and Joel G Fletcher. Dual-and multi-energy ct: principles, technical approaches, and clinical applications. *Radiology*, 276(3):637–653, 2015.

[4] Daniela Muenzel, Heiner Daerr, Roland Proksa, Alexander A Fingerle, Felix K Kopp, Philippe Douek, Julia Herzen, Franz Pfeiffer, Ernst J Rummeny, and Peter B Noël. Simultaneous dual-contrast multi-phase liver imaging using spectral photon-counting computed tomography: a proof-of-concept study. *European radiology experimental*, 1(1):25.

[5] Katsuyuki Taguchi and Jan S Iwanczyk. Vision 20/20: Single photon counting x-ray detectors in medical imaging. *Medical physics*, 40(10), 2013.

[6] Steven Wayne Tilley II, Wojciech Zbijewski, and Joseph Webster Stayman. Model-based material decomposition with a penalized nonlinear least-squares ct reconstruction algorithm. *Physics in medicine and biology*, 2018.

[7] Matthew Tivnan, Wenying Wang, and J. Webster Stayman. Preconditioned algorithms for model-based iterative reconstruction and material decomposition from spectral ct data. 2020.

[8] Matthew Tivnan, Wenying Wang, Steven Tilley II, Jeffrey H Siewerdsen, and J Webster Stayman. Optimized spatial-spectral ct for multi-material decomposition. In *Fully 3D*, volume 11072, page 1107211, 2019.

[9] W Wang, M Tivnan, G Gang, S Tilley, and J Stayman. Generalized local impulse response prediction in model-based material decomposition of spectral ct. In *Medical Physics*, volume 46, pages E257–E257.

[10] Lifeng Yu, Jodie A Christner, Shuai Leng, Jia Wang, Joel G Fletcher, and Cynthia H McCollough. Virtual monochromatic imaging in dual-source dual-energy ct: radiation dose and image quality. *Medical physics*, 38(12):6371–6379, 2011.

Dynamic Fluence Modulation using Proton CT for Low–dose Imaging in Particle Therapy

Jannis Dickmann, Christina Sarosiek, George Coutrakon, Simon Rit, Nick Detrich, Victor Rykalin, Mark Pankuch, Robert P. Johnson, Reinhard W. Schulte, Katia Parodi, Guillaume Landry, and George Dedes

***Abstract*—Dynamic fluence modulation for computed tomography (CT), i.e. the acquisition of tomographic images with variable, patient- and task-specific fluence fields, offers the potential to substantially reduce local imaging dose. In particular, volume-of-interest (VOI) imaging allows to limit imaging dose to a clinically relevant volume and reduce it elsewhere. In the context of particle therapy, where tomographic data is required for treatment planning the VOI is the treatment beam path. VOI imaging is of particular interest for particle therapy given the very low integral out-of-VOI treatment dose. Proton CT imaging allows for a direct measurement of the proton stopping power with an increased accuracy and a decreased imaging dose compared to x-ray-based CT. In addition, frequent imaging is required to verify patient positioning and to monitor potential anatomical changes, which over the course of a treatment may compromise the planned dose. In this work, we evaluate the performance of a fluence-modulated proton CT algorithm for low-dose in-room imaging. This would allow for recalculation or replanning of the treatment dose according to the anatomy of the day with out-of-VOI dose below 1 mGy. We performed a simulation study and acquired experimental data using a prototype proton CT scanner. By employing a bow-tie-like fluence modulation aiming for constant noise, imaging dose was reduced by 9%. For a VOI imaging task, out-of-VOI dose was reduced by 41% and substantially below 1 mGy. This may pave the way for daily imaging prior to every treatment session aiming to eventually reduce safety margins in particle therapy, thus further reducing normal tissue exposure to therapeutic doses.**

I. Introduction

IMAGE-GUIDED particle therapy requires an accurate volumetric representation of the proton (or heavier ion) stopping power for calculation and optimization of the therapeutic dose. Ideally, images are acquired at minimal imaging dose prior to each treatment fraction and in treatment position to avoid suboptimal tumor coverage due to positioning errors or anatomical changes. The current clinical practice is to use single or (less frequently) dual energy x-ray computed tomography (xCT) and a subsequent calibration mapping photon attenuation to the relative (to water) stopping power (RSP). This procedure leads to intrinsic errors of approx. 1% (dual energy) to 3% (single energy) in RSP [1]. To account for such uncertainties, additional safety margins are added to the planning target volume, which necessarily results in a higher dose delivered to healthy tissue or even organs-at-risk during treatment. Furthermore, dual energy xCT imaging, when available, cannot be easily performed in treatment position, and, due to timing and imaging dose constraints, is not performed prior to every treatment fraction at most centers.

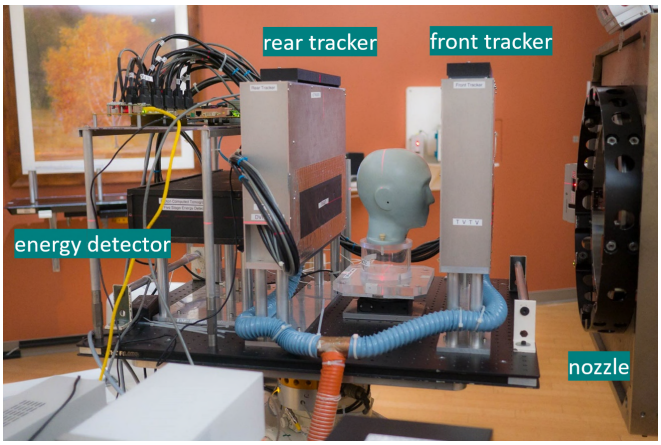


Fig. 1. Prototype proton CT scanner mounted on top of a robotic arm in front of the treatment beam nozzle. Two tracking detectors measure positions and directions of each proton and an energy detector measures its residual energy.

Alternatively, a direct determination of the RSP is possible using proton computed tomography (pCT) by measuring the energy loss of high-energy (e.g. 200 MeV for cranial indications) protons traversing the patient, as originally proposed by Cormack [2]. The energy loss of a single proton is directly linked to a line integral through RSP, which can therefore be reconstructed using iterative or analytical image reconstruction accounting for the curved path of the protons [3]. RSP accuracy of images acquired using the pre-clinical prototype pCT scanner [4], [5] shown in fig. 1 is reported to be competitive to state-of-the-art dual energy xCT with optimal spectral separation [6]. At the same time, future pCT scanners could readily be rotated with the proton gantry and images be

J. Dickmann, Prof. Dr. K. Parodi, Prof. Dr. G. Landry, and Dr. G. Dedes are with the Ludwig-Maximilians-Universität München (LMU Munich), Germany. Prof. Dr. G. Landry is also with the University Hospital, LMU Munich, Germany and the German Cancer Consortium (DKTK), Munich, Germany. C. Sarosiek and Prof. Dr. G. Coutrakon are with the Northern Illinois University, USA. Dr. S. Rit is with CREATIS and University of Lyon, France. N. Detrich. Dr. V. Rykalin and Dr. M. Pankuch are with the Northwestern Medicine Chicago Proton Center, USA. N. Detrich is also with IBA, Louvain-La-Neuve, Belgium. Dr. V. Rykalin is also with Proton VDA Inc., Naperville, USA. Prof. Dr. R. P. Johnson is with University of California, Santa Cruz, USA. Prof. Dr. R. W. Schulte is with Loma Linda University, USA.

Corresponding author: jannis.dickmann@lmu.de.

acquired in treatment position just before treatment. Moreover, dose efficiency is improved compared to xCT [7] and typical pCT imaging doses are at 1 to 2 mGy per tomography and therefore at the level of cone-beam xCT, yet with the dosimetric accuracy of dual energy xCT. The rate of double-strand DNA breaks by pCT may also be reduced compared to xCT at equivalent physical dose [8], [9]. This makes pCT a good candidate for frequent pre-treatment recalculation and optional replanning of the therapeutic dose, without accumulation of critical imaging doses.

A further reduction of imaging dose can be achieved by using dynamic fluence field modulation, originally proposed for xCT [10], [11]. Fluence-modulated imaging aims at inhomogeneous and changing spatial distributions of radiation fluence within one projection, thus concentrating imaging dose within a volume-of-interest (VOI), where image noise is low. Elsewhere, image noise is elevated, but imaging dose reduced. This is particularly meaningful in the context of particle therapy, where low noise (and high RSP accuracy) is only required within the treatment beam path (i.e. the VOI) and image noise (and RSP accuracy) is irrelevant elsewhere. At the same time, imaging dose outside the VOI must be kept low to maintain the low dose to normal tissue achievable with proton therapy – and is irrelevant inside the VOI, where therapeutic dose dominates. While experimental implementation of fluence modulation in xCT is challenging [12], [13], the incident proton fluence can readily be modulated by scanning the object with proton pencil beams using the treatment beam delivery system and modulating the dwell time or beam current of each pencil beam. The typical size of a proton pencil beam is about 10 mm (FWHM), which allows for a fine modulation of the fluence.

In this work, we evaluate a previously presented optimization algorithm for fluence-modulated pCT (FMpCT) [14] for its ability to reduce imaging dose and achieve prescribed image noise distributions. We employ both a Monte Carlo simulation of the imaging set-up and experimental data.

II. Materials & Methods

A. Image noise for proton CT

The pCT prototype scanner investigated in this work records position, direction and residual energy information for every proton incident to the detector – an operation mode called "list-mode." Variance in list-mode pCT projections is proportional to the number of particles per pixel in the projection (referred to as "counts"), because the projection value is the average water-equivalent path length (WEPL) of a set of protons incident to each pixel. For a given number of counts, noise is primarily affected by the protons' stochastic loss of energy, known as "energy straggling," in the object and also in the detector. Moreover, multiple Coulomb scattering (MCS) of protons and the accuracy of the proton path estimate through tracking impact noise. Finally, the energy spread of the accelerator can increase noise. All these contributions were quantified in a previous publication [15] and can be predicted precisely using a Monte Carlo simulation of the prototype pCT scanner shown in fig. 1 and used in this study. It is important to note that MCS is object-dependent and heterogeneities strongly increase noise, distorting the bijective relation between line integral values and noise levels which can be expected in xCT. Therefore, any fluence modulation must take into account prior knowledge about the object to be imaged, for example from a previous diagnostic xCT or pCT scan, or even MRI converted to xCT [16]. Given noise levels in the projection, image noise can be calculated using variance reconstruction [17], [18], which is an error propagation through the linear reconstruction operation of filtered backprojection.

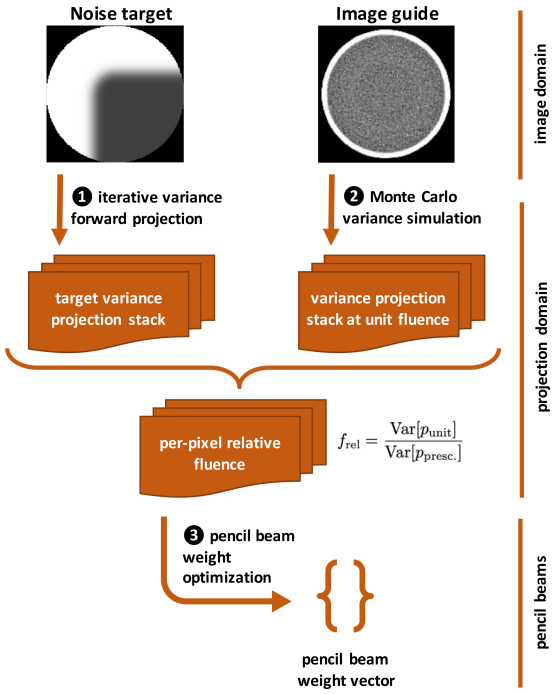


Fig. 2. Three-step projection-based optimization workflow for fluence-modulated pCT: given an image noise target and an image guide, the algorithm calculates a pencil beam weight vector that results in the desired image noise target for a given object.

B. Fluence optimization algorithm

The fluence-modulation optimization algorithm presented in [14] calculates a set of pencil beam weights (i.e. their relative dwell time) that achieve a prescribed image noise target. The workflow is depicted in fig. 2 and consists of three steps.

At first, an iterative variance forward projection is employed, that finds a strictly positive (and thus physical) stack of variance projections $\mathrm{Var}[p_{\mathrm{presc.}}]$, resulting in the prescribed image noise target when input to the variance reconstruction.

In a second step, a Monte Carlo simulation is used to generate a set of variance projections $\mathrm{Var}[p_{\mathrm{unit}}]$, which correspond to imaging the object at some unit fluence (e.g. uniform fluence). This incorporates prior knowledge about the object to be imaged. Since variance is inversely proportional to the number of counts, the required fluence modulation (relative to unit fluence) is calculated as $f_{\mathrm{rel}} = \mathrm{Var}[p_{\mathrm{unit}}] / \mathrm{Var}[p_{\mathrm{presc.}}]$, which would already be sufficient if it were possible to modulate fluence pixel-by-pixel.

A third step therefore includes knowledge about the finite extent of each pencil beam. A linear combination of a set of pre-

defined pencil beams is calculated, such that the summed fluence is equal to

$$F_{\mathrm{FMpCT}} = F_{\mathrm{unit}} \cdot f_{\mathrm{rel}} = F_{\mathrm{unit}} \cdot \mathrm{Var}[p_{\mathrm{unit}}] / \mathrm{Var}[p_{\mathrm{presc.}}], \quad (1)$$

where F_{unit} is the unit fluence of all pencil beams. This optimization problem is solved by minimizing the summed squared pixel-wise deviation between eq. (1) and a linear combination of pencil beams. The pencil beam model for this optimization (spot size, divergence) was derived from experimental data. Pencil beam centers were chosen to be on a regular grid as shown in the last row of fig. 3 (a). The large spacing in horizontal direction was compensated by shifting the fluence pattern by a quarter of this spacing, resulting in a homogeneous fluence when summing counts of two opposing projections (analogous to the quarter detector shift in xCT).

C. *Simulation study*

A Monte Carlo simulation study was performed using a dedicated Geant4 simulation platform [19], which we recently validated for its accuracy in predicting RSP noise [15]. A water filled cylindrical PMMA container (diameter 150.5 mm) was simulated, for which a physical counterpart existed. Three scenarios were investigated: (1) acquisition with uniform fluence (i.e. unit fluence), which is current standard for pCT acquisitions; (2) prescription of constant noise throughout the water cylinder; and (3) an FMpCT acquisition with a VOI in one quadrant mimicking the beam path of a two-field treatment with therapeutic beams from 90 and 180 degrees. Fluence patterns were optimized using the algorithm in section II.B. Imaging doses were scored for every projection and summed.

D. *Experimental acquisitions*

Experimental data were acquired using the pCT prototype scanner [4] shown in fig. 1 and located at a fixed beam line of the Northwestern Medicine Chicago Proton Center, which is capable of pencil beam scanning. The object was rotated for a full 360 degrees during the acquisition. An interface to the

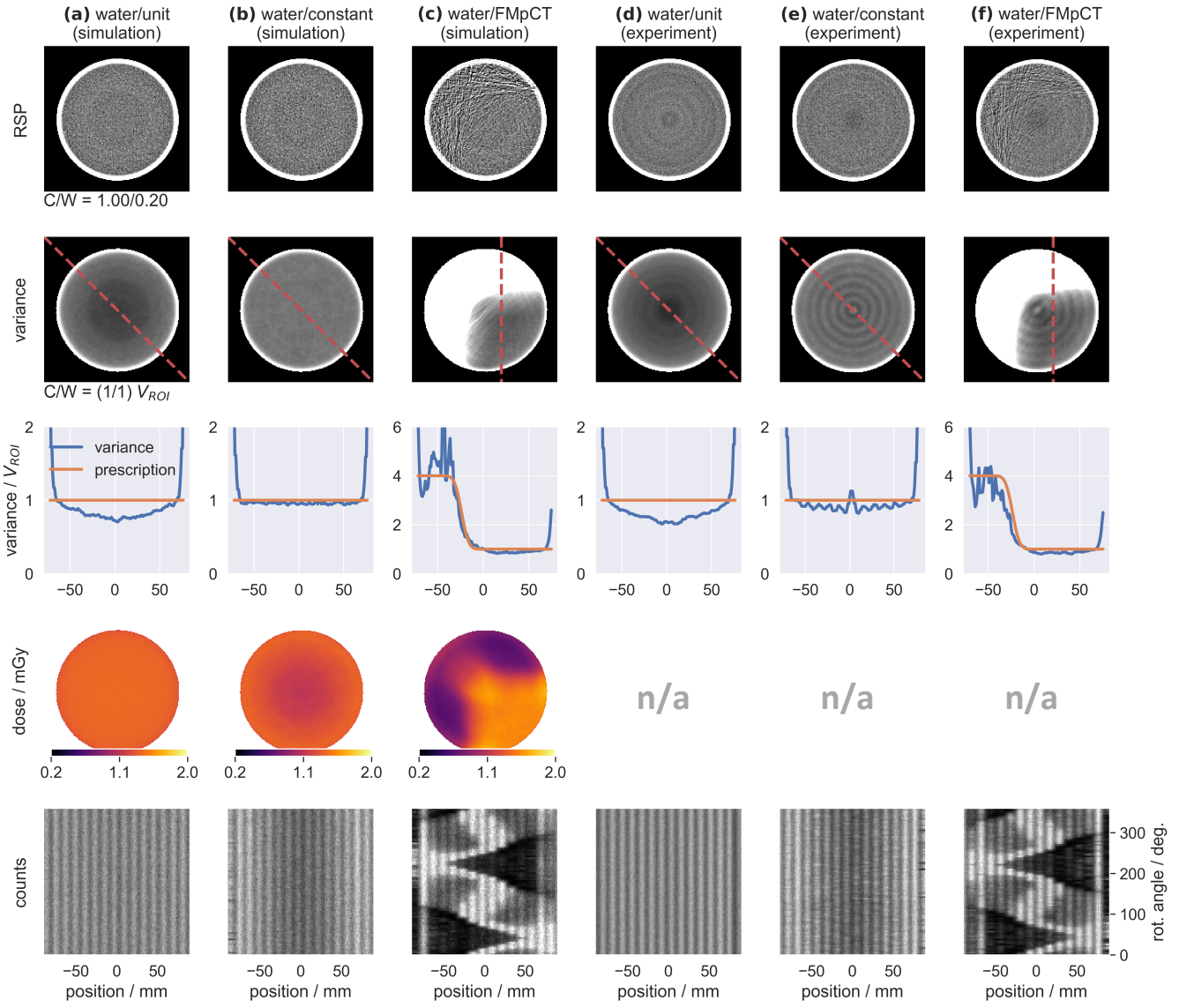


Fig. 3. Simulation (a, b, c) and experimental (d, e, f) pCT acquisitions for unit fluence (a, d), a constant noise (b, e) and an FMpCT noise prescription (c, f). RSP, variance, simulated imaging dose and counts sinograms are displayed.

beam control system was developed to automatically deliver fluence patterns. The same water filled phantom as in the simulation study was used. Noise targets and optimized fluence patterns were also identical.

III. Results & Discussion

Results are summarized in fig. 3, where RSP maps, the variance reconstruction, profiles through the variance reconstruction (indicated by the dashed line), imaging doses and counts sinograms are shown. Simulation study results are in fig. 3 (a) to (c) and experimental results in fig. 3 (d) to (f).

In fig. 3 (a) unit fluence was delivered as can be seen in the counts sinogram. This resulted in a reduced image variance in the center of the object, which is expected due to multiple Coulomb scattering and was discussed in a previous publication [15]. Imaging dose is constant throughout the object because the energy loss for high energy protons (far from the Bragg peak) is relatively constant. Fig. 3 (b) shows an optimized fluence plan aiming for constant variance in the image, which was successful as can be seen in the variance map and profile. While the change in the RSP image is not directly obvious, the counts sinogram shows reduced fluence in the center of the object, which at the same place resulted in a reduced dose. This is the equivalent of a bow-tie filter in xCT, but the fluence profile is fundamentally different due to different noise characteristics in pCT. At equal peak noise level, dose was reduced by 9% using fluence modulation, compared to the unit fluence scan. The effect of the FMpCT plan in fig. 3 (c) can already be appreciated in the RSP map. The variance map and profile show that fluence modulation was successful and variance follows the prescription with good accuracy. At equal peak noise level in the VOI, dose outside of the VOI was reduced by 41%, which is a substantial dose saving. In [14], an anthropomorphic head phantom was also studied showing similar dose savings for a more heterogeneous geometry.

Experimental data in fig. 3 (d) to (f) agree well with results from the simulation study. Unfortunately, the alignment of the scanner with respect to the coordinate system of the beam line was not perfectly reproduced in between the unit fluence acquisition and the two fluence modulations. This caused the quarter shift of the pencil beam pattern to be spoiled and resulted in rings in the variance map due to changing proton statistics. Except for this minor disagreement, the variance maps using fluence modulation were as expected from simulations. While variance rings partially impaired the constant noise target in fig. 3 (e), the FMpCT acquisition in fig. 3 (f) agreed well with the prescription, and the prescribed variance difference was larger than the ringing caused by the misalignment. Imaging doses were not determined for the experimental dataset, but dose savings can be expected to be comparable to those determined in the simulation study, especially given that fluence sinograms were very similar.

IV. Conclusion

In this work, we generated fluence-modulated pCT delivery patterns using a dedicated optimization algorithm for FMpCT. Patterns were calculated for a simple water-filled cylinder and for two noise prescriptions: constant noise throughout the image as an equivalent to xCT bow-tie filters, and a volume-of-interest imaging task with application for particle therapy. The performance of the optimization algorithm was evaluated both in a simulation study and in an experiment using a prototype pCT scanner. Dose savings in the simulation study were substantial, highlighting the potential use of fluence modulation in pCT. Experimental feasibility was demonstrated but leaves room for technical improvements not related to the optimization algorithm.

Acknowledgment

This work was supported by the German Research Foundation (DFG) under grant #388731804. Additional support from the Munich-Centre for Advanced Photonics, the Bavaria-California Technology Center and the Franco-Bavarian university cooperation center is gratefully acknowledged.

References

[1] H. Paganetti, “Range uncertainties in proton therapy and the role of Monte Carlo simulations,” *Phys. Med. Biol.*, 57, 11, R99–R117, 2012.
[2] A. M. Cormack, “Representation of a Function by Its Line Integrals, with Some Radiological Applications,” *J. Appl. Phys.*, 34, 9, 2722–2727, 1963.
[3] S. Rit *et al.*, “Filtered backprojection proton CT reconstruction along most likely paths,” *Med. Phys.*, 40, 3, 031103, 2013.
[4] R. P. Johnson *et al.*, “A Fast Experimental Scanner for Proton CT: Technical Performance and First Experience With Phantom Scans,” *IEEE Trans. Nucl. Sci.*, 63, 1, 52–60, 2016.
[5] V. A. Bashkirov *et al.*, “Novel scintillation detector design and performance for proton radiography and computed tomography,” *Med. Phys.*, 43, 2, 664–674, 2016.
[6] G. Dedes *et al.*, “Experimental comparison of proton CT and dual energy x-ray CT for relative stopping power estimation in proton therapy,” *Phys. Med. Biol.*, 64, 16, 165002, 2019.
[7] R. W. Schulte *et al.*, “Density resolution of proton computed tomography,” *Med. Phys.*, 32, 4, 1035–1046, 2005.
[8] S. Meyer *et al.*, “Dosimetric accuracy and radiobiological implications of ion computed tomography for proton therapy treatment planning,” *Phys. Med. Biol.*, 2019.
[9] L. Yasui *et al.*, “DNA Double-Strand Break Induction and Repair of Proton Computed Tomography in Normal Human Cells,” in *RSNA Annual Meeting*, 2019.
[10] S. A. Graham *et al.*, “Intensity-modulated fluence patterns for task-specific imaging in cone-beam CT,” in *Proceedings of SPIE*, 2007, 651003.
[11] S. Bartolac *et al.*, “Fluence field optimization for noise and dose objectives in CT,” *Med. Phys.*, 38, 7, S2–S17, 2011.
[12] J. W. Stayman *et al.*, “Fluence-field modulated x-ray CT using multiple aperture devices,” *Proceedings of SPIE*, 9783, 97830X, 2016.
[13] S. Huck *et al.*, “Optimized intensity modulation for a dynamic beam attenuator in x-ray computed tomography,” in *Proceedings of SPIE*, 2019, 1094824, p. 75.
[14] J. Dickmann *et al.*, “An optimization algorithm for dose reduction with fluence-modulated proton CT,” *Med. Phys.*, 47, 4, 1895–1906, 2020.
[15] J. Dickmann *et al.*, “Prediction of image noise contributions in proton computed tomography and comparison to measurements,” *Phys. Med. Biol.*, 64, 14, 145016, 2019.
[16] S. Neppl *et al.*, “Evaluation of proton and photon dose distributions recalculated on 2D and 3D Unet-generated pseudoCTs from T1-weighted MR head scans,” *Acta Oncol.*, 58, 10, 1429–1434, 2019.
[17] A. Wunderlich and F. Noo, “Image covariance and lesion detectability in direct fan-beam x-ray computed tomography,” *Phys. Med. Biol.*, 53, 10, 2471–2493, 2008.
[18] M. Rädler *et al.*, “Two-dimensional noise reconstruction in proton computed tomography using distance-driven filtered back-projection of simulated projections,” *Phys. Med. Biol.*, 63, 21, 215009, 2018.
[19] V. Giacometti *et al.*, “Software platform for simulation of a prototype proton CT scanner,” *Med. Phys.*, 44, 3, 1002–1016, 2017.

Author List

A

Abella Monica 474
Abella Mónica 118
Aichert André 392
Akino Naruomi 408
Alves Frauke 608
Amato Carlo 280
Anastasio Mark A. 618
Andreyev Andriy 562
Anjomrouz Marzieh 554

B

Badali Daniel S. 178
Badea Cristian T. 182, 284
Baek Jongduk 626
Baek Sumin 510
Bai Chuanyong 566
Bai Mei 578
Balles A. 514
Bao Yuan 158
Bazrafkan Shabab 380, 486, 518
Beister Marcel 582
Benig Lily 554
Bennett N. Robert 404, 574
Berdon Patricia M. 474
Berker Yannick 590
Betcke Marta M. 328
Beuing Oliver 348
Bian Zhaoying 134, 138
Birkhold Annette 416
Bismark Richard N. K. 348
Bliznakova Kristina 442
Blum Nele 396
Bochynek Thomas 392
Boone John 194, 442
Brambilla Andrea 272
Bresler Yoram 308
Briel Andreas 280
Brunner Stefan E. 430
Brönnimann Christian 280
Butler Anthony 554
Butzer Jochen 94
Buzug Thorsten M. 396, 470, 482
Byl Achim 256

C

Cai Jianmei 364
Cao Qian 558
Cao Wenchao 260
Capostagno S. 236
Carlsson Gudrun Alm 506
Chang Shaojie 376
Chattaraj Rajarshi 170
Chen Baiyu 30, 332
Chen Buxin 534, 542
Chen Xi 376
Chen Zhiqiang .. 190, 268, 304, 324, 384, 598
Chernoglazov Alex 554
Cheslerean-Boghiu Theodor 166
Cho Hyuncheol 198
Cho Minkook 130
Cho Seungryong 320, 356, 602
Choi Da-in 356
Choi Jiyoung 130
Chouzenoux Emilie 454
Cierniak Robert 586
Clackdoyle Rolf 46, 526, 530
Clark Darin P. 182, 284
Claus Bernhard 150
Colbeth Richard E. 62, 404
Cossairt Oliver 392
Coussat Aurélien 526
Coutrakon George 642
Cruz-Bastida Juan P. 618

D

De Beenhouwer Jan ... 90, 154, 380, 486, 518
De Cecco C. N. 292
Dedes George 642
Defrise Michel 526
den Harder Johan Michiel 622
Deng Xiaojuan 42
Deng Yifan 66
Dennerlein Frank 264

Desbat Laurent 46, 526, 530
Desco Manuel 118, 474
Detrich Nick 642
di Franco Francesca 442
Diakite Daouda 490
Dickmann Jannis 642
Dierolf Martin 336
Dirckx Joris J.J. 90
Dittmann J. Graetz 514
Djurabekova Nargiza 328
Donath T. 514
Dreier Mark 470
Du Muge 232
Duan Jiayu 364
Döring Annika 94

E
Eckermann Marina 106
Effland Alexander 332
Ehtiati T. 236
Elhamiasl Masoud 368
Endrizzi Marco 162
Erath Julien 412
Eulig Elias 252, 372
Evans Connor J. 462

F
Fahrig Rebecca 416
Fang Chengyijue 434
Fang Wei 598
Fella C. 514
Felsner Lina 502
Fessler Jeffrey A. 118, 224
Fournier Clarisse 272
Fournié Eric 252, 412
Franke Danielle 280
Frenkel Michael 122, 186
Frohn Jasper 608
Fuller Michael 612
Fulton Roger 248
Fung George S.K. 214
Funk Tobias 178

G
Gac Nicolas 490
Gagnon Daniel 566
Gang Grace J. 400, 466, 630
Gao Hao 352
Gao Hewei 50, 66, 190, 232, 324, 574
Gao Qi 138
Gao Yongfeng 38, 296
Geagan Michael 170
Gehrke Nicole 280
Gilat-Schmidt Taly 218, 522
Gindrier Nicolas 46
Gkoumas Spyridon 280
Gnudi Paolo 590
Goldberg Andrew 328
Gondrom-Linke Sven 94
Gong Zhaowen 202
Gottschalk Tristan M. 276
Graf vom Hagen Christoph 562
Gullberg Grant T. 612
Guo Zhitao 288
Gutjahr Ralf 174

H
Hadhazi Daniel 494
Hagen Charlotte K. 86, 162
Han Minah 626
Han R. 634
Han Yoseob 340
Haneda Eri 150
Haninger Stefan 142
Hansen Lasse 482
Harbecke Markus 94
Hardt Jennifer 256
Hawkes David 328
He Liuchun 158
Heinrich Mattias P. 482
Heinze Sarah 256
Herbst Magdalena 210, 582
Hernandez Andrew M. 442
Hernandez-Giron Irene 622
Heylen Rob 244
Hoang A-T. 518
Hofmann Christian 396
Holbrook Matt D. 182
Holmes T. W. 292
Horvath Gabor 494
Hou Wei 296
Huang Jing 134
Huang Qiu 612
Huang Xiaokun 550
Huck Sascha Manuel 214
Hwang Jaehong 356

I
Iskender Berk 308

J
Jain Amit 566
Jang Seokhwan 126
Jang Tae Jun 388
Jang Taejun 198
Jensen Tor 546

Jeon Kiwan 206
Jeong Uijin 320
Jia Xun 538
Jiang Xiao 316
Jin Xin 268
Jo Hwanhui 320, 602
Johnson Robert P. 642
Jolivet Frédéric 272
Jordan P. 218
Jung Hyunuk 538
Juntunen Mikael 570

K

Kachelrieß Marc 50, 146, 252, 256, 280, 344, 372, 412, 458, 590
Kappler Steffen 210, 582
Karp Joel S. 300
Katsevich Alexander 122, 186
Kazantsev Daniil 450
Khan Amara 608
Kim Byeongjoon 626
Kim Doil 126
Kim Ho Kyung 110
Kim Jinwoo 110
Kim Junyoung 340
Kim Kangcheol 198
Kim Sally J. 612
Klein Laura 256, 280, 458
Klein Philipp 482
Kleinszig G. 634
Knaup Michael 372
Kobler Erich 30, 332
Koch Philipp 470
Kolehmainen Ville 570
Konings Oleg 178
Korwarschik Markus 416
Kreher Björn W. 276, 634
Kunze H. 634
Kwon Taejin 356

L

Landry Guillaume 642
Lasser Tobias 142, 166, 336, 478
Lebedev Sergej 252
Lee Daeyeon 170
Lee Donghyeon 320, 602
Lee Duhgoon 130
Lee Inje 356
Lee Jaesung 356
Lee Kyoung-Yong 126, 130
Lee Okkyun 114, 510
Lee Sang-Hwy 388
Lee Sung Min 388
Lee Sungmin 198
Lehmann Mathias 240
Lesaint Jérôme 272
Li Danyang 134, 138
Li Hongwei 42, 360
Li Junyuan 62
Li Liang 268, 598
Li Mengfei 352
Li Mengzhou 426
Li Qin 224
Li Sui 134, 138
Li Yang 364
Liang Kaichao 232, 304, 384
Liang Wuyang 268
Liang Zhengrong 38, 296
Lim Jihwan 356
Lim Sunho 602
Lin Claire Yilin 224
Liu Leening P. 170
Liu Stephen Z. 558
Liu Yanyan 78
Liu Yinong 232
LoBato Denae 554
Long Guy 328
Long Wenting 102
Long Yong 228
Lowe Chiara 554
Lu Jiang 224
Lu Minghui 62, 404
Lu Siming 38
Lu Yaguang 578
Létang Jean Michel 526

M

Ma Genwei 98, 360
Ma Jianhua 134, 138
Ma Yiqun Q. 62, 400
Maaß Nicole 264
Magnusson Maria 506
Maier Andreas K. 174, 264, 276, 416, 502
Maier Joscha 252, 280, 372, 412
Malusek Alexandr 506
Man Bruno De 150
Martelli Maxime 490
Martínez Cristóbal 118
Maslowski A. 218
Matsuura Masakazu 408
Maughan-Jones Charlotte 86
Meaney Alexander 570
Mei Kai 300
Mertins Alfred 470
Messmer Philippe 240

Mettivier Giovanni 442
Michielsen Koen 438
Moghiseh Mahdieh 554
Moitra Parikshit 546
Moriakov Nikita 438
Morton Ed 186
Mou Xuanqin 38, 344, 364, 376
Muller Sarah 372
Munro Peter 122, 240
Myers Kyle J. 224
Müller D. 514

N
Nam Jeonghyeon 114
Nauwynck M. 486
Nguyen Hung 530
Nguyen Van 90, 518
Noo Frédéric 34
Noël Peter B. 170, 260, 300, 630
Nuyts Johan 244, 248, 368

O
Olivo Alessandro 86, 162
Ongie Greg 618
Ostadhossein Fatemeh 554
Otazo Ricardo 30, 332

P
Pack Jed D. 150
Pan Dipanjan 546, 554
Pan Huiying 54
Pan Xiaochuan 522, 534, 542, 618
Pankuch Mark 642
Panta Raj 554
Park Hyoung Suk 206
Park Junyoung 130
Parodi Katia 214, 642
Patwari Mayank 174
Paul Emily R. 522
Paysan Pascal 122, 146, 240
Pekel Erdal 336
Pelt Daniël M. 86
Perkins Amy E. 300
Pesquet Jean-Christophe 454
Peterlik Igor 122, 146, 240
Petersilka Martin 412
Petra Stefania 458
Pfeiffer Franz 166, 336
Piol Alessandro 474
Plamondon Mathieu 240
Pluta Piotr 586
Pock Thomas 332
Pourmorteza A. 292
Preuhs Alexander 416, 502
Principi S. 218

Q
Quan Guotao 78, 82

R
Rahman Obaidullah 462
Raja Aamir 554
Raupach Rainer 174
Ravishankar Saiprasad 228
Rawlik Michał 446
Ray Priyanka 546
Reddin Janet 300
Reichardt Marius 608
Reiser Ingrid S. 618
Ren Yan 102
Renders Jens 154
Riddell Cyril 454
Riess Christian 502
Rit Simon 46, 58, 526, 642
Ritschl Ludwig 210, 582
Rivas Sara Teruel 368
Robisch Anna-Lena 106
Roche i Morgó Oriol 86, 162
Rodesch Pierre-Antoine 58
Roeder Ryan K. 462
Rose Georg 348
Roser Philipp 416, 502
Rotkopf Lukas 458
Rubenstein Michael 392
Rundle David S. 426
Russ Tom 400
Russo Paolo 442
Rykalin Victor 642

S
Salditt Tim 106, 608
Salvo Koen 368
Sanctorum Joaquim G. 90
Sandborg Michael 506
Santeler Scott 546
Sarno Antonio 442
Sarosiek Christina 642
Sauer Ken 462
Savanier Marion 454
Sawall Stefan 252, 256, 280, 372, 458
Schaart Dennis R. 430
Schad Lothar R. 400
Scheuerman Joshua 300
Schiffers Florian 392
Schlemmer Heinz-Peter 256, 458
Schmitt Michael H. 498

Schulte Reinhard W. 642
Schweizer Bernd 590
Sechopoulos Ioannis 438
Seghers Dieter 122, 146, 240
Sehgal Chandra 170
Seo Jin Keun 198, 388
Seo Youngho 612
Shapira Nadav 170, 300, 630
Shechter Gilad 312
Shen Chenyang 538
Shen Le 304
Shepherd Luke 328
Sheth N. 634
Shi Daxin 70
Shi Linxi 404
Shi Yongyi 38, 296
Shiroma Amy 404
Si-Mohamed Salim 58
Sidky Emil Y. 522, 534, 542, 618
Siewerdsen Jeffrey H. 194, 210, 236, 400, 558, 634
Sijbers Jan 90, 154, 380, 486, 518
Siltanen Samuli 570
Simon Sven 202
Sisniega A. 236, 634
Sitek Arkadiusz 562
Sodickson Aaron 30
Sodickson Daniel K. 30, 332
Son Kihong 126
Soons Joris 380
Stadelmann Christine 106
Stampanoni Marco 446
Star-Lack Josh 62, 404
Stayman J. Webster .. 62, 236, 400, 466, 558, 630, 638
Sterpin Edmond 368
Stierstorfer Karl 214, 252, 412
Stille Maik 396, 470, 482
Streekstra Geert J. 622
Strobel Norbert 416
Strzelecki Adam 122, 146, 240
Sun Kaicong 202
Sun Mingshan 404
Sung Yongjin 612
Susenburger Markus 146
Syben Christopher 416, 502

T

Taguchi Hiroki 408
Taguchi Katsuyuki 422
Tan Jiaxing 38
Tang Xiangyang 102
Tao Weijie 612
Tedgren Asa Carlsson 506
Tellenbach David 478
Teuwen Jonas 438
Thompson William 186
Thüring Thomas 280
Tian Jiao 82
Tivnan Matthew 62, 466, 630, 638
Toennes Christian 400
Toivanen Jussi 570
Tripathi Indu 554
Trousset Yves 454
Tuohimaa T. 514
Töpperwien Mareike 106

U

Uneri Ali 194, 634

V

Vagdargi P. 634
van der Meer Franziska 106
van der Sar Stefan J. 430
van Gogh Stefano 446
van Heteren A. H. 486
Van Nieuwenhove Vincent 380
Van Wassenbergh Sam 90, 518
Varallo Antonio 442
Veldkamp Wouter J. H. 622
Vijayan R. 634
Vittoria Fabio 162
Vogt Sebastian 210, 634
Vöth Tim 412

W

Wadeson Nicola 450
Wang Adam 218, 404, 574
Wang Ge 426
Wang Jing 550
Wang Sen 50
Wang T. 634
Wang Tong 360
Wang Wenying 62, 466, 630, 638
Wang Yi 82
Wang Zhentian 446
Wareing T. 218
Wattenberg Maximilian 482
Weber Thomas 210
Wehrse Eckhard 458
Wein Wolfgang 142
Weiss Clifford R. 236, 400
Weller David 470
Wieczorek Matthias 142
Wu Chengpeng 190

Wu P. 634
Würfl Tobias 264, 392, 502

X
Xia Dan 534, 542
Xie Huiqiao 102
Xing Yuxiang 50, 190, 232, 304, 324, 384
Xu Jingyan 34
Xu Qiong 376

Y
Yang Xikai 228
Yang Yunhuan 434
Yao Yidi 268
Ye Jong Chul 340
Yoon Seongjin 122, 186
Yu Hengyong 288
Yu Zhicong 566
Yu Zhou 408
Yun Hye Sun 388
Yun Hyesun 198

Z
Zabić Stanislav 158
Zbijewski Wojciech 62, 210, 236, 558
Zeng Dong 134, 138
Zeng Gengsheng L. 74, 594
Zeng Rongping 224
Zhang Jin 62, 404
Zhang Li 50, 190, 304, 324, 384
Zhang Tao 324, 574
Zhang Weibin 54
Zhang Yicheng 546
Zhang Yinghui 360
Zhang You 550
Zhang Zheng 534, 542
Zhao Chumin 210
Zhao Hao 78
Zhao Shusen 54, 98
Zhao Xing 42, 54, 98, 360
Zhao Yunsong 352
Zheng Ao 50, 384
Zheng Xuehang 228
Zhi Shaohua 344
Zhou Jian 408
Zhou Weimin 618
Zhu Lei 316, 434
Zhu Manman 138
Zhu Xiaohua 190
Zhuang Jingwen 578
Ziener Christian Herbert 458
Zong Chunguang 598

Schedule CT Meeting 2020

	Monday *August 3*	**Tuesday** *August 4*	**Wednesday** *August 5*	**Thursday** *August 6*	**Friday** *August 7*
10:00 - 11:40 CEST	**Test Session**	**Image Reconstruction** Opening remarks + 5 talks *(Page 30)*	**Deep Learning** Basic short course: Deep Learning + 3 talks *(Page 224)*	**Photon Counting** Basic short course: Photon Counting + 3 talks *(Page 422)*	
		Coffee	**Coffee**	**Coffee**	
12:00 - 13:40 CEST		**Dual and Multi-Energy Imaging** 5 talks *(Page 50)*	**Motion Estimation and Compensation** 5 talks *(Page 236)*	**Breast CT** Introduction: Breast CT + 4 talks *(Page 434)*	**Observer** Basic short course: Observer + 3 talks *(Page 618)*
		Lunch	**Lunch**	**Lunch**	**Lunch**
14:40 - 16:20 CEST		**Poster Session 1** Poster FF + 35 posters *(Page 70)*	**Poster Session 2** Poster FF + 36 posters *(Page 256)*	**Poster Session 3** Poster FF + 36 posters *(Page 450)*	**Novel CT Technologies** 4 talks + Fully 3D 2021 & CT Meeting 2022 previews *(Page 630)*
		Coffee	**Coffee**	**Coffee**	
16:40 - 18:20 CEST	**Test Session**	**Diagnostic and Interventional** Plenary: Intelligent clinical CT + 3 talks *(Page 210)*	**Artifact and Noise Reduction** 5 talks *(Page 400)*	**Security and Phase Contrast CT** 5 talks *(Page 594)*	