

MODEL-BASED ITERATIVE RECONSTRUCTION  
IN CONE-BEAM COMPUTED TOMOGRAPHY:  
ADVANCED MODELS OF IMAGING PHYSICS AND PRIOR INFORMATION

by  
Hao Dang

A dissertation submitted to Johns Hopkins University in conformity with the  
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland  
March, 2017

© 2017 Hao Dang  
All Rights Reserved

# Abstract

Cone-beam computed tomography (CBCT) represents a rapidly developing imaging modality that provides three-dimensional (3D) volumetric images with sub-millimeter spatial resolution and soft-tissue visibility from a single gantry rotation. CBCT tends to have small footprint, mechanical simplicity, open geometry, and low cost compared to conventional diagnostic CT. Because of these features, CBCT has been used in a variety of specialty diagnostic applications, image-guided radiation therapy (on-board CBCT), and surgical guidance (e.g., C-arm based CBCT). However, the current generation of CBCT systems face major challenges in low-contrast, soft-tissue image quality – for example, in the detection of acute intracranial hemorrhage (ICH), which requires a fairly high level of image uniformity, spatial resolution, and contrast resolution. Moreover, conventional approaches in both diagnostic and image-guided interventions that involve a series of imaging studies fail to leverage the wealth of patient-specific anatomical information available from previous scans. Leveraging the knowledge gained from prior images holds the potential for major gains in image quality and dose reduction.

Model-based iterative reconstruction (MBIR) attempts to make more efficient use of the measurement data by incorporating a forward model of physical detection processes. Moreover, MBIR allows incorporation of various forms of prior information into image reconstruction, ranging from image smoothness and sharpness to patient-specific anatomical information. By leveraging such advantages, MBIR has demonstrated improved tradeoffs between image quality and radiation dose in multi-detector CT in the past decade and has recently shown similar promise in CBCT. However, the full potential of MBIR in CBCT is yet to be realized.

This dissertation explores the capabilities of MBIR in improving image quality (especially low-contrast, soft-tissue image quality) and reducing radiation dose in CBCT. The presented work

encompasses new MBIR methods that: i) modify the noise model in MBIR to compensate for noise amplification from artifact correction; ii) design regularization by explicitly incorporating task-based imaging performance as the objective; iii) mitigate truncation effects in a computationally efficient manner; iv) leverage a wealth of patient-specific anatomical information from a previously acquired image; and v) prospectively estimate the optimal amount of prior image information for accurate admission of specific anatomical changes. Specific clinical challenges are investigated in the detection of acute ICH and surveillance of lung nodules. The results show that MBIR can substantially improve low-contrast, soft-tissue image quality in CBCT and enable dose reduction techniques in sequential imaging studies. The thesis demonstrates that novel MBIR methods hold strong potential to overcome conventional barriers to CBCT image quality and open new clinical applications that would benefit from high-quality 3D imaging.

## **Thesis Committee Members**

Jeffrey H. Siewerdsen, Ph.D. (Primary Advisor)

Professor, Department of Biomedical Engineering

Johns Hopkins University

J. Webster Stayman, Ph.D. (Co-advisor)

Assistant Professor, Department of Biomedical Engineering

Johns Hopkins University

Jerry L. Prince, Ph.D.

Professor, Department of Electrical and Computer Engineering

Johns Hopkins University

Katsuyuki Taguchi, Ph.D.

Associate Professor, Department of Radiology and Radiological Science

Johns Hopkins University

...the imposing edifice of science provides a challenging view of what can be achieved  
by the accumulation of many small efforts in a steady objective and dedicated search for truth.

(Charles H. Townes, *Nobel Banquet Speech*)

Pursue some path, however narrow and crooked,  
in which you can walk with love and reverence.

(Henry David Thoreau)

# Acknowledgements

My graduate training began in 2010, when I was soon to make my best decision during doctoral training in joining the I-STAR Lab. Now I am proud to arrive at the end of this long, long journey, with the benefit of a lot of espresso, Koco bibimbap, and well spent GPU cycles. However, none of this would have been possible without the tremendous amount of support I have received.

First, I would like to express my deepest gratitude to my advisor Prof. Jeffrey H. Siewerdsen, who is undoubtedly an extraordinary scientist and a dedicated educator. His passion, dedication, and insight on making scientific innovations and clinical impact have greatly inspired me and shaped my scientific thinking. I am very grateful to have enjoyed a systematic training from him on how to conduct scientific research and deliver findings to the community at a high standard. I would also like to express gratitude to my co-advisor Prof. J. Webster Stayman, who introduced me to the playground of statistical reconstruction and many other fascinating things in life. I had the privilege of working with such an inspiring and cheerful mentor side-by-side, and I much enjoyed the days in which we worked together on solving reconstruction problems and continued those interesting topics on the drive home after work. Moreover, I would like to thank Profs. Katsuyuki Taguchi and Jerry L. Prince for their guidance and valuable feedback on this thesis.

I am also very grateful to have directly worked with brilliant scientists in the lab. My thanks first go to Drs. Wojciech Zibewski, Alejandro Sisniega, and Jennifer Xu, who contributed great expertise to many aspects of the work in Chapters 2-4. Special thanks to their daily companionship, especially in after-lunch espresso. I would also like to thank Dr. Adam S. Wang, who shared with me a lot of his broad knowledge in science, which was instrumental to my academic growth and contributed directly to the work in Chapter 5. I am also happy to acknowledge Drs. Yoshito Otake and Ali Uneri for providing efficient GPU libraries for the work in this thesis and Dr. Sebastian Schafer for assisting with C-arm imaging in separate work on automatic registration.

I would like to acknowledge the contributions of my clinical collaborators, including Drs. Nafi Aygun (Neuroradiology), Vassilis E. Koliatsos (Neurology), and Marc S. Sussman (General and Thoracic Surgery) for their assistance in cadaver experiments and patient studies as well as invaluable insight on clinical applications of imaging technology. I also appreciate the funding support from National Institute of Health through the grants R01-127444, 2R01-CA-112163, and R01-EB-018896 and scholarship support from the Siebel Scholars Foundation. I am also grateful for academic-industry research partnership with: 1) Carestream Health (Rochester NY), including Drs. Xiaohui Wang, John (Ian) Yorkston, and Mr. David H. Foos for their expertise and support in providing the prototype cone-beam CT scanner for the work in Chapters 2-4; and 2) Varian Medical Systems (Palo Alto CA), including Drs. Rick Colbeth, Sungwon Yoon, and Mr. Edward Shapiro for their expertise and support in providing the flat-panel detectors in the work in Chapters 5-6.

I am very grateful to have gotten to know, worked with, and enjoyed life with members from diverse academic and cultural backgrounds in the I-STAR and AIAI Labs. These wonderful people include fellow students Sajendra Nithiananthan (thanks for introducing me to this lab in 2010), Steven Tilley II (thanks for the camaraderie on image reconstruction projects), Qian Cao, Sarah Ouadah, Michael Ketcha, Sureerat (Ja) Reaungamornrat, Yifu Ding, Michael Mow, Xiaoxuan (Esme) Zhang, Daniel Mirota, Wen Liu, Jongheun Yoo, Eugenio Marinetto, research scientists Amir Pourmorteza, Shiyu Xu, Michael Brehler, Matthew Jacobson, Grace Gang, Aswin Mathews, Tharindu De Silva, Joseph Goerres, Hao Zhang, Amir Manbachi, Abdullah Al Muhit, and new faces Pengwei Wu, Runze Han, Wenying Wang, and Andrew Mao. Additional thanks to my best friends outside the lab who have made this Charm City my new home – countless joyful memories from hitting a tennis volley or pressing a DSLR shutter button. Finally, I would like to dedicate this thesis to my family on the other side of the Pacific for their love and support and to my love Lu, to whom my every heartbeat belongs.

# Table of Contents

Abstract.....	ii
Acknowledgements.....	vi
List of Tables .....	xv
List of Figures .....	xvi
List of Abbreviations .....	xxiv
1. Introduction.....	1
1.1 Medical imaging: importance, challenges, and opportunities.....	1
1.1.1 Radiation dose reduction.....	2
1.1.2 Dedicated imaging systems.....	3
1.1.3 Imaging for guided interventions .....	4
1.1.4 Prior information in sequential studies .....	5
1.2 Cone-beam computed tomography .....	7
1.2.1 Principles and technology .....	7
1.2.2 Imaging applications.....	9
1.2.3 Challenges and opportunities .....	11
1.3 Filtered backprojection .....	12
1.3.1 Idealized continuous model .....	12
1.3.2 Filtering and backprojection .....	14
1.4 Model-based iterative reconstruction.....	16

1.4.1 Advantages of MBIR .....	17
1.4.2 Models in CT reconstruction.....	18
1.4.3 Discrete forward model.....	18
1.4.4 Noise model .....	21
1.4.5 Algebraic reconstruction technique .....	22
1.4.6 Total variation minimization.....	23
1.4.7 Maximum-likelihood estimation.....	25
1.4.8 Penalized-likelihood estimation.....	26
1.4.9 Penalized weighted least-squares estimation .....	28
1.5 Model-based iterative reconstruction in cone-beam CT .....	31
1.5.1 Thesis statement.....	33
1.5.2 Thesis outline .....	33
2. Statistical Reconstruction with a Post-Artifact-Correction Noise Model .....	38
2.1 Introduction.....	38
2.1.1 Clinical motivation and challenges .....	38
2.1.2 Acknowledgements and unique contributions .....	41
2.2 Theoretical methods.....	42
2.2.1 PWLS image reconstruction with a noise model for generalized data transformation .	42
2.2.2 Effect of log transformation on variance .....	44
2.2.3 Effect of scatter correction on variance .....	45
2.2.4 Effect of beam hardening correction on variance .....	46

2.2.5 Effect of both scatter and beam hardening corrections: Modified PWLS weights .....	49
2.2.6 Optimization approach .....	50
2.3 Experimental methods .....	52
2.3.1 Experiments on a CBCT test-bench and head phantom.....	52
2.3.2 Evaluation of image quality .....	54
2.4 Results.....	56
2.4.1 Effect of artifact corrections on variance .....	56
2.4.2 Selection of image regularization parameters .....	57
2.4.3 Image reconstruction with beam hardening correction.....	59
2.4.4 Image reconstruction with scatter correction .....	60
2.4.5 Image reconstruction with scatter and beam hardening corrections .....	62
2.5. Conclusions and discussion .....	65
3. Task-Based Image Reconstruction .....	69
3.1 Introduction.....	69
3.1.1 Motivation.....	69
3.1.2 Acknowledgements and unique contributions .....	72
3.2 Methods .....	74
3.2.1 Statistical reconstruction for head imaging.....	74
3.2.2 Task-based performance prediction .....	76
3.2.3 Task-based regularization .....	80
3.2.4 CBCT simulation studies .....	82

3.2.5 Experiments on a CBCT test-bench.....	84
3.3 Results.....	85
3.3.1 Local MTF and NPS .....	85
3.3.2 Task-based regularization (mid-frequency detection task) .....	87
3.3.3 Task-based regularization (various tasks).....	91
3.3.4 Image reconstructions .....	92
3.4 Conclusions and discussion .....	95
4. Multi-Resolution Reconstruction to Mitigate Image Truncation Effects .....	99
4.1 Introduction.....	99
4.1.1 Motivation.....	99
4.1.2 Acknowledgements and unique contributions .....	102
4.2 Methods .....	103
4.2.1 Multi-resolution PWLS for high-quality head imaging .....	103
4.2.2 Optimization approach for multi-resolution PWLS .....	107
4.2.3 Experimental studies.....	108
4.2.4 Multi-resolution PWLS: parameter selection .....	111
4.2.5 Imaging performance and computational complexity.....	112
4.3 Results.....	113
4.3.1 Scatter correction with a head holder model.....	113
4.3.2 Single-resolution PWLS .....	114
4.3.3 Multi-resolution PWLS.....	116

4.3.3.1 Regularization parameters.....	116
4.3.3.2 Downsampling factor.....	118
4.3.3.3 Reconstruction field-of-view .....	119
4.3.4 Computation time.....	120
4.3.5 Comparison of reconstruction methods .....	121
4.4 Conclusions and discussion .....	123
5. Incorporation of Prior Images in Statistical Reconstruction.....	126
5.1 Introduction.....	126
5.1.1 Prior-image-based reconstruction .....	126
5.1.2 Acknowledgement and unique contributions.....	129
5.2 Methods .....	130
5.2.1 Penalized Likelihood Estimation .....	130
5.2.2 Deformable Prior Image Registration, Penalized-Likelihood Estimation .....	131
5.2.3 Strategy for solving dPIRPLE.....	134
5.2.4 Computational complexity and implementation.....	138
5.2.5 Cadaver experiments.....	139
5.2.6 Evaluation methods.....	142
5.3 Results.....	143
5.3.1 Parameter Selection for dPIRPLE.....	144
5.3.2 Convergence properties of dPIRPLE .....	147
5.3.3 Alternating maximization schedule .....	151

5.3.4 Reconstruction results .....	153
5.3.5 Performance with varying sparsity .....	155
5.3.6 Capture range of the deformable registration in dPIRPLE .....	157
5.4 Conclusions and discussion .....	159
6. Control of Change Admission in Prior-Image-Based Reconstruction .....	164
6.1 Introduction.....	164
6.1.1 Balancing prior information with new data .....	164
6.1.2 Acknowledgement and unique contributions.....	166
6.2 Methods .....	167
6.2.1 Regularization design in prior-image-based reconstruction (PIBR) .....	167
6.2.2 Approximate analytical solution of PIBR.....	168
6.2.3 Selection of an operating point .....	171
6.2.4 Predictive performance metric .....	173
6.2.5 Spatially varying prior image strength map.....	174
6.2.6 Digital phantoms and simulation studies .....	175
6.2.7 Computational complexity and implementation .....	177
6.3 Results.....	177
6.3.1 Evaluation of approximate analytical solution.....	177
6.3.2 Evaluation of predictive performance metric.....	179
6.3.3 Location dependence of regularization design in PIBR and evaluation of spatially varying prior image strength map .....	180

6.3.4 Attenuation, shape, and size dependence of regularization design in PIBR .....	184
6.3.5 Evaluation of regularization design in lung nodule surveillance .....	186
6.3.6 Evaluation of regularization design in a nodule disappearance scenario.....	188
6.4 Conclusions and discussion .....	190
7. Conclusions.....	193
7.1 Contributions .....	194
7.2 Future work.....	198
Bibliography .....	200
Curriculum Vitae .....	231

## List of Tables

2.1	Pseudocode for solving PWLS* using OS-SQS.	51
2.2	Summary of major parameters in the OS-SQS algorithm for PWLS*.	52
3.1	Spatially varying regularization design for maximal task-based performance.	82
4.1	Pseudocode for solving the multi-resolution PWLS reconstruction using OS-SQS.	107
5.1	Pseudocode for the dPIRPLE algorithm.	137
5.2	Summary of major parameters in dPIRPLE.	144
5.3	Computation time at various settings of $T$ .	152
6.1	Summary of different ways of selecting the operating point for the approximate analytical solution of PIRPLE.	173

# List of Figures

1.1	Two potential methods for dose reduction in sequential imaging studies: sparse sampling (fewer projections acquired in the scan) and reduced fluence (lower dose per projection).	6
1.2	Illustration of a projection along a certain angle in a continuous model assuming a continuous object and a continuous detector system.	13
2.1	Artifact corrections in CBCT of the head. (a) CBCT image reconstructed using FBP without artifact corrections. (b) CBCT image with artifact corrections (scatter and beam hardening, denoted by the superscript $SB$ ) showing a strong improvement in image uniformity but an amplification of noise by more than a factor of 2. (c) Flowchart of the artifact correction process and image reconstruction methods (FBP and PWLS) investigated in this work.	40
2.2	Experimental methods. (a) Benchtop CBCT system with the x-ray source and FPD in a geometry emulating a compact head scanner. (b) Anthropomorphic head phantom incorporating simulated intracranial hemorrhage. The occipital portion can be removed to access inside the interior of the cranium. (c) A photograph inside the head phantom during assembly, showing the gelatin mixture (brain), wax (ventricles), and plastic spheres (hemorrhage).	54
2.3	Illustration of image quality metrology. (a) Two axial slices in which the spatial resolution, noise, and CNR were computed. The spatial resolution and contrast were computed by fitting an ESF to voxel values in $60^\circ$ fans centered on a sphere of contrast equivalent to blood. Noise was computed as the standard deviation of voxel values in a homogeneous ROI immediately adjacent to the sphere. (b) Example sigmoid fit to the measured ESF, from which spatial resolution was characterized in terms of the ESF width, $\varepsilon$ , and contrast was given by the parameter $c$ .	55
2.4	Example illustration (in a lateral projection view) of the change in variance associated with each step of artifact correction. The change in variance is shown for corrections associated with: (a) beam hardening correction for water ( $\eta_w$ ); (b) beam hardening correction for bone ( $\eta_b$ ); (c) total beam hardening correction ( $\eta_w \eta_b$ ); (d) scatter correction $(1+SPR)^2$ ; and (e) all scatter and beam hardening corrections.	57
2.5	Selection of regularization parameters with respect to noise-resolution tradeoffs. (a) PWLS* images in a ROI in the $Z_1$ axial slice using different $\delta$ values. Images were reconstructed using the fully-corrected dataset with spatial resolution matched at $\varepsilon = 0.65$ mm. (b-c) Noise-resolution tradeoff and CNR-resolution tradeoff for PWLS* using different $\delta$ values. The error bars are based on the standard deviation of the noise in 6 neighboring axial slices. The solid curves are second-order polynomial fits intended merely	58

as a guide to trends evident in the underlying data points. A value of  $\delta = 5$  HU was selected as a conservative choice giving a reasonable level of noise reduction and edge preservation without introducing an unnatural appearance of patchy over-regularization.

2.6	Reconstructions with beam-hardening correction. (a) FBP reconstruction with no corrections. (b) FBP reconstruction of beam-hardening-corrected data. (c) PWLS reconstruction of beam-hardening-corrected data with conventional weights. (d) PWLS* reconstruction of beam-hardening-corrected data with modified weights. The proposed method (PWLS* <sup>B</sup> ) maintains the quality of artifact correction and gives a factor of $\sim 2$ improvement in noise compared to FBP <sup>B</sup> . (Resolution was matched at $\varepsilon = 0.70$ mm for fair comparison.)	59
2.7	Reconstructions with scatter correction - top row at matched spatial resolution and bottom row at matched CNR. From left to right: FBP reconstruction of uncorrected and scatter-corrected projection data; PWLS of scatter-corrected data with conventional weights; and PWLS* reconstructions of scatter-corrected data with modified weights.	61
2.8	Noise-resolution tradeoffs for the FBP, PWLS, and PWLS* reconstruction methods operating on scatter-corrected projection data. The levels of matched spatial resolution and CNR used in Fig. 2.7 are marked by the dashed lines in (b).	62
2.9	Reconstructions with scatter and beam hardening corrections - top row at matched spatial resolution and bottom row at matched CNR. As in Fig. 2.7, from left to right: FBP, PWLS, and PWLS* reconstructions of fully corrected projection data.	63
2.10	Noise-resolution tradeoffs for the FBP, PWLS, and PWLS* reconstruction methods operating on fully-corrected (scatter and beam hardening) projection data. The levels of matched spatial resolution and CNR used in Fig. 2.9 are marked by the dashed lines in (b).	63
2.11	Axial (top row), coronal (middle row), and sagittal (bottom row) CBCT images of the head phantom with scatter and beam hardening corrections, reconstructed by FBP, PWLS, and PWLS*. The skull base presents a challenging region for which proper account of scatter and beam hardening corrections is essential to high-quality reconstruction. The spatial resolution in each case was matched ( $\varepsilon = 0.40$ mm) at the largest sphere in the axial slice $Z_1$ .	64
3.1	Illustration of nonuniform spatial resolution (ESF width, $\varepsilon$ ) and noise (standard deviation, $\sigma$ ) at three regions in an anthropomorphic head phantom (1 - near the cranium, 2 - at the center of the cranial vault, and 3 - near the skull base) in an image reconstructed by PWLS with a quadratic penalty. The phantom was scanned on a CBCT test-bench at a dose of 24 mGy. Grayscale window for axial images 1-3 is [-60 160] HU.	71

3.2	Three task functions investigated in the current work: (a) low-frequency, (b) mid-frequency, and (c) high-frequency. The mid-frequency task was modeled as a difference of two Gaussians with 50 HU contrast, corresponding to a low-contrast feature (acute ICH) with characteristic length of $\sim 1$ mm.	80
3.3	Head phantom and CBCT test-bench. (a) Sagittal, axial, and coronal slices of the 3D digital head phantom used in the simulation studies, which contains realistic bone attenuation and uniform soft tissue. Five locations of interest are denoted 1-5. (b) Photographs of the CBCT test-bench and an anthropomorphic head phantom filled with materials emulating brain, ventricles, and acute ICH.	84
3.4	Local MTF and NPS in 3D images of the head reconstructed using PWLS with a conventional penalty (constant $\beta = 10^{6.4}$ ). Each exhibits shift-variance and anisotropy at locations defined in Fig. 3.3(a). The top two rows show “axial” MTF( $f_x, f_y$ ) and NPS( $f_x, f_y$ ), and the bottom two rows show “sagittal” MTF( $f_y, f_z$ ) and NPS( $f_y, f_z$ ), the latter demonstrating the null cone of unsampled frequencies in regions (e.g., location 5) far from the central slice.	86
3.5	Maps of the regularization parameter $\beta$ for the three penalty methods. (a-b) The conventional penalty employs a constant scalar value $\beta (= 10^{6.4})$ , chosen to yield the highest $d'$ averaged over all regions of the brain). (c-d) The certainty-based uniform resolution penalty involves the product of a scalar $\beta$ and a spatially-varying certainty term $\kappa_j^2$ (chosen to yield uniform PSF width of 0.95 mm). (e-f) The task-based penalty adjusts the $\beta$ map to maximize $d'$ everywhere in the brain (chosen here for a mid-frequency task, $W_{Mid}$ ).	88
3.6	Spatial resolution and noise for the three penalty methods. (a-f) Spatial resolution is described by the FWHM of the local PSF, and (g-l) noise is given by the local standard deviation in voxel values. The conventional penalty yields (a-b) nonuniform spatial resolution and (g-h) fairly uniform noise. The certainty-based penalty yields (c-d) uniform spatial resolution and (i-j) slightly stronger nonstationarity in the noise. The task-based penalty yields spatially varying resolution and noise characteristics that adjust in a manner to maximize local $d'$ (chosen here for the mid-frequency task, $W_{Mid}$ ).	89
3.7	Local detectability for the three penalty methods. The top two rows (a-f) show maps of $d'$ (mid-frequency detection task) for the (a-b) conventional constant penalty; (c-d) uniform resolution penalty; and (e-f) task-based penalty. The bottom two rows (g-j) show the ratio of detectability index for the (g-h) uniform resolution penalty and (i-j) task-based penalty to that of the conventional penalty. The uniform resolution penalty increases $d'$ near the skull base by $\sim 10\%$ but degrades $d'$ near the periphery by $\sim 5\%$ . The task-based penalty improves $d'$ by up to $\sim 12\%$ near the skull base and meets or exceeds that of other methods at every location in the head.	90

3.8	Detectability index $d'$ computed as a function of regularization parameter $\beta$ at locations (a) near the skull base, (b) in the center of the brain, and (c) near the cranium. Each plot shows $d'$ computed for a low-frequency task (denoted as $d'_{Low}$ ), a mid-frequency task ( $d'_{Mid}$ ), and a high-frequency task ( $d'_{High}$ ).	92
3.9	Visualization of simulated ICH lesions using three regularization methods. The conventional penalty ( $\beta=10^{6.4}$ ) exhibited a somewhat over-smoothed image near the skull base. The uniform resolution penalty yielded a more uniform image appearance (uniform PSF) and improved conspicuity near the skull base, with slight tradeoff near the cranium. The task-based penalty further improved conspicuity at each location, particularly near the skull base.	93
3.10	PWLS reconstructions of a head phantom containing a simulated ICH of 12 mm diameter and 50 HU contrast in regions near the cranium (a, c, e) and skull base (b, d, f). (a,b) Conventional penalty ( $\beta=10^{5.6}$ ). (c,d) Uniform resolution penalty. (e,f) Task-based penalty.	94
4.1	Artifacts caused by lateral truncation of the head support in CBCT of the head. (a) CAD drawing of a patient with head supported by a carbon-fiber head holder during a CBCT scan. (b) PWLS image of an anthropomorphic head phantom without a head holder. The circular inserts within the central region of the cranium span a range of contrast including that of ICH. (c) PWLS image of the same, with a U-shaped carbon-fiber head holder in place during the scan (evident beneath the posterior of the head). (d) Illustration of RFOV and SFOV for circular orbit CBCT.	100
4.2	Illustration of fine and coarse regions in multi-resolution reconstruction (only x-y plane shown here). The dashed circle denotes the SFOV. The fine region ( $\mu_F$ ) is a 3D rectangle that contains the head, and the coarse region ( $\mu_C$ ) is the space outside the fine region that contains the head holder (depicted as a black, U-shaped arc posterior to the head). The volume encompassing both the fine and coarse regions is the RFOV.	104
4.3	Experimental setup. (a) Photograph of the CBCT test-bench, head phantom, and head holder. (b) Illustration of three locations at which the head holder was positioned during the experiments. (c-e) Axial images superimposed with a representation of the head holder at each location. (f) Axial image illustrating structures and ROI used for image quality assessment. The central circular insert was used to compute spatial resolution (edge spread function) and contrast, and the nearby rectangular ROI was used to compute noise.	109
4.4	Scatter correction (a-f) without and (g-l) with the head holder included in the Monte Carlo model. (a-c) FBP reconstructions without a head holder model exhibit shading and streaks in the (d-f) difference images from (m) “truth”. Including the head holder in scatter correction reduces such residual errors as shown in (g-i) and difference images (j-l).	114

4.5	Nominal parameter selection for single-resolution PWLS. (a) ESF and CNR as a function of the regularization strength $\beta$ (in the absence of truncation). (b) Axial slice of a single-resolution PWLS image using $\beta = 10^{2.4}$ , exhibiting a reasonable balance between ESF and CNR and taken as the “truth” image for subsequent PWLS reconstructions.	115
4.6	Single-resolution PWLS with a carbon-fiber head holder positioned at three locations posterior to the head. RFOV is $(400 \times 480 \times 480)$ voxels. (d-f) Difference images between (a-c) and the “truth” image of Fig. 4.5(b).	116
4.7	Accuracy of multi-resolution PWLS reconstructions as a function of coarse region regularization strength $\beta_C$ and downsampling factor (DS) at three locations.	117
4.8	Multi-resolution PWLS reconstruction for various choices of coarse region regularization strength, $\beta_C$ . Images (a-c) show the fine region, with the left half showing the PWLS image (grayscale window: $[-300, 200]$ HU) and the right half showing the difference from truth (grayscale window: $[-1100 -700]$ HU). Images (d-f) show the coarse region (outside cyan box). The head holder was at location 2, the DS was 4, and the RFOV was $1000^3$ voxels with 0.5 mm isotropic voxel size.	118
4.9	Multi-resolution PWLS reconstruction for various choices of coarse region voxel size, characterized by DS. Images (a-c) show the fine region, with the left / right presentation of the PWLS (grayscale window: $[-300, 200]$ HU) and difference image (grayscale window: $[-1100 -700]$ HU) as in Figure 4.8. The head holder was at location 3, the RFOV was $1000^3$ voxels (with fine region voxel size = 0.5 mm isotropic), and $\beta_C$ was chosen to minimize RMSD for each DS.	119
4.10	Multi-resolution PWLS reconstruction for various choices of RFOV. Images (a-c) show the fine region, with the left / right presentation of the PWLS (grayscale window: $[-300, 200]$ HU) and difference image (grayscale window: $[-1100 -700]$ HU) as in Figs. 8-9. The head holder was at location 2, the DS was set to 4, and $\beta_F$ was adjusted slightly to maintain constant noise-resolution performance: (a) $\beta_F = 10^{2.35}$ , (b) $\beta_F = 10^{2.38}$ , and (c) $\beta_F = 10^{2.40}$ . The parameter $\beta_C$ was chosen to minimize RMSD for each RFOV (after selecting $\beta_F$ ).	120
4.11	(a) Measured computation time (averaged over one forward projection and one backprojection) and measured memory usage as a function of RFOV. (b) Time per iteration (i.e., for all subsets) as a function of RFOV for single-resolution and multi-resolution PWLS reconstruction at different DS levels. Substantial speedup can be seen compared to the single-resolution approach at DS = 2 and to multi-resolution approach at DS = 4.	121
4.12	Comparison of single-resolution and multi-resolution PWLS reconstruction. (a-b) Single-resolution PWLS reconstruction using the basic RFOV $(400 \times 480 \times 480)$ voxels. (c-d) Single-resolution PWLS reconstruction using an extended RFOV $(600^3)$ voxels assuming an isotropic	122

	voxel size of 0.5 mm). (e-f) Multi-resolution PWLS reconstruction using the same extended RFOV. The symbols $\varepsilon$ and $t$ denote RMSD and computation time, respectively, quantifying the reduction in artifact using an extended RFOV and the benefit to computation time using the multi-resolution method.	
5.1	Flow chart for solving the dPIRPLE objective function using an alternating maximization approach.	134
5.2	(a) Experimental setup on the CBCT test-bench for cadaver experiments. (b) Simulation of lung tumor growth via petroleum jelly injection into the cadaver lung. A semiopaque rendering of a generic skeleton is overlaid on the photograph to illustrate the anatomical position of the cadaver.	140
5.3	Three views of the patient-specific prior image, formed by PLE ( $p_R = 2$ , $\beta_R = 10^6$ ) reconstruction of the fully sampled dataset before the injection. Two zoomed-in regions in each view correspond to prior image (left) and current anatomy (right).	140
5.4	Example optimization of $(\beta_P, \beta_R)$ from a pair of 1D parameter sweeps (dashed lines) compared to the true 2D optimum (asterisk). Results here used the global RMSE as a metric; however, very similar results are seen when using local RMSE.	146
5.5	(a, b) Deformation field estimated by dPIRPLE using 20 projections. The image below the deformation field is a merged image from the prior image and the current anatomy. The yellow and blue color correspond to positive and negative image value differences between the prior image and the current anatomy, while the original greyscale is used in the region where the image value differences are small. (c, d) Error vectors between deformation field estimated by dPIRPLE and the “true” deformation field approximated by registering the prior image with current anatomy. (e, f) Plot of the magnitude of error vectors (with a trace of the body outline superimposed).	148
5.6	Axial and sagittal views of a sequence of registration residual errors in dPIRPLE using 20 projections over $190^\circ$ . The value $z$ reflects the estimation at the $z^{\text{th}}$ alternation of registration updates.	149
5.7	(a) The dPIRPLE objective function difference analyzed as a function of the cumulative number of image updates, $T_z=50$ . (b) Local RMSE versus iteration number for all approaches.	150
5.8	Local RMSE computed versus the cumulative number of image updates for dPIRPLE at various $T$ values.	152
5.9	Reconstruction results of FBP, PLE ( $\beta_R=10^3$ ), rigid PIRPLE ( $\beta_P=10^4$ , $\beta_R=10^{3.5}$ ), single-registration dPIRPLE ( $\beta_P=10^4$ , $\beta_R=10^{3.5}$ ), and dPIRPLE ( $\beta_P=10^4$ , $\beta_R=10^{3.5}$ ) with 20 projections and 1.25 mAs/projection. The local RMSE of the reconstruction result is displayed in each case. Note the accurate estimate of nodule and lung structures (e.g. airways) in dPIRPLE.	154

5.10	PLE, PIPEL, rigid PIRPLE and dPIRPLE reconstruction images at different levels of projection sparsity at low exposure (0.1 mAs/projection).	156
5.11	Global image accuracy [(a) RMSE] and local image quality [(b) SSIM] for PLE, PIPEL, rigid PIRPLE, and dPIRPLE reconstruction computed as a function of sparsity at low exposure (0.1 mAs/projection).	157
5.12	RMSE between the registered prior image and the current anatomy as a function of the mean displacement after both the initial rigid registration and after both rigid and deformable registration. Each vertical pair of data points correspond to one instance of random rigid motion that was added to the prior image before registration. The dashed line indicates the delineation between success and failure cases at a mean displacement of 125 mm.	158
6.1	Approximation of the modified penalty function $f$ (equivalent to the Huber function) with a quadratic function $g$ about an operating point $\tau_i$ . The operating point is selected either within (a) or outside of (b) the quadratic neighborhood $([-\delta, \delta])$ of the function $f$ .	170
6.2	(a) Ellipse phantom with attenuation ( $\text{mm}^{-1}$ ): 0.021 (background ellipse), 0.041 (bright circular insert), and 0.001 (dark circular insert). An anatomical change is introduced in one or both of Location I and Location II in the subsequent scan. The dashed circles illustrate the Change ROIs used in the predictive performance metric. (b) Thorax phantom generated from an axial slice of a CT scan of a cadaver torso. A uniform circular lung nodule was introduced in the subsequent scan in the center or the periphery of the lung to emulate a lung nodule surveillance scenario.	176
6.3	PIRPLE reconstruction, Ideal estimate, P+C estimate (1 iteration), and P+C estimate (5 iterations) of a simulated circular solitary pulmonary nodule at various $\beta_P$ . Note that too small $\beta_P$ or too large $\beta_P$ leads to little benefits or false features (false negative) in PIRPLE reconstruction. Only 20 projections equally distributed over $190^\circ$ were used. (Grayscale window: $[0 \ 0.04] \text{ mm}^{-1}$ .)	179
6.4	Evaluation of the proposed predictive performance metric at different $\beta_P$ (with a uniform log spacing of $10^{0.1}$ ) using either PIRPLE reconstruction or one of the image estimation methods. Note that all the methods yielded almost the same optimal $\beta_P$ , while the proposed method (using P+C estimate) does not require full image reconstruction.	180
6.5	Comparison of PIRPLE reconstructions and the prospective performance metric for the same change at two different locations. A clear difference in the optimal $\beta_P$ (over an order of magnitude) between the two locations can be seen both in the PIRPLE reconstructions and in the metric curves. These results motivate the design of a spatially varying $\beta_P$ map. (Grayscale window: $[0.03 \ 0.052] \text{ mm}^{-1}$ for Location I and $[-0.02 \ 0.052] \text{ mm}^{-1}$ for Location II.)	182

6.6	(a-b) Spatially varying $\beta_P$ map generated using traditional exhaustive search (a) or the proposed method (b). The $\beta_P$ map in (b) exhibits good agreement with the $\beta_P$ map in (a). (c) PIRPLE reconstruction of the same change at both locations using a scalar $\beta_P$ optimized for Location I ( $\hat{\beta}_P^I = 10^{3.2}$ ), a scalar $\beta_P$ optimized for Location II ( $\hat{\beta}_P^{II} = 10^{4.5}$ ), and the spatially varying $\beta_P$ map in (b). The parameter $\varepsilon$ stands for the RMSE of the PIRPLE reconstruction with respect to the truth image. (Grayscale window: [0.03 0.052] mm <sup>-1</sup> for reconstructed images at Location I, [-0.02 0.052] mm <sup>-1</sup> for reconstructed images at Location II, and [-0.01 0.01] mm <sup>-1</sup> for all the difference images.)	183
6.7	Illustration of the dependence of the optimal $\beta_P$ on the attenuation (a), shape (b), and size (c) of the anatomical change. One anatomical change was introduced to Location II of the ellipse phantom in Fig. 6.2(a), and only one of the three properties of the change mentioned above was varied at a time. The optimal $\beta_P$ was estimated by evaluating the proposed metric at different $\beta_P$ with a uniform log spacing of $10^{0.01}$ . A negative exponential function was fit to the data points in (a) to help illustrate the relationship in (a). (Grayscale window: [0 0.052] mm <sup>-1</sup> .)	186
6.8	Spatially varying $\beta_P$ map generated using traditional exhaustive search (a) or the proposed method (b) for optimally admitting a solitary pulmonary nodule everywhere in both sides of the thorax cavity. The optimal $\beta_P$ was estimated on an image grid with a spacing of 20 voxels in each dimension and then interpolated into a $\beta_P$ map using radial basis functions. The $\beta_P$ map in (b) exhibited good agreement with the $\beta_P$ map in (a) while reducing the computation time by a factor of $\sim 20$ .	187
6.9	Image reconstruction of a solitary pulmonary nodule which was not present in a baseline exam (e) but appeared in the periphery of the right lung in the subsequent exam (a). (b-d) A ROI of the images reconstructed by FBP, PIRPLE using a suboptimal scalar $\beta_P$ , and PIRPLE using the spatially varying $\beta_P$ map in Fig. 6.8(b). (f-h) Difference image between each of the images in (b-d) and the current anatomy. The parameter $\varepsilon$ stands for the RMSE of the difference image in (f-h). (Grayscale window: [0 0.04] mm <sup>-1</sup> for reconstructed images and [-0.01 0.01] mm <sup>-1</sup> for difference images.)	188
6.10	(a) PIRPLE reconstruction of the current anatomy (labeled as “Truth”) at various $\beta_P$ . A lung nodule was present in the prior image but not present in the current anatomy. Note that too small $\beta_P$ or too large $\beta_P$ leads to little benefits or false features (false positive) in PIRPLE reconstruction. (Grayscale window: [0 0.04] mm <sup>-1</sup> ) (b) Evaluation of the proposed metric at different $\beta_P$ in the nodule disappearance scenario. Note that all the methods yielded almost the same optimal $\beta_P$ , while the proposed method (P+C estimate) does not require full image reconstruction.	189

## List of Abbreviations

AEC	Automatic Exposure Control
ALARA	As Low As Reasonably Achievable
ART	Algebraic Reconstruction Technique
ASGD	Adaptive Stochastic Gradient Descent
CBCT	Cone-Beam Computed Tomography
CG	Conjugate Gradient
CHO	Channelized Hotelling Observer
CMOS	Complementary Metal-Oxide Semiconductor
CNR	Contrast-to-Noise Ratio
CRT	Conformal Radiotherapy
CT	Computed Tomography
CTV	Clinical Target Volume
DBT	Digital Breast Tomosynthesis
DOF	Degree-of-Freedom
dPIRPLE	deformable Prior Image Registration, Penalized-Likelihood Estimation
DS	Downsampling Factor
EPID	Electronic Portal Imaging Device
ESF	Edge Spread Function
FBP	Filtered-Backprojection
FDK	Feldkamp-Davis-Kress
FFD	Free-Form Deformation
FOV	Field of View
FPD	Flat-Panel Detector
FWHM	Full Width at Half Maximum
GPU	Graphics Processing Unit
ICH	Intracranial Hemorrhage
ICU	Intensive Care Unit

IGRT	Image-Guided Radiation Therapy
IGS	Image-Guided Surgery
IMRT	Intensity-Modulated Radiation Therapy
KCR	Known-Component Reconstruction
MBIR	Model-Based Iterative Reconstruction
MDCT	Multi-Detector Computed Tomography
ML	Maximum Likelihood
MLC	Multi-Leaf Collimator
MRI	Magnetic Resonance Imaging
MTF	Modulation Transfer Function
NCCU	Neuroscience Critical Care Unit
NPS	Noise-Power Spectrum
NPW	Non-Prewhitening
OS-SQS	Separable Quadratic Surrogate with Ordered Subsets
PET	Positron Emission Tomography
PIBR	Prior-Image-Based Reconstruction
PICCS	Prior Image Constrained Compressed Sensing
PIPLE	Prior Image, Penalized-Likelihood Estimation
PIRPLE	Prior Image Registration, Penalized Likelihood Estimation
PL	Penalized-Likelihood
PLE	Penalized-Likelihood Estimation
PSF	Point Spread Function
PTV	Planning Target Volume
PW	Prewhitening
PWLS	Penalized Weighted Least-Squares
RFOV	Reconstruction Field of View
RMSD	Root Mean Square Difference
RMSE	Root Mean Square Error
ROI	Region-of-Interest

SAD	Source-to-Axis Distance
SART	Simultaneous Algebraic Reconstruction Technique
SDD	Source-to-Detector Distance
SFOV	Scan Field of View
SPECT	Single Photon Emission Computed Tomography
SPR	Scatter-to-Primary Ratio
SQS	Separable Quadratic Surrogate
SSD	Sum of Squared Differences
SSIM	Structural Similarity
TCM	Tube Current Modulation
TFT	Thin-Film Transistor
TV	Total Variation

# Chapter 1

## 1. Introduction

### 1.1 Medical imaging: importance, challenges, and opportunities

Medical imaging involves the interaction of some form of energy with tissues in the body to form a spatially distributed measurement of a particular quantity – in x-ray imaging and computed tomography (CT), for example, the spatial distribution of electron density, which is in turn related to attenuation coefficient and physical density. Applications include both the diagnosis of disease and the guidance of therapeutic intervention. The particular forms of energy applied in such imaging procedures cover a wide range of the electromagnetic spectrum (and mechanical / acoustic energy), including radiofrequency waves in magnetic resonance imaging (MRI), visible light in endoscopy and microscopy, x-rays in mammography, fluoroscopy, and CT, gamma rays in nuclear medicine, and sound waves in ultrasound imaging.

Noninvasive medical imaging is among the primary advances in 20<sup>th</sup> and 21<sup>st</sup> century medicine – for example, largely replacing exploratory surgery. In fact, x-rays were being used for diagnosis and image guidance within a month of their serendipitous discovery by Wilhelm Röntgen in 1895.<sup>1</sup> The following century saw vast technological advances in medical imaging, including the introduction of new imaging modalities and substantially improved imaging capabilities. Such technological advances have greatly shaped the evolution of medical practice and have established an imperative role of medical imaging in many aspects of today's medical practice. The following sections discuss a few rising challenges and opportunities in medical imaging in the century ahead.

### 1.1.1 Radiation dose reduction

The human body receives background ionizing radiation constantly from sources such as cosmic radiation and naturally occurring radioactive materials (e.g., radon). Some medical imaging modalities also cause ionizing radiation to the human body. For example, the radiation dose associated with a typical CT scan (1-14 mSv) is comparable to that from annual background radiation (1-10 mSv depending on geographic locale).<sup>2</sup> The advances in CT imaging capabilities since ~2001 have provided faster and more accurate diagnosis and interventional guidance, but they have also led to greatly increased CT usage. For example, the number of CT scans increased from ~13 million in 1990 to ~64 million in 2006.<sup>3</sup> Although the health risks associated with low-level radiation (<100 mSv) such as radiation from a CT scan is small (with the degree of risk representing a prominent area of debate),<sup>4</sup> it is nonetheless desirable to keep the radiation dose in a CT scan as low as reasonably achievable (ALARA) such that a particular diagnostic task can still be reliably accomplished.

To reduce radiation dose in CT, the radiology community has worked to implement the principles of ALARA in CT imaging, which guides clinicians to select the proper amount of scan dose by taking the specific patient attenuation (e.g., patient size) and specific imaging task into consideration.<sup>5-7</sup> In addition, a number of dose reduction techniques have been developed and now available on commercial CT scanners.<sup>7,8</sup> One such technique is tube current modulation (TCM),<sup>9</sup> which modulates the x-ray tube current in a CT scan in the longitudinal ( $z$ ) direction and/or angular (e.g., anterior-posterior versus lateral) direction to adapt to changes in patient attenuation. Automatic exposure control (AEC)<sup>10</sup> is a more comprehensive technique, which automatically determines and delivers an amount of dose to the patient using preprogramming and/or a near real-time feedback of detector signal. A new category of dose reduction technique is given by model-based iterative reconstruction (MBIR), which has demonstrated strong potential to improve image

quality and reduce radiation dose in CT.<sup>11–13</sup> This dissertation develops and evaluates a number of new MBIR methods that can potentially reduce radiation dose in CT.

### **1.1.2 Dedicated imaging systems**

An exciting new area in medical imaging is the development of dedicated imaging systems for specialty applications. While conventional whole-body imaging systems [e.g., multi-detector CT (MDCT), MRI, positron emission tomography (PET), and single photon emission CT (SPECT)] also have scan protocols optimized for different applications (e.g., head, heart, lung, and abdomen), the development of dedicated imaging systems provides more flexibility in the design and optimization of the system for specific applications. Dedicated imaging systems tend to have smaller footprint, higher portability, and lower cost compared to whole-body imaging systems. Moreover, because such systems are designed to accomplish a relatively narrow range of diagnostic tasks, they have the potential to outperform conventional, general-purpose imaging systems.

An example of a dedicated imaging system is x-ray mammography, which has many unique attributes compared to a general radiography system, including relatively low x-ray tube voltage for maximizing soft-tissue contrast in the breast, high spatial resolution of the detector for improving microcalcification detection, and the use of a compression paddle for reducing scatter, dose, and detector dynamic range requirements.<sup>4</sup> Another type of dedicated system for breast imaging is digital breast tomosynthesis (DBT), which attempts to reduce the “superposition” problem in x-ray mammography by acquiring multiple projections of the breast at several angular positions and synthesizing images at a few in-focus planes. DBT has demonstrated improved diagnostic accuracy for breast cancer when combined with x-ray mammography.<sup>14</sup> Moreover, since the introduction of flat-panel detectors (FPDs) in the late 1990s, dedicated CBCT systems with a FPD have been developed for specialty applications in a variety of specialty applications, including dental, maxillofacial, otologic,<sup>15–18</sup> breast,<sup>19–24</sup> and musculoskeletal imaging.<sup>25–29</sup> Such CBCT

systems are discussed in greater detail in Sec. 1.2. Even in MRI, dedicated systems have been of great interest due to the high installation and management costs, large footprint, and patient discomfort associated with current whole-body MRI systems. Among such developments are dedicated low-field and intermediate-field MRI systems for musculoskeletal imaging that have demonstrated diagnostic accuracy similar to high-field whole-body MRI while bringing higher cost-effectiveness and patient comfort.<sup>30–32</sup>

### **1.1.3 Imaging for guided interventions**

Another exciting trend in medical imaging is the increasing role of imaging to guide high-precision radiation therapy and surgery. Radiation therapy uses high-energy radiation (e.g., x-rays, gamma rays, and charged particles) to kill cancer cells while sparing nearby healthy tissue. Image-guided radiation therapy (IGRT) is often regarded as another major advance in radiation therapy following: 1) the introduction of 3D conformal radiotherapy (CRT) with multi-leaf collimators (MLC) in the 1980s that enabled delivery of radiation to precisely defined target areas; 2) and the introduction of intensity-modulated radiation therapy (IMRT) in the 1990s that enabled sophisticated treatment planning.<sup>33</sup> A major challenge in radiation therapy is that geometric uncertainties caused by changes in patient setup between each treatment fraction and/or by patient motion during the treatment can hinder accurate delivery of the treatment plan. Early development in IGRT involved adding radiographic systems to existing radiation treatment machines and using oblique radiographic images generated by the radiographic systems to correct for geometric uncertainties.<sup>33</sup> This was followed by the integration of electronic portal imaging devices (EPIDs)<sup>34</sup> into the treatment system, which could generate images of the anatomy and field shape simultaneously using the megavoltage treatment beam, although the contrast of the anatomy in the megavoltage image was relatively poor. On-board kilovoltage CBCT was introduced in the early 2000s,<sup>35–37</sup> providing 3D volumetric imaging of the anatomy with improved soft-tissue contrast and has grown to become the modern

standard of care. The use of CBCT for radiation therapy guidance will be further discussed in Sec. 1.2. Moreover, other imaging modalities such as MDCT, MRI, and PET can potentially be integrated into radiation therapy. Such systems are currently under technical development and/or early clinical evaluation.<sup>38–41</sup>

Image-guided surgery (IGS) usually involves the use of imaging, tracking, registration, and visualization techniques to provide the surgeon with real-time images of the patient or real-time location information of surgical tools with respect to the patient. IGS has been evolving in the past three decades along with the improvement of medical imaging and computing power. Nowadays, IGS systems for many clinical areas have been developed.<sup>42</sup> For example, image-guided neurosurgery has been practiced for over 25 years and has become the standard of care in many surgical centers.<sup>43</sup> Because of image-guided neurosurgery, a number of procedures now do not require craniotomy. Because significant brain shift may occur in some neurosurgeries,<sup>44</sup> the use of intra-operative imaging such as ultrasound, C-arm CBCT, or MRI can provide up-to-date anatomical information for surgical navigation compared to only using pre-operative images.<sup>44,45</sup> Orthopedics is another clinical field in which IGS has been used extensively, including the use of intra-operative imaging (usually radiographs or CT) to guide pedicle-screw insertion, hip replacement, knee replacement, and fracture alignment.<sup>46–48</sup> Another clinical field that can potentially benefit from IGS is image-guided cardiac interventions, including systems that use intra-operative real-time MRI, combine intra-operative x-ray angiography and MRI, or register intra-operative ultrasound to pre-operative MRI.<sup>49–51</sup>

#### **1.1.4 Prior information in sequential studies**

There are many clinical scenarios in which sequential imaging studies are conducted. For example, in diagnostic radiology, patients with suspicious lung nodules or cancer often undergo a series of longitudinal chest CT scans to assess the tumor growth rate<sup>52</sup> and/or monitor the therapy response.<sup>53</sup>

In image-guided interventions, pre-operative CT or MRI studies are often acquired, and multiple intra-operative imaging studies (e.g., using a C-arm CBCT) may be acquired for up-to-date tracking, registration, and visualization.<sup>42,54–57</sup> Traditionally, such sequential studies are conducted through a series of *complete* acquisitions, and the cumulative radiation dose in such studies raises concerns for both the patient and the surgical staff.

To address this problem, many dose reduction methods have been developed (in addition to typical dose reduction methods used in *complete* acquisitions such as TCM and AEC). Two straightforward methods are illustrated in Fig. 1.1, including: 1) reduction in the amount of data acquired (i.e. sparse sampling), and 2) reduction of the x-ray technique (e.g., reduced tube current). However, such dose reduction techniques in the absence of countermeasures to mitigate noise and/or artifacts tend to experience a reduction in overall image quality of the reconstructed image. For example, sparse sampling makes the image reconstruction problem more ill-conditioned and leads to streak artifacts in the reconstructed image, while lowering fluence increases quantum noise in the projection data and the reconstructed image.

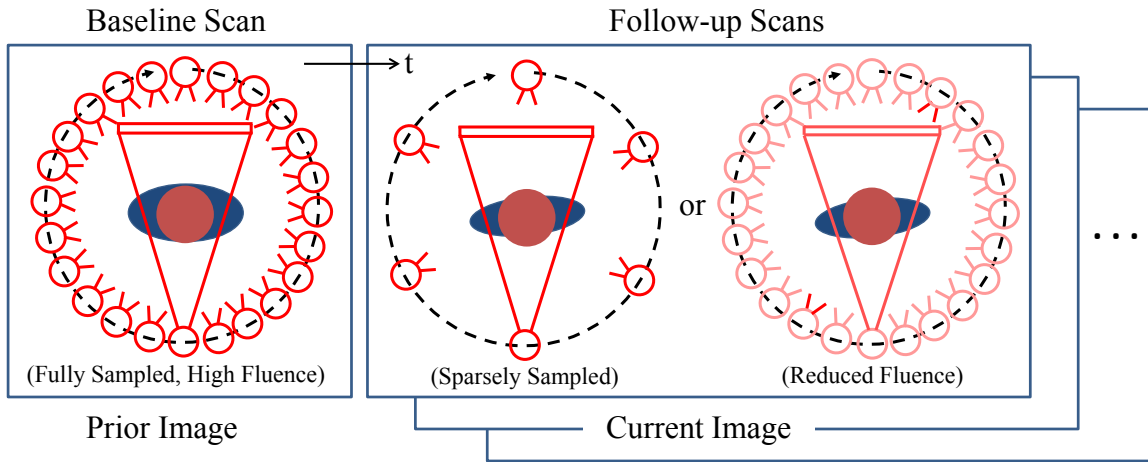


Figure 1.1: Two potential methods for dose reduction in sequential imaging studies: sparse sampling (fewer projections acquired in the scan) and reduced fluence (lower dose per projection).

Patient-specific prior information, such as a prior image of the same patient acquired in a previous imaging study, is often available in sequential imaging studies. Such information contains a tremendous amount of patient-specific anatomical information that could potentially be leveraged in image reconstruction, presenting increased opportunities for dose reduction. While conventional analytical reconstruction methods such as filtered-backprojection (FBP) are usually derived using data only from the current scan, MBIR allows the incorporation of prior information into image reconstruction through regularization and thereby represents a promising approach to leverage patient-specific prior information. However, the use of patient-specific prior images in image reconstruction also introduces new challenges. For example, possible patient motion between imaging studies needs to be compensated to ensure accurate use of patient-specific prior information. Moreover, a proper balance between information from a prior image and from the current scan is needed to ensure accurate admission of anatomical changes from the prior image. New MBIR methods will be introduced in Chapter 5 and 6 to leverage patient-specific prior information and address these new challenges.

## **1.2 Cone-beam computed tomography**

### **1.2.1 Principles and technology**

In a broad sense, a CT system whose scan geometry has a fairly large cone angle may be referred to as cone-beam CT (CBCT). A major advantage of using a large cone angle in CT compared to a small cone angle or a fan-beam geometry is the increased coverage of the object by x-ray in each projection. Such a configuration allows imaging of a full volume from a single orbit about the patient.

Arguably, modern MDCT with a detector comprising a large number of detector rows amounts to a “cone-beam” imaging system. The number of detector rows in MDCT has

dramatically increased in the past two decades. For example, a modern MDCT scanner may have as many as 128, 320, or even 640 detector rows, providing longitudinal coverage of 8-16 cm of coverage.

This section focuses on CBCT systems based on a large-area detector, usually a large-area FPD. Large-area FPDs were initially developed for radiographic / fluoroscopic imaging in the 1990s.<sup>58-61</sup> FPDs include two general varieties of detection mechanism – indirect FPDs and direct FPDs.<sup>4</sup> The former type converts incident x-ray photons into light photons in the scintillator (usually gadolinium oxysulfide,  $\text{Gd}_2\text{O}_2\text{S:Tb}$ , or cesium iodide,  $\text{CsI:Tl}$ ) and then converts light photons into electronic charge using an array of amorphous silicon ( $\text{a-Si:H}$ ) photodiodes and thin-film transistors (TFTs). The latter type directly converts incident x-ray photons into electron-hole pairs (typically in amorphous selenium,  $\text{a-Se}$ , or another photoconductor) and then detects electrons using arrays of high-voltage bias electrodes together with TFTs.

FPD-based CBCT systems have a number of distinct features compared to MDCT. First, FPD-based CBCT systems have a (nearly) full cone-beam geometry (i.e., the cone angle is comparable to the fan angle) and provide volumetric image reconstruction from a circular orbit. This is because a modern FPD has a large number of detector “rows”. For example, an indirect FPD used in studies below has 3072 pixels in the vertical direction (detector “rows”) and 3072 pixels in the horizontal direction. Second, the pixel size in FPDs is usually small and isotropic (e.g., a pixel pitch of 139  $\mu\text{m}$  for the indirect FPD mentioned above), so FPD-based CBCT systems are capable of providing images with isotropic, sub-mm spatial resolution. Third, FPD-based CBCT usually allows more flexibility in the design and optimization of the system for specialty applications compared to MDCT. Some design processes leverage the understanding of 3D imaging performance provided by cascaded systems modeling,<sup>62-65</sup> which are extensions from previous work in 2D.<sup>66-72</sup> Fourth, FPD-based CBCT systems tend to have mechanical simplicity, open geometry, small footprint, portability, and low cost compared to MDCT systems. These features

allow them to be deployed outside radiology suites such as in doctors' offices, surgical rooms, and intensive care units.

### **1.2.2 Imaging applications**

FPD-based CBCT systems (simply referred to as CBCT below) have been developed since the 2000s and since then have found application in many areas in diagnostic imaging and image-guided interventions.

One important area of applications of CBCT is dental, maxillofacial, and otologic imaging. Dedicated CBCT systems for these applications have been commercially available and widely used in the clinical environment. These systems provide spatial and contrast resolution suitable to visualization of high-contrast anatomical structures in these applications (e.g., for dental implant placement planning or for determining osseous lesion in head and neck) at doses comparable to or less than those for MDCT.<sup>15–18</sup>

Dedicated CBCT systems have also been developed for imaging of upper and lower extremities. For example, one system has been designed and optimized using cascaded systems analysis without costly repeated experimentation,<sup>73</sup> and a prototype has been developed through a combined process of design specification, image quality assessment, clinical feedback, and revision.<sup>25</sup> As demonstrated by this system and other systems,<sup>26–29</sup> dedicated extremity CBCT can provide excellent visualization of bone detail and good visibility of soft-tissue suitable to a broad spectrum of musculoskeletal applications. In addition, dedicated extremity CBCT allows for imaging of the lower extremities during weight-bearing, which opens new possibilities to study degenerative joint diseases in the lower extremities.

Dedicated CBCT for breast imaging is another promising area of application. The development of dedicated breast CBCT began in the early 2000s.<sup>19,20,74</sup> Current systems provide

spatial and contrast resolution suitable to soft tissue delineation and calcification detection in the breast.<sup>22,23</sup> These systems typically involve dose levels comparable to digital mammography but provide 3D volumetric images of the breast as opposed to 2D projection images in mammography. More recent developments leveraging complementary metal-oxide semiconductor (CMOS) detectors<sup>75</sup> and photon-counting detectors<sup>24</sup> in breast CBCT have shown improved visualization of microcalcifications and reduced dose, which promises clinical translation of low-dose breast CBCT.

CBCT has also played an important role in the advance of IGRT. CBCT began to be integrated with radiation treatment systems in the early 2000s.<sup>35–37</sup> The system integration usually involves incorporation of a kilovoltage x-ray source and a FPD mounted onto the treatment gantry that supports the megavoltage treatment source. Such CBCT systems provide 3D volumetric images of the patient compared to 2D projection images in earlier radiographic systems and improved soft-tissue contrast resolution compared to portal imaging. CBCT is widely used in IGRT mainly to guide patient setup before the treatment. This also allows the use of a smaller safety margin between the planning target volume (PTV) and clinical target volume (CTV) and the use of dose escalation during the treatment.<sup>33,76–78</sup>

CBCT also represents a promising imaging modality for image-guided surgery. Conventional surgical guidance systems rely on pre-operative images of the patient (e.g., CT or MRI) for inter-operative guidance.<sup>42</sup> One problem associated with this workflow is that changes in the patient anatomy during surgery due to resection (e.g., resection of bones in the sinuses to access pituitary tumor in skull base surgery) or patient deformation (e.g., brain midline shift during surgery) are not reflected in the pre-operative images used in surgical guidance. CBCT – for example, implemented on a mobile surgical C-arm – provides the ability to acquire high-quality 3D volumetric images of the up-to-date anatomy during surgical intervention. Integration of such systems with tracking (optical and electromagnetic), registration (rigid and deformable), and visualization (3D rendering and virtual reality) in conventional surgical guidance systems offers

potential benefits in a variety of interventions, including head and neck, spine, orthopaedic, thoracic, and abdominal surgeries.<sup>35,55,56,79</sup>

### **1.2.3 Challenges and opportunities**

A major challenge in CBCT is that the image quality of the current generation of CBCT systems tends to be insufficient for low-contrast, soft-tissue visualization.<sup>18</sup> CBCT has shown great promise for a number of applications, particularly for imaging of bony structures, but low-contrast, soft-tissue visualization requires a fairly high level of uniformity and a low level of noise in the image. Image uniformity in CBCT is often impaired by artifacts that arise from non-idealities in FPDs, x-ray scatter (which is increased in cone-beam geometry compared to fan-beam geometry), beam hardening, cone-beam sampling (resulting in so-called “cone-beam” artifacts), data truncation, and patient motion. Moreover, image noise in CBCT is affected by a number of factors, including scan dose, artifact correction, and the selection of image reconstruction methods. Further improvement in low-contrast, soft-tissue visualization in CBCT is necessary for certain applications in diagnostic imaging (e.g., detecting hemorrhage in the brain) and image-guided interventions (e.g., using CBCT images acquired during radiation treatment for adaptive treatment planning such as tumor delineation and dose calculation).

Closely related to the challenge of low-contrast, soft-tissue visualization in CBCT is the challenge of accumulated radiation dose when using CBCT in sequential imaging studies. Although CBCT systems equipped with the state-of-the-art FPDs tend to involve lower dose than MDCT,<sup>17,22,26,77,79</sup> the accumulated radiation dose in sequential imaging studies is a concern to the patient and potentially a limiting factor for more extensive use of CBCT in sequential studies. For example, if the dose associated with each CBCT scan could be reduced, then the clinician will be better able to acquire repeated CBCT scans in the interest of increased precision in patient setup and re-planning.

Along with the aforementioned challenges in current CBCT systems, opportunities for improving image quality and reducing dose in CBCT arise together with ongoing advances in detector technology, artifact correction, and MBIR. For example, new FPDs using CMOS image sensors have exhibited a number of desirable characteristics, including lower electronic noise and higher frame rate compared to conventional FPDs, presenting a potential for reducing noise and motion artifacts in CBCT images.<sup>80</sup> Another example is the use of a graphics processing unit (GPU) for accelerated Monte Carlo based scatter correction, allowing accurate scatter correction with a reasonably fast computation time.<sup>81,82</sup> Further opportunities related MBIR will be discussed in Sec. 1.4.

### **1.3 Filtered backprojection**

Filtered backprojection (FBP) is an analytical tomographic reconstruction method. The basic FBP method for parallel-beam geometry was used to reconstruct images in the first-generation CT scanners in the 1970s.<sup>83</sup> As the evolution of CT technology, modified FBP methods have been developed and become commercially available for fan-beam geometry (third-generation CT scanners) and cone-beam geometry (multi-detector CT scanners). FBP has come to represent the most prevalent method for 3D image reconstruction and often presents the standard of comparison for newly developed reconstruction methods such as MBIR. This section briefly reviews FBP, including basic principles, extension for practical systems, and drawbacks.

#### **1.3.1 Idealized continuous model**

In CT, x-rays are transmitted through the object and measured on a detector array. An idealized continuous model assumes both the object and detector array to be continuous. Although this model differs from the real world in which the detector array is discrete and comprised of individual

detector elements each with a finite width, the continuous assumption is convenient for mathematical derivation of the FBP reconstruction method.

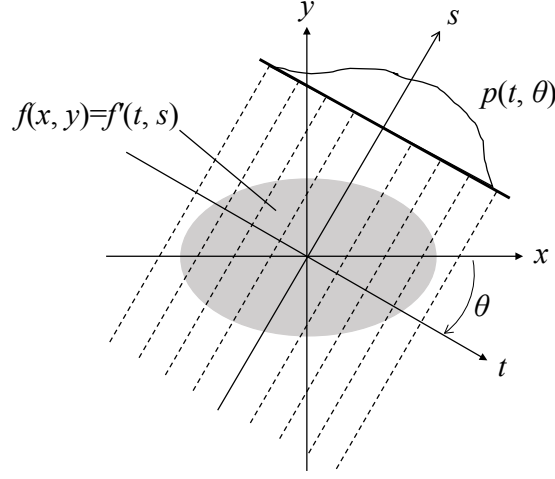


Figure 1.2: Illustration of a projection along a certain angle in a continuous model assuming a continuous object and a continuous detector system.

An illustration of the continuous model is shown in Fig. 1.2. In this example, a 2D object in parallel-beam geometry is considered. Let  $f(x, y)$  denote an object containing attenuation values in the coordinate system  $(x, y)$ , which can also be represented by  $f'(t, s)$  in the rotated coordinate system  $(t, s)$ . The two coordinate systems can be related by:

$$\begin{cases} t = x \cos \theta + y \sin \theta \\ s = -x \sin \theta + y \cos \theta \end{cases} \quad (1.1)$$

Assuming a continuous detector system is oriented at the angle  $\theta$ , a projection of the object in the rotated coordinate system  $(t, s)$  can be written:

$$p(t, \theta) = \int_{-\infty}^{\infty} f'(t, s) ds \quad (1.2)$$

The resulting projection  $p(t, \theta)$  is the line integral of the object  $f'(t, s)$  along a specific line determined by  $t$ , which models the physical process in CT in which a projection measures the total

amount of attenuation by the object along a specific radiographic ray. The Eq. (1.2) is also known as the 2D Radon transform.

### 1.3.2 Filtering and backprojection

Given a set of measured projections, a straightforward approach to reconstruct the image is to project the measured values from the projection domain back to the image domain. This process is called backprojection, which paints the entire ray path with the same value as in the measured projection. However, simple backprojection of all the projections yields a blurred version of the true object – specifically, a convolution of the true object with a  $1/r$  low-pass filter in spatial domain. On the contrary, FBP removes such  $1/r$  blur by multiplying with a ramp function (a high-pass filter) in the frequency domain.

The basic principles of FBP are illustrated in the 2D, parallel-beam example in Fig. 1.2. Let  $P(\omega, \theta)$  denote 1D Fourier transform of a projection at angle  $\theta$ :

$$P(\omega, \theta) = FT\{p(t, \theta)\} \quad (1.3)$$

FBP filters the projection by multiplying with a ramp function  $|\omega|$  in the Fourier domain and yields a filtered projection  $g(t, \theta)$  in the spatial domain after inverse Fourier transform:

$$g(t, \theta) = \int_{-\infty}^{\infty} P(\omega, \theta) |\omega| e^{j2\pi\omega t} d\omega \quad (1.4)$$

An image is then computed by backprojecting the filtered projection  $g(t, \theta)$  at all angles:

$$f(x, y) = \int_0^\pi d\theta \int_{-\infty}^{\infty} P(\omega, \theta) |\omega| e^{j2\pi\omega t} d\omega \quad (1.5)$$

Derivation of the FBP result in Eq. (1.5) has been described in many textbooks<sup>83,84</sup> and is not further detailed here. However, it is worth mentioning that a number of extensions from Eq. (1.5) need to be made for practical systems, including: 1) application of a window function to obtain

a band-limited ramp filter and thereby reduce amplification of high-frequency noise; 2) adjusting the shape and cutoff of the window function to achieve a particular, desired noise-resolution tradeoff; and 3) padding each projection with a sufficient number of zeros prior to Fourier transform and filtering operation to avoid “wrap-around” effects. Another necessary extension is the discretization of the object and the detector (which are assumed continuous in this section) and the implementation of a discrete backprojector, which will be discussed in Sec. 1.4.3.

The FBP results above consider parallel-beam geometry, whereas modified FBP methods have also been developed for fan-beam geometry and cone-beam geometry. An image of the object scanned in fan-beam geometry can be reconstructed either using fan-beam FBP methods (which contain additional weights associated with the fan angle compared to parallel-beam FBP methods) or by first rebinning the fan-beam projections into parallel-beam projections and then using parallel-beam FBP methods.<sup>83,84</sup> For cone-beam geometry, one of the most popular direct reconstruction algorithm is the Feldkamp-Davis-Kress (FDK) algorithm,<sup>85</sup> which is similar to fan-beam FBP methods in that it also contains additional weights associated with the cone angle. The FDK algorithm will be used to perform FBP reconstruction of cone-beam CT data in this dissertation. Similar to fan-beam geometry, cone-beam projections can also be reconstructed by first rebinning into tilted fan- or parallel-beam projections and then using fan- or parallel-beam FBP methods.<sup>86,87</sup> The development of analytical cone-beam reconstruction methods for improved accuracy, computational efficiency, and/or nonconventional trajectories is an ongoing research area.<sup>88–90</sup>

FBP also has a number of drawbacks. Although it has been widely used in the past four decades, it may not ultimately be the best choice for CT reconstruction, especially when taking into account the challenges / opportunities described in Sec. 1.1 and 1.2. For example, FBP does not consider the statistical distribution (i.e., the “noise”) associated with the measurements and instead treats all the measurements equally. This makes the image quality in FBP susceptible to increased

noise in the measurements and impedes the use of dose reduction techniques. Moreover, a number of physical effects are not considered or modeled in FBP, such as the x-ray focal spot size, detector footprint, and voxel size, which can lead to reduced image quality in some scenarios. Furthermore, the derivation of FBP assumes sufficient angular sampling of the object in a circular orbit. As a result, it is difficult to adopt new sampling schemes such as noncircular orbits and/or angular undersampling in FBP, which would otherwise introduce artifacts such as streaks and shading in the image. These and other drawbacks of FBP have motivated the development of MBIR, which is introduced in greater detail in the next section.

## **1.4 Model-based iterative reconstruction**

Whereas FBP offers a closed-form analytical solution to reconstruction of an object from its projections, MBIR usually involves formulating an objective function based on a forward model and solving it using iterative algorithms. Iterative algorithms are needed because either MBIR does not have a closed-form solution or the closed-form solution is computationally impractical. Use of MBIR for transmission and emission tomography dates to the origin of these modalities in the 1970s. For example, it appears that the first paper on the maximum likelihood method for transmission tomography was published by Rockmore and Macovski in 1977.<sup>91</sup> MBIR began to be used in a clinical setting in emission tomography in the 1990s because of the relatively high level of noise in the measurements, advances in computationally efficient algorithms and computing power, and the small image matrix sizes in emission tomography compared to transmission tomography.<sup>4,92</sup> MBIR has become an increasingly active research area for transmission tomography since the 2000s. Since 2010, MBIR algorithms have become commonly available (even if as an optional feature) on commercial MDCT systems in the clinical setting<sup>4,13</sup> and have shown great promise in improving image quality and/or reducing radiation dose for CBCT systems.<sup>93</sup>

This section first discusses the advantages of MBIR in comparison to FBP, then introduces some basic models in MBIR including the discrete forward model and noise model, and finally introduces a few widely used MBIR methods. The basic concepts described below also provide background material for the development of new MBIR methods in subsequent chapters.

### **1.4.1 Advantages of MBIR**

MBIR has a number of advantages over conventional analytical reconstruction methods such as FBP. First, MBIR attempts to make more efficient use of the measurements by incorporating a forward model of the physical detection processes, including noise in the measurements,<sup>94,95</sup> extended source and detector footprints / blur,<sup>96–99</sup> spectral effects,<sup>11,12,100–102</sup> scatter,<sup>101,103</sup> and complex geometries.<sup>104,105</sup> Second, MBIR allows incorporation of additional sources of information in the form of regularization terms, including general image properties such as smoothness and edge-preservation<sup>106–109</sup> and patient-specific anatomical information from previous imaging studies.<sup>110,111</sup>

The advantages of MBIR have made it better suited than FBP to many of the challenges / opportunities mentioned in Sec. 1.1 and 1.2. For example, MBIR has demonstrated significantly improved tradeoffs between image quality and radiation dose compared to FBP in MDCT<sup>12,13</sup> and the potential for similar improvements in CBCT.<sup>93</sup> Moreover, for dedicated imaging systems and systems for image-guided interventions, MBIR allows accurate modeling of many non-ideal effects that would otherwise adversely affect image quality. For example, MBIR with explicit modeling of focal spot blur, detector blur, and correlated noise has yielded improved noise-resolution tradeoff in dedicated extremity CBCT.<sup>99</sup> To take another example, MBIR has allowed the use of nonconventional scan orbits to improve the imaging performance compared to a circular orbit on a robotic C-arm.<sup>105</sup> Furthermore, MBIR provides a useful means in sequential imaging studies to

incorporate patient-specific prior image information into the reconstruction of a subsequent image.<sup>110,111</sup>

### **1.4.2 Models in CT reconstruction**

Two points regarding modeling in CT reconstruction deserve mention. First, MBIR as defined in this dissertation is not alone among image reconstruction methods in requiring a model of some physical process or aspect of the system in image formation. For example, even FBP involves at least a model of the detector array, object, and the backprojection process. One advantage of MBIR over FBP is more efficient use of the available data through incorporation of more sophisticated models of the physical detection processes. Second, while a more sophisticated model is likely to capture more aspects of the real image formation process, that alone does not guarantee a preferable image reconstruction method. For example, a more sophisticated model may also require greater effort to construct (e.g., measurement of physical parameters in the model), compute, and analyze. Instead, considerable effort is given to identify the most useful approximation of reality based on the specific problem being solved. As Dr. George Box said in the book *Statistics for Experimenters*<sup>112</sup>:

*"The most that can be expected from any model is that it can supply a useful approximation to reality: All models are wrong; some models are useful."*

### **1.4.3 Discrete forward model**

Although the continuous model introduced in Sec. 1.3.1 is convenient for derivation of the FBP process, its assumption on continuous measurements deviates from the real situation in which only discrete measurements are collected. A more realistic model includes discrete measurements on a pixelated detector array. Moreover, it is often common to adopt a discrete model of the object in

image reconstruction using a linear combination of basis functions, considering that the reconstructed image will ultimately be viewed on a digital display with discrete pixels. The discretization of the object most commonly takes the form of voxels, while other basis functions such as natural pixels<sup>113</sup> and spherically symmetric elements (“blobs”)<sup>114</sup> can be used.

A discrete forward model based on discrete measurements and discrete voxels in the image is introduced as follows. Let  $\mu$  denote a discrete image of attenuation values that has been ordered into a  $N_\mu \times 1$  vector ( $N_\mu$  is the total number of voxels in the image), and let  $l$  denote discrete line integrals that have been similarly ordered into a  $N_y \times 1$  vector ( $N_y$  is the total number of measurements). The forward projection operation is represented by a  $N_\mu \times N_y$  system matrix  $\mathbf{A}$ , whose element  $\mathbf{A}_{ij}$  denotes the line segment along the  $i^{\text{th}}$  ray through the  $j^{\text{th}}$  voxel. The backprojection operation is denoted as  $\mathbf{A}^T$ . The forward projection expressed in the continuous model in Eq. (1.3) can be re-written in a discrete form:

$$l_i = \sum_{j=1}^{N_\mu} \mathbf{A}_{ij} \mu_j \quad (1.6)$$

Now consider  $\bar{y}$  denoting a  $N_y \times 1$  vector representing the mean measurements, which can be related to the line integrals  $l$  via the Beer-Lambert Law:

$$\bar{y}_i = g_i \exp(-l_i) + r_i \quad (1.7)$$

where  $g$  denotes a  $N_y \times 1$  vector representing measurement-dependent gains, and  $r$  denotes a  $N_y \times 1$  vector representing other detected signals (e.g., x-ray scatter). Equations (1.6) and (1.7) can also be written in vector form:

$$l = \mathbf{A}\mu \quad (1.8)$$

$$\bar{y} = \mathbf{D}\{g\} \exp(-l) + r \quad (1.9)$$

where  $\mathbf{D}$  is an operator that converts a vector into a diagonal matrix. Equations (1.8) and (1.9) (or Equations (1.6) and (1.7)) together are referred to as the *forward model* in MBIR, which relates the image estimate  $\mu$  to the mean measurements  $\bar{y}$ . Note that this particular forward model assumes mono-energetic x-rays and independent measurements. One may formulate a more sophisticated forward model that considers more physical effects in image formation such as poly-energetic x-rays.<sup>100,101</sup>

Many methods have been developed to implement the forward projection (denoted  $\mathbf{A}$ ) and backprojection (denoted  $\mathbf{A}^T$ ) operations. One such method is the “voxel-driven” approach, which, for example in backprojection, computes the intersection of the ray passing through the voxel of interest and the detector array, then interpolates the projection value at the intersection using neighboring projection samples, and then copies the interpolated value back to the voxel of interest. Methods of accelerating the voxel-driven approach have been developed, such as pre-upsampling the projection samples before 1D interpolation.<sup>115,116</sup> Another method is the “ray-driven” approach, which, for example in forward projection, traces each ray through the image and approximates the line integrals as a weighted sum of the values of all the voxels close to the ray. One weighting mechanism weights each voxel value by the intersection length within the voxel, which can be efficiently computed using methods such as that developed by Siddon.<sup>116,117</sup> More recently developed approaches make more extensive modeling of the footprints of the image voxel and detector pixel than the “voxel-driven” and “ray-driven” approaches. For example, the “distance-driven” approach by De Man and Basu<sup>97,118</sup> maps the boundaries of the image voxel and detector pixel to a common axis, and projects the data from one set of boundaries to another via a 1D kernel operation. The “separable footprints” approach by Long *et al.*<sup>98</sup> maps the image voxel onto the detector array and approximates the voxel footprint functions as 2D separable functions (trapezoid or rectangular function), which greatly simplifies the calculation of integration over a detector pixel. In addition to computational efficiency, the “separable footprints” approach has also shown

improved accuracy over more conventional approaches especially for cone-beam geometry.<sup>98</sup> It is therefore used for forward projection and backprojection operations in MBIR in this dissertation.

#### 1.4.4 Noise model

One of the main advantages of MBIR is the ability to incorporate a noise model in image reconstruction. Noise in CT can come from a variety of sources, such as quantum noise associated with the detection of x-ray photons, electronic noise (e.g., thermal noise and shot noise) in the electronic detector systems, and structure noise associated with detector readout in parallel channels.<sup>4</sup> The proper selection of a noise model often depends on the specific type of systems and applications. Under the assumption of independent measurements, a simple and commonly used noise model assumes quantum noise as the dominating noise source, which leads to Poisson-distributed measurements:

$$y_i \sim \text{Poisson} \{ \bar{y}_i \} \quad (1.10)$$

where  $y$  and  $\bar{y}$  denote the actual measurements and mean measurements respectively. If one also considers additive electronic noise in the measurements and models it as independent Gaussian noise, a Poisson-Gaussian mixture model<sup>119</sup> can be written:

$$y_i \sim \text{Poisson} \{ \bar{y}_i \} + \text{N} \{ 0, \sigma_e^2 \} \quad (1.11)$$

where  $\sigma_e^2$  denotes the variance of the Gaussian noise. This model is expected to be especially useful when the level of detected photons is low and the effects of electronic noise need to be considered. Moreover, a Gaussian noise model with nonuniform variances may be used to account for multiple noise sources:

$$y_i \sim \text{N} \{ \bar{y}_i, \sigma_i^2 \} \quad (1.12)$$

where  $\sigma_i^2$  denotes the variance associated with the  $i^{\text{th}}$  measurement.

In addition to modeling noise in the measurements, in some cases, modeling the noise in the data in the form of line integrals is needed in MBIR. One method to obtain a noise model in the line integrals is to propagate the noise characteristics from the measurements to the line integrals through a data transformation. Such data transformation can be simply a log transformation or a log transformation plus other data processing steps, such as artifact correction and/or sinogram denoising. The former case will be discussed in Sec. 1.4.9, and the latter case will be covered in Chapter 2.

### 1.4.5 Algebraic reconstruction technique

The algebraic reconstruction technique (ART) is a generic tomographic reconstruction method that formulates the image reconstruction problem as a system of linear equations. In the context of CT image reconstruction, ART attempts to solve the system of linear equations:

$$l = \mathbf{A}\mu \quad (1.13)$$

where  $l$  here denotes *measured* line integrals computed from the measurements  $y$ , as opposed to *estimated* line integrals computed by forward projecting the image estimate  $\mu$  in Eq. (1.8). While the original ART algorithm<sup>120,121</sup> solves Eq. (1.13) sequentially by updating the image volume on a ray-by-ray basis, a variant of the ART algorithm named simultaneous ART (SART)<sup>122</sup> updates the image volume simultaneously using all the rays, showing a computational advantage over the original ART algorithm. The simultaneous update equation of SART for CT image reconstruction can be written:

$$\hat{\mu}_j = \hat{\mu}_j + \frac{\sum_{i=1}^{N_y} a_{ij} \frac{l_i - [\mathbf{A}\hat{\mu}]_i}{\sum_{j=1}^{N_\mu} a_{ij}}}{\sum_{i=1}^{N_y} a_{ij}} \quad (1.14)$$

which involves computing the mismatch between the measured and estimated line integrals and backprojecting the mismatch to the image estimate after some normalization.

Neither ART nor SART assumes uniformly distributed angular sampling in a circular orbit as assumed in FBP. Therefore, ART can be useful for CT systems that acquire projections only at several angular positions, such as digital tomosynthesis systems.<sup>4</sup> However, a drawback of ART is that it does not consider noise in the measurements (which is also the case for FBP) and attempts to estimate an image volume that directly matches the noisy measurements in CT. As a result, ART often produces images with a fairly high level of noise when the scan dose is low.<sup>123</sup> In addition, ART assumes that data consistency can be enforced through a linear model (as in Eq. (1.13)), which makes it difficult to include nonlinear effects such as polyenergetic x-ray into image reconstruction.

#### 1.4.6 Total variation minimization

Total variation (TV) minimization was first introduced by Rudin *et al.*<sup>124</sup> to remove noise in images. One advantage of using TV minimization in image denoising is that it can smooth the noise while preserving sharp edges and not cause over/under shooting and ripples.<sup>108,124</sup> Because of this, TV minimization has also been used to suppress noise in image reconstruction for a variety of systems, including tomosynthesis, MDCT, CBCT, micro-CT, and nuclear medicine.<sup>109,125–127</sup>

The TV term is computed based on the gradient information in the image, which can be expressed:

$$\text{TV}(f) = \sum_{x,y,z} \sqrt{(f_{x+1,y,z} - f_{x,y,z})^2 + (f_{x,y+1,z} - f_{x,y,z})^2 + (f_{x,y,z+1} - f_{x,y,z})^2} \quad (1.15)$$

where  $f$  denotes a 3D image whose voxels are indexed by the subscripts  $x$ ,  $y$ , and  $z$ . One way to use TV minimization in CT image reconstruction is to treat TV minimization as the objective function and treat data consistency as an equality or inequality constraint,<sup>109,127</sup> resulting in a constrained optimization problem. One example is an algorithm developed by Sidky and Pan for CBCT,<sup>109</sup> whose objective function can be written:

$$\mu = \arg \min_{\mu} \text{TV}(\mu) \quad \text{subject to} \quad \|\mathbf{A}\mu - l\| \leq \varepsilon \quad (1.16)$$

where  $\text{TV}(\mu)$  denotes the TV term defined in Eq. (1.15) and computed for the image estimate  $\mu$ , and  $\varepsilon$  is a small positive number. The TV term is minimized by steepest descent with an adaptive step-size, and the constraints (including a non-negativity constraint not explicitly shown in Eq. (1.16)) are enforced using a method named projection onto convex sets. Although this algorithm still uses a linear model as in ART, it appears to handle the data inconsistency problem (discussed in Sec. 1.4.5) better than ART for two reasons. First, an inequality constraint is used in this algorithm intended to accommodate sources of data inconsistency such as noise. Second, the TV minimization can help choose a unique solution from all the solutions that satisfy the inequality constraint.<sup>109</sup> This algorithm has shown reduced cone-beam artifacts and improved image quality under angular undersampling in CBCT.<sup>109</sup> However, similar to ART, this algorithm and many other TV-based reconstruction methods<sup>125–127</sup> still do not use a noise model when enforcing data consistency and instead attempt to suppress noise in the reconstructed image by minimizing TV. Previous work has shown that images reconstructed by TV minimization tend to exhibit unphysical patchy appearance when the scan dose is low and a large amount of TV minimization is applied to suppress noise in the image.<sup>128</sup>

### 1.4.7 Maximum-likelihood estimation

This section and the following two sections (Sec. 1.4.8 and Sec. 1.4.9) introduce MBIR methods that consider the statistical distribution (i.e., the noise) in the measurements. When a statistical noise model is used, maximum likelihood (ML) estimation is a natural approach to estimate an image from a particular measurement realization. The objective function of ML estimation can be written:

$$\hat{\mu} = \arg \max_{\mu} \log L(y; \mu) \quad (1.17)$$

where  $L(y; \mu)$  denotes the likelihood of observing the particular measurements  $y$  given an image estimate  $\mu$ . The ML method seeks an image  $\mu$  that maximizes the likelihood of having observed the particular measurements  $y$ . Assuming independent measurements, the likelihood term can be computed by multiplying the conditional probability density function  $p(y_i | \mu)$  for each measurement  $y_i$ :

$$L(y; \mu) = p(y | \mu) = \prod_{i=1}^{N_y} p(y_i | \mu) \quad (1.18)$$

If one assumes Poisson noise in the measurements, a Poisson conditional probability can be written:

$$p(y_i | \mu) = \exp[-\bar{y}_i(\mu)] \frac{[\bar{y}_i(\mu)]^{y_i}}{y_i!} \quad (1.19)$$

by substituting the Poisson conditional probability in Eq. (1.19) and the forward model in Eq. (1.7) (assuming other signals  $r = 0$ ) into the likelihood in Eq. (1.18), the Poisson log-likelihood can be expressed after dropping constant terms:

$$\log L(y; \mu) \cong \sum_{i=1}^{N_y} h_i([\mathbf{A}\mu]_i) = \sum_{i=1}^{N_y} y_i \log[g_i \exp(-[\mathbf{A}\mu]_i)] - g_i \exp(-[\mathbf{A}\mu]_i) \quad (1.20)$$

where  $h_i$  represents a marginal log-likelihood for each measurement  $y_i$  and  $[\cdot]_i$  denotes an operator that selects the  $i^{\text{th}}$  element of a vector.

Because image reconstruction in CT is usually an ill-conditioned problem, direct maximization of the log-likelihood in Eq. (1.20) often leads to overly noisy images. Although it is possible to obtain a less noisy image using methods such as stopping the iterative algorithm before convergence<sup>129</sup> or post-smoothing the image,<sup>130</sup> another approach is to include a penalty term in the objective function. This approach is called regularization. Using regularization not only improves the conditioning of the image reconstruction problem but also allows one to enforce desired properties in the reconstructed image. Applying regularization to an objective with a likelihood-based data consistency term is often referred to as penalized-likelihood estimation, which will be discussed in the next section.

### 1.4.8 Penalized-likelihood estimation

Penalized-likelihood (PL) estimation considers the statistical distribution (i.e., the noise) in the measurements and enforces desired properties in the reconstructed image through regularization. The objective function of PL estimation can be written:

$$\hat{\mu} = \arg \max_{\mu} \log L(y; \mu) - R(\mu) \quad (1.21)$$

where  $R(\mu)$  denotes a penalty term. The PL estimation can also be interpreted from the perspective of Bayesian statistics. That is, the PL objective function above can be derived by assuming a prior distribution of the image  $\mu$  proportional to  $e^{-R(\mu)}$  and computing the maximum *a posteriori* estimation of the image  $\mu$  following Bayes' rule. This derivation is straightforward and can be found in textbooks,<sup>123</sup> so it will not be repeated here.

The penalty term can take a variety of forms. One commonly used penalty term<sup>123</sup> that penalizes discrepancies between neighboring voxel values is:

$$R(\mu) = \frac{1}{2} \beta \sum_{j \in N_\mu} \sum_{k \in N_j} \omega_{jk} \varphi(\mu_j - \mu_k) \quad (1.23)$$

where the difference between voxel  $\mu_j$  and its neighboring voxel  $\mu_k$  is penalized by a potential function (sometimes referred to as the penalty function)  $\varphi$  and weighted by a directional weight  $\omega_{jk}$ . The weighted function value is summed over all neighboring voxels in a neighborhood (denoted  $N_j$ ) and then summed over all the voxels (denoted  $N_\mu$ ) in the image. The scalar  $\beta$  is called the regularization parameter, which controls the balance between the log-likelihood term and the penalty term.

The penalty formulation in Eq. (1.23) is fairly flexible in allowing one to enforce different desired properties in the image. First, one could use different  $\beta$  values to control the amount of smoothing applied to the image. Second, by changing the directional weight  $\omega_{jk}$ , one can choose different neighborhood systems and weight neighboring voxels accordingly. For example, in a 2D image, one may follow a 4-neighborhood system and set  $\omega_{jk} = 1$  for horizontal and vertical neighboring voxels and 0 for other voxels. Or, one may pursue an 8-neighborhood system by also setting  $\omega_{jk} = 1/\sqrt{2}$  for diagonal neighboring voxels. Third, the selection of a potential function  $\varphi$  is vital in controlling the noise-resolution properties in the image. One of the simplest choices is a quadratic potential function written as:

$$\varphi(x) = \frac{1}{2} x^2 \quad (1.24)$$

where  $x$  is a scalar input. A quadratic potential function penalizes more for larger neighborhood differences including edges in the image and thereby tends to discourage edges (sometimes interpreted as oversmoothing of the image). In comparison, many potential functions such as those developed by Huber<sup>131</sup> and Lange<sup>106</sup> become linear after some point to penalize less heavily than

the quadratic function and thereby suppress noise while preserving edges. The Huber penalty function is:

$$\varphi(x) = \begin{cases} \frac{1}{2\delta} x^2 & |x| < \delta \\ \left| x - \frac{\delta}{2} \operatorname{sgn}(x) \right| & |x| \geq \delta \end{cases} \quad (1.25)$$

where  $\operatorname{sgn}(\cdot)$  denotes the sign function and  $\delta$  denotes the range of the quadratic neighborhood near the origin.

Regularization design in MBIR has been an active research area.<sup>132–138</sup> Chapter 3 discusses in depth a method for regularization design for improving imaging performance with respect to specific tasks in CT image reconstruction. Chapters 5 and 6 introduce a new type of regularization that incorporates patient-specific prior images acquired from previous imaging studies into image reconstruction.

### 1.4.9 Penalized weighted least-squares estimation

The Poisson log-likelihood in the ML estimation and PL estimation is a complex function and can become more complex if a more sophisticated noise model is used. It is therefore natural to ask if there is a simpler data fit method that requires simpler optimization algorithms but still yields adequate images. One such method is penalized weighted least-squares (PWLS) estimation, whose objective function is:

$$\hat{\mu} = \arg \min_{\mu} \frac{1}{2} \|\mathbf{A}\mu - l\|_{\mathbf{W}}^2 + R(\mu) \quad (1.26)$$

where  $l$  denotes the *measured* line integrals and  $\mathbf{W}$  is the diagonal weighting matrix with the  $i^{\text{th}}$  diagonal element  $\mathbf{W}_i$  representing the statistical weight for the  $i^{\text{th}}$  measured line integral. The statistical weights in PWLS estimation are usually chosen as the inverse of the variance of the line

integrals, so PWLS estimation assigns higher weights to rays that have lower variance (i.e., less noisy and of higher fidelity) and assigns lower weights to rays that have higher variance (i.e., more noisy and of lower fidelity). This weighting mechanism makes more efficient use of the measurements than FBP, which does not consider noise in the measurements and treats all measurements equally.

A common choice of statistical weights in PWLS estimation is the measurements ( $y$ ),<sup>13,139</sup> written as:

$$\mathbf{W} = \mathbf{D}\{y\} \quad (1.27)$$

The weighted least-squares approach using the statistical weights above are referred to as the data-weighted least-squares.<sup>123</sup> Interestingly, the data-weighted least-squares may be interpreted from different perspectives. First, if one assumes the measured line integral ( $l_i$ ) is computed from the measurement ( $y_i$ ) through a log transformation, it is easy to show that the inverse of the variance is approximately the mean measurement ( $\bar{y}_i$ ), which can be approximated by the measurement ( $y_i$ ) (as derived in Chapter 2). Therefore, the data-weighted least-squares can be interpreted as a weighting strategy when the data transformation from the measurements to line integrals is a log transformation. In situations where measurements undergo additional data-processing steps such as artifact correction and de-noising in addition to a log transformation, the variance may differ from that in the data-weighted least-squares case and a new weighting strategy may be preferred. A method is introduced in Chapter 2 to design statistical weights given a general data transformation.

Second, the data-weighted least-squares can also be interpreted as a quadratic approximation to the Poisson log-likelihood. This is because the data-weighted least-squares can also be derived from the Poisson log-likelihood in Eq. (1.20) using second-order Taylor expansion.

The derivation has been previously reported<sup>139</sup> and is illustrated here for reader's convenience. First, Eq. (1.20) becomes as follows after dropping a constant term:

$$\log L(y; \mu) \cong -\sum_{i=1}^{N_y} y_i [\mathbf{A}\mu]_i + g_i e^{-[\mathbf{A}\mu]_i} \quad (1.28)$$

Taking the second-order Taylor expansion of  $e^{-[\mathbf{A}\mu]_i}$  about the measured line integral  $[\mathbf{A}\mu]_i = l_i$  yields:

$$\log L(y; \mu) \approx -\sum_{i=1}^{N_y} y_i [\mathbf{A}\mu]_i + g_i \left\{ e^{-l_i} + ([\mathbf{A}\mu]_i - l_i)(-1)e^{-l_i} + \frac{1}{2}([\mathbf{A}\mu]_i - l_i)^2 e^{-l_i} \right\} \quad (1.29)$$

Rearranging the definition of the measured line integral  $l_i = -\log(y_i/g_i)$  into  $e^{-l_i} = y_i/g_i$  and substituting into the equation above yields:

$$\log L(y; \mu) \approx -\sum_{i=1}^{N_y} (1+l_i) y_i + \frac{1}{2}([\mathbf{A}\mu]_i - l_i)^2 y_i \quad (1.30)$$

Finally, dropping the first term (not a function of the image estimate) in the summation yields the data-weighted LS:

$$\begin{aligned} \log L(y; \mu) &\approx -\sum_{i=1}^{N_y} \frac{1}{2}([\mathbf{A}\mu]_i - l_i)^2 y_i \\ &= -\|\mathbf{A}\mu - \mathbf{l}\|_{\mathbf{W}}^2 \quad \text{where } \mathbf{W} = \mathbf{D}\{y\} \end{aligned} \quad (1.31)$$

Therefore, the data-weighted least-squares can be regarded as a quadratic approximation to Poisson log-likelihood. Previous work has shown that such quadratic approximation to Poisson log-likelihood generally holds well for the x-ray fluence range in diagnostic CT.<sup>139</sup>

Interestingly, a data-weighted least-squares approach can also be interpreted as maximum likelihood estimation of the image assuming Gaussian noise in the measured line integrals – more specifically, assuming a Gaussian distribution of the measured line integral ( $l_i$ ) with variance equal

to the inverse of the measurement data ( $y_i$ ). The proof for this result is straightforward. First, the assumption on a Gaussian distribution of the measured line integral can be expressed:

$$l_i \sim N\left([\mathbf{A}\mu]_i, y_i^{-1}\right) \quad (1.32)$$

where  $N(\cdot)$  denotes a Gaussian distribution with the first and second operands corresponding to its mean and variance. Then, the likelihood of having observed a particular measured line integral  $l_i$  given an image  $\mu$  can be expressed:

$$L(l; \mu) = p(l | \mu) = \prod_{i=1}^{N_y} p(l_i | \mu) \quad (1.33)$$

Note that this is the same as Eq. (1.18) except that the observed variables are now line integrals  $l$  instead of measurements  $y$ . Substituting the probability density function defined in Eq. (1.32) into Eq. (1.33) yields:

$$\log L(l; \mu) = \log \prod_{i=1}^{N_y} \frac{1}{\sqrt{2y_i^{-1}\pi}} \exp\left(-\frac{(l_i - [\mathbf{A}\mu]_i)^2}{2y_i^{-1}}\right) \quad (1.34)$$

Finally, rearranging the right-hand side of the equation above and dropping constant terms yield the data-weighted LS:

$$\begin{aligned} \log L(l; \mu) &\cong \sum_{i=1}^{N_y} -\frac{([\mathbf{A}\mu]_i - l_i)^2}{2y_i^{-1}} \\ &= -\|\mathbf{A}\mu - l\|_{\mathbf{W}}^2 \quad \text{where } \mathbf{W} = \mathbf{D}\{y\} \end{aligned} \quad (1.35)$$

## 1.5 Model-based iterative reconstruction in cone-beam CT

CBCT represents a rapidly developing imaging modality finding application in a broad range of diagnostic imaging procedures (e.g., dental, maxillofacial, otologic, extremities, and breast)

and image-guided interventions (e.g., CBCT for IGRT and image-guided surgery). CBCT systems equipped with a FPD typically provide 3D volumetric image reconstruction with sub-mm spatial resolution and soft-tissue visibility in a circular orbit. Moreover, CBCT systems tend to have small footprint, mechanical simplicity, open geometry, portability, and low cost compared to MDCT.

However, the current generation of CBCT systems face challenges in low-contrast, soft-tissue visualization. One example clinical application – detection of acute intracranial hemorrhage at the point-of-care – may greatly benefit from the use of a dedicated head CBCT system. However, such an imaging task requires a fairly high level of contrast resolution (40-80 HU), spatial resolution (0.5-10 mm), and image uniformity, which poses major challenges to the current generation of CBCT systems. Moreover, the use of CBCT in sequential imaging studies is somewhat limited by the accumulated radiation dose in sequential studies. For example, if the dose associated with each CBCT scan can be reduced, then CBCT can be acquired with greater frequency in support of high-precision treatment.

MBIR represents a promising approach to improving image quality and reducing radiation dose in CT and CBCT. MBIR attempts to make more efficient use of the measured data by incorporating a forward model of the physical detection process. Moreover, MBIR allows incorporation of various forms of prior information into the image reconstruction process, typically through regularization. Such prior information ranges from general, desirable image properties such as image smoothness and edge preservation to patient-specific information such as prior images from previous imaging studies. The capability of MBIR to improve image quality and reduce radiation dose has been demonstrated in MDCT and holds similar – or potentially greater – promise in CBCT.

### 1.5.1 Thesis statement

Recognizing the potential of MBIR in CBCT leads us to the following thesis statement and research objective:

Model-based iterative reconstruction (MBIR) methods can improve image quality and/or reduce radiation dose in cone-beam computed tomography (CBCT) through modeling of imaging physics, development of advanced regularization methods, and incorporation of patient-specific prior image information.

The thesis develops a body of work that develops and investigates novel MBIR methods and the associated advantages in improving image quality and reducing radiation dose in CBCT. Specific techniques adopted in this work include: 1) more efficient use of the data by incorporating models of imaging physics; 2) development of advanced regularization techniques to optimize imaging performance with respect to a specific task; and 3) incorporation of patient-specific prior image information that is often available in sequential imaging studies but not leveraged by conventional image reconstruction methods. Specific clinical applications are considered, and novel MBIR methods for CBCT are developed and investigated with respect to image quality and computational efficiency.

### 1.5.2 Thesis outline

The thesis is broadly divided into two parts: first, Chapters 2-4 involve MBIR algorithms that incorporate modeling of imaging physics and advanced regularization; and second, Chapters 5-6 involve MBIR algorithms that incorporate patient-specific prior image information. Specifically:

**Chapter 2** introduces the clinical need for point-of-care detection of acute intracranial hemorrhage (ICH), which is associated with a number of neurological pathologies. Although dedicated CBCT systems are potentially well suited for point-of-care imaging, ICH detection poses major challenges in image quality to the current generation of CBCT systems. A recently developed artifact correction framework overcomes major sources of image non-uniformity and streaks, but also causes an amplification of image noise reconstructed by FBP. This chapter first proposes a general framework for modeling the effects of data corrections on the noise model and then applies the framework to the two dominant artifact correction steps (scatter and beam hardening corrections) required for high-quality CBCT of the head. A MBIR method is then developed that integrates the underlying variations in measurements into a PWLS framework and accounts for noise characteristics following artifact correction by modified statistical weights. The proposed PWLS method is compared to FBP as well as PWLS using conventional statistical weights, and the image quality is evaluated in CBCT images of an anthropomorphic head phantom emulating acute ICH.

**Chapter 3** investigates a novel MBIR approach based on the simple premise that a medical image is always produced to accomplish a particular clinical task (or tasks). Task-based assessment of medical imaging performance has provided a basis for the design and optimization of a variety of medical imaging systems, but comparatively less effort has been made to incorporate task-based modeling of imaging performance into the process of CT image reconstruction. Regularization in conventional MBIR methods is often formulated in generic terms to encourage smoothness and/or sharpness without explicit formulation of the task. This chapter develops a MBIR method that incorporates a model for task-based imaging performance directly and explicitly in the selection of regularization parameters to maximize task-based image quality. Moreover, a framework is introduced that designs a spatially varying penalty to maximize local task-based image quality at every location in the image. The proposed task-

based image reconstruction is evaluated in the context of CBCT of the head introduced in Chapter 2. The imaging tasks considered in Chapter 3 include a mid-frequency task emulating detection of a low-contrast ICH as well as tasks emphasizing other frequency content (low- and high- frequency).

**Chapter 4** addresses a common practical challenge that can confound imaging performance in CBCT – namely, lateral truncation of the projection data by anatomy and/or patient support mechanisms outside the field of view (FOV) of the detector. Such effects present a challenge for FBP and are particularly problematic for MBIR, because MBIR attempts to solve for an image estimate that best matches all of the measurements. For the specific application of head imaging treated in the previous two chapters, a head holder is typically used to support the head and minimize motion during the scan but can be partially truncated in the lateral direction, introducing artifacts that could hinder ICH detection. Increasing the reconstruction FOV beyond the scan FOV can mitigate truncation effects but also increase the computational cost, especially when the truncated object is relatively far from the scan FOV. This chapter introduces a multi-resolution reconstruction approach to mitigate truncation effects by extending the reconstruction FOV without major increase in computational burden. This multi-resolution approach is incorporated into the PWLS reconstruction framework developed in Chapter 2. The method is evaluated in CBCT scans of a head phantom with varying degrees of realistic data truncation by a carbon-fiber head support, focusing on a particular form of data truncation encountered in CBCT of the head but offering a potentially general solution for other scenarios in CT or CBCT as well.

**Chapter 5** considers MBIR in clinical scenarios involving a sequence of imaging studies. Conventionally, each scan is treated in isolation, and the accumulated radiation dose can be a concern to both the patient and surgical staff. Knowledge of patient-specific anatomy gained from an image previously acquired in sequential studies (referred to herein as the "prior image")

can be leveraged in MBIR, thereby improving image quality and/or reducing dose. This chapter introduces a MBIR method that incorporates a Poisson noise model and a high-quality patient-specific prior image to reconstruct images from sparse and/or noisy measurements. Patient motion between the prior image and the current anatomy is estimated jointly through an alternating maximization strategy. Experiments are performed on a cadaver specimen emulating a lung nodule surveillance scenario. A number of aspects of the proposed algorithm are evaluated, including convergence properties, registration accuracy, and the scheduling of registration / reconstruction updates. The performance of the proposed algorithm is compared to alternatives under various conditions of data sparsity and dose.

**Chapter 6** extends the MBIR method of Chapter 5 in relation to a key question regarding to what extent prior image information should be used (via the regularization term associated with incorporation of prior image information) to achieve accurate image reconstruction. For example, using too little prior image information fails to yield a significant benefit with low fidelity data, while using too much prior image information can obscure anatomical changes and produce false structures in the resulting image. This chapter introduces a novel method that prospectively estimates the optimal prior image strength for PIBR without heuristics or exhaustive search. In addition, a spatially varying map of prior image strength is proposed to optimally admit changes everywhere in the image and thereby ensure accurate reconstructions without *a priori* knowledge of the change location. The proposed methodology is evaluated in an ellipse phantom and in a realistic thorax phantom emulating a lung nodule surveillance scenario. The dependence of optimal prior image regularization strength on various properties of the changes (i.e., attenuation, shape, and/or size of features changing between the prior and current image) is also investigated. The proposed method provides a means for prospective patient-, change-, and data-specific customization of the prior image strength to ensure reliable reconstruction of specific anatomical changes.

**Chapter 7** summarizes conclusions, interprets the findings of this dissertation within the broader context of MBIR research and clinical application, and identifies potential avenues of future investigation.

## **Chapter 2**

# **2. Statistical Reconstruction with a Post-Artifact-Correction Noise Model**

## **2.1 Introduction**

### **2.1.1 Clinical motivation and challenges**

Intracranial hemorrhage (ICH) can impart devastating disability with poor prognosis and a high rate of mortality.<sup>140</sup> ICH is associated with many neurological pathologies, including traumatic brain injury, hemorrhagic stroke, postsurgical hemorrhage, and aneurysm. Non-contrast-enhanced multi-detector CT (NC-MDCT) is the most prevalent front-line imaging modality for diagnosis and monitoring of acute ICH, which provides high sensitivity to the presence of fresh blood in the brain. ICH typically presents in NC-MDCT as hyperdense during the hyperacute and acute stages (~40-80 HU contrast within ~3 days) and becomes hypodense during the subacute and chronic stages (-5 to 20 HU contrast in ~10-20 days or longer).<sup>141,142</sup> MRI is the gold standard for diagnosis of chronic-stage ICH and has recently been shown to be potentially suited to the diagnosis of acute ICH as well, although barriers still exist in cost, time, and access.<sup>143,144</sup> Transport of patients in the intensive care unit or neurological critical care unit to either a NC-MDCT or MRI suite requires time and dedicated personnel and brings potential risk to the patient. For instance, such transport has been reported to lead to an average transport time of 50-80 min outside the critical care environment and a 71% incidence of adverse events.<sup>145-147</sup> The importance of diagnosis and

monitoring of acute ICH combined with the risk associated with patient transport motivates the development of point-of-care imaging of acute ICH.

CBCT with a flat-panel detector (FPD) has emerged in the past decade as an invaluable tool for a variety of specialty diagnostic applications, including imaging of the breast,<sup>19,148</sup> musculoskeletal extremities,<sup>25,29</sup> and head and neck.<sup>17,18</sup> While CBCT systems can be well suited to point-of-care applications (offering a small footprint, open adaptable geometry, and relatively low cost), current CBCT systems face significant challenges with respect to image quality in low-contrast, soft tissue imaging tasks such as ICH. Among factors limiting soft-tissue image quality are increased levels of image artifacts and noise when compared to MDCT.

Recent work by Sisniega *et al.*<sup>149</sup> has demonstrated a promising framework for artifact correction in CBCT of the head and brain, including correction of x-ray scatter, beam hardening, image lag, and low-frequency glare/off-focal radiation. The artifact correction framework overcomes major sources of image non-uniformity and streaks and provides a level of image quality potentially suitable to detection of ICH; however, such corrections also cause an amplification of noise in images reconstructed by conventional 3D FBP. As shown in Fig. 2.1(a-b), although the artifact correction framework removes >300 HU of deterministic bias (cupping and streaks primarily associated with scatter and beam hardening) in a head phantom emulating intraparenchymal hemorrhage, the image noise increases by more than a factor of two.<sup>149</sup> Potential solutions for noise reduction include sinogram denoising by adaptive filtering techniques<sup>150</sup> and estimation of the ideal sinogram by minimizing a cost function.<sup>151</sup> Alternatively, the problem can be approached with a MBIR framework, which has demonstrated major improvements in CBCT image quality over conventional FBP for low-contrast, soft tissue imaging tasks.<sup>93</sup>

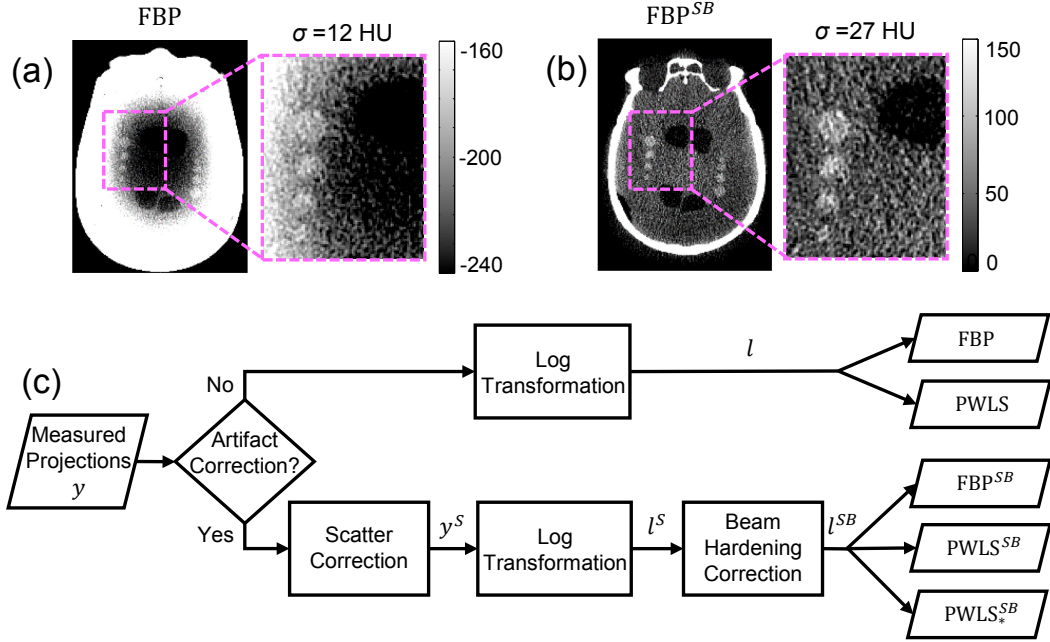


Figure 2.1: Artifact corrections in CBCT of the head. (a) CBCT image reconstructed using FBP without artifact corrections. (b) CBCT image with artifact corrections (scatter and beam hardening, denoted by the superscript  $SB$ ) showing a strong improvement in image uniformity but an amplification of noise by more than a factor of 2. (c) Flowchart of the artifact correction process and image reconstruction methods (FBP and PWLS) investigated in this work.

In this chapter, an MBIR approach is developed that includes the effects of artifact corrections on the underlying noise model, thereby maintaining the benefit of artifact corrections while overcoming the associated noise penalty. The underlying variations in measurements are integrated into a PWLS reconstruction approach, and the noise characteristics following each data correction are accounted by modified weights. Regularization with a Huber penalty on image roughness was used to further improve noise-resolution tradeoffs. Previous analogous work<sup>152</sup> investigated a PWLS objective that included the variations in measurement noise characteristics with scatter correction and quadratic regularization, showing improved image quality in chest CT.

First, a general framework for modeling the effect of data corrections on the noise model is proposed. The framework is then applied to the two dominant correction factors required for high-quality CBCT of the head – scatter and beam hardening (using bone and water segmentation).

Image quality is evaluated in CBCT images of a head phantom emulating intra-parenchymal hemorrhage. The proposed PWLS method using modified weights is compared to FBP as well as PWLS using conventional weights, and the improvements to image noise and/or spatial resolution are quantified.

### **2.1.2 Acknowledgements and unique contributions**

The methods and results reported in this chapter were reported in conference proceedings and journal articles as follows:

- (1) H. Dang, J. W. Stayman, J. Xu, A. Sisniega, W. Zbijewski, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, J. H. Siewerdsen, "Regularization design for high-quality cone-beam CT of intracranial hemorrhage using statistical reconstruction," *SPIE Medical Imaging*, San Diego, CA, Vol. 9783, 97832Y (2016).
- (2) H. Dang, J. W. Stayman, A. Sisniega, J. Xu, W. Zbijewski, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, J. H. Siewerdsen, "Statistical reconstruction for cone-beam CT with a post-artifact-correction noise model: application to high-quality head imaging," *Physics in Medicine and Biology*, 60 (16), 6153-6175 (2015).

with permission from the publishers for reproduction of content in this dissertation. The author's primary contributions in this work were as follows: derivation of the statistical weights following artifact correction; implementation of the reported algorithm; physical experimentation; and quantitative analysis. The author gratefully acknowledges the contributions of coauthors, including: Dr. Alejandro Sisniega for his assistance in head phantom build, data acquisition, and artifact correction; Dr. Jennifer Xu and Dr. Wojciech Zbijewski for their help on modeling of physical processes in CT; Dr. Xiaohui Wang for his input on iterative reconstruction; and Dr. Nafi Aygun for his clinical expertise and advice on brain imaging. The author also gratefully acknowledges all

the other developers (besides the author) of CudaTools, which is a CUDA-based parallel computing toolkit developed within the I-STAR laboratory - specifically, the leading developer Dr. Ali Uneri and core developer Mr. Steven Tilley II. Valuable discussion with Dr. Adam S. Wang on quantitative assessment of soft-tissue image quality in CT, Dr. Grace J. Gang on nonuniformities in CT images, and Dr. John (Ian) Yorkston on flat-panel detectors is also gratefully acknowledged. The work was supported by research collaboration with Carestream Health (Rochester NY).

## 2.2 Theoretical methods

### 2.2.1 PWLS image reconstruction with a noise model for generalized data transformation

The PWLS method<sup>139</sup> weighs the residual error for each measurement by the fidelity of that measurement – specifically an estimate of the inverse of the variance of each measurement, thereby reducing the contributions of low-fidelity measurements and consequently reducing noise in the reconstructed image. The PWLS objective function has been described in Sec. 1.4.3 and is given here again for reader’s convenience:

$$\hat{\mu} = \arg \min_{\mu} \frac{1}{2} \|\mathbf{A}\mu - \mathbf{l}\|_{\mathbf{W}}^2 + R(\mu) \quad (2.1)$$

Recall that  $\mathbf{W}$  is the diagonal weighting matrix with the  $i^{\text{th}}$  diagonal element  $\mathbf{W}_i$  representing the fidelity of the  $i^{\text{th}}$  measurement. In this chapter, a regularization term  $R(\mu)$  was chosen that penalizes first-order neighborhood differences in the image  $\mu$  using the Huber penalty function,<sup>131</sup> which is quadratic within a neighborhood  $([-\delta, \delta])$  and linear for larger differences. The strength of regularization is controlled by the scalar  $\beta$  inside  $R(\mu)$ .

The line integrals in Eq. (2.1) are typically derived from the raw measurements through a number of steps. Such steps include a log transformation to convert from the raw measurement

domain to the line integral domain. In many situations, the measured data are also processed to correct for effects such as x-ray scatter<sup>82,153–156</sup> and beam hardening<sup>157,158</sup> and/or to reduce noise in the measurements (e.g., sinogram smoothing<sup>159</sup>). Such processes lead to potential changes in the noise characteristics of the measurements, which need to be accommodated in the PWLS weighting terms. The processing of the measured data (including log transformation and processes associated with artifact correction) is first modeled as a general function  $f$ , giving processed line integrals expressed as:

$$l_i = f(y_i) \quad (2.2)$$

under the common assumption of independent measurements  $y_i$ . The variance of the line integrals can be derived using first-order Taylor expansion assuming the function  $f$  is differentiable at mean measurements<sup>160</sup> and the second and higher order terms in the Taylor expansion are negligible:

$$\text{var}(l_i) = \text{var}[f(y_i)] = \text{var}\left[f(\bar{y}_i) + \dot{f}(\bar{y}_i)(y_i - \bar{y}_i) + O(y_i^2)\right] \quad (2.3a)$$

where  $\dot{f}$  is the derivative of the function  $f$  and  $\bar{y}$  is a vector of the mean measurements. The approximation in Eq. (2.3a) can be simplified by removing the terms that have zero variance:

$$\text{var}(l_i) \approx \text{var}[\dot{f}(\bar{y}_i)y_i] \quad (2.3b)$$

Since the term  $\dot{f}(\bar{y}_i)$  is a constant, it can be taken out of the variance to yield the final expression of the variance:

$$\text{var}(l_i) \approx [\dot{f}(\bar{y}_i)]^2 \text{var}(y_i) \quad (2.3c)$$

Equation (2.3c) shows that applying a number of operations (modeled as the function  $f$ ) on the raw measurements results in a scaling of the variance by a factor  $[\dot{f}(\bar{y}_i)]^2$ . The PWLS weights can be computed as the inverse of the variance given in Eq. (2.3c):

$$\mathbf{W}_i = \frac{1}{\text{var}(l_i)} \approx \frac{1}{[\dot{f}(\bar{y}_i)]^2} \frac{1}{\text{var}(y_i)} \quad (2.4)$$

which includes the changes in variance associated with the processing operations,  $f$ .

### 2.2.2 Effect of log transformation on variance

First, consider a scenario in which no artifact correction is applied. Assuming Poisson noise in the raw measurements, the forward model in this scenario is written as:

$$y_i \sim \text{Poisson}(\bar{y}_i), \quad \bar{y}_i = g_i \exp(-[\mathbf{A}\boldsymbol{\mu}]_i) \quad (2.5)$$

where  $g_i$  represents measurement-dependent gains. In this scenario, the line integrals in Eq. (2.1) can be computed from the raw measurements simply by log transformation. This process can be simply modeled as:

$$f_l(y_i) = \log\left(\frac{g_i}{y_i}\right) \quad (2.6a)$$

$$l_i = f(y_i) = f_l(y_i) \quad (2.6b)$$

where the log transformation is modeled as function  $f_l(y_i)$  and the general form for  $f$  in Eq. (2.2) is taken in this section to refer specifically to log transformation. Using Eq. (2.3c), and since  $\dot{f}(\bar{y}_i) = -1/\bar{y}_i$ , the variance of the line integrals is:

$$\begin{aligned} \text{var}(l_i) &\approx \left[\dot{f}(\bar{y}_i)\right]^2 \text{var}(y_i) \\ &= \frac{1}{\bar{y}_i^2} \text{var}(y_i) \end{aligned} \quad (2.7)$$

where variations associated with  $g$  have been ignored and accurate measurement-dependent gains<sup>161</sup> have been assumed. Thus, log transformation simply scales the variance by a factor  $1/\bar{y}_i^2$ , giving:

$$\mathbf{W}_i = \frac{1}{\text{var}(l_i)} \approx \frac{\bar{y}_i^2}{\text{var}(y_i)} = \bar{y}_i \approx y_i \quad (2.8)$$

which are the conventional PWLS weights equal to the mean measurements under the usual Gaussian approximation of a Poisson noise model.<sup>13,139</sup> In practice, actual measurement values are typically used in place of the mean measurement values. With additional processing of the

measurements, however, the above assumption may no longer hold, as shown in the following sections that apply the general form of Eq. (2.3c) to the specific cases of scatter correction and beam hardening correction.

### 2.2.3 Effect of scatter correction on variance

Scatter correction typically involves subtraction of a scatter fluence estimate from the measurements, where the scatter fluence may be estimated from the measurements,<sup>154–156</sup> by analytical models,<sup>153</sup> and/or by Monte Carlo calculation.<sup>82,149</sup> Assuming Poisson noise in the raw measurements and including mean scatter fluence as part of the mean measurements, the forward model in this scenario is written as:

$$y_i \sim \text{Poisson}(\bar{y}_i), \quad \bar{y}_i = g_i \exp(-[\mathbf{A}\boldsymbol{\mu}]_i) + \bar{S}_i \quad (2.9)$$

where  $\bar{S}_i$  represents the mean scatter fluence. In this scenario, the line integrals in Eq. (2.1) can be computed by subtracting a scatter fluence estimate from the raw measurements followed by log transformation. Assuming ideal scatter estimation in which the scatter estimate equals the mean scatter fluence, the process can be modeled as:

$$l_i^S = f(y_i) = f_l(y_i - \bar{S}_i) = \log\left(\frac{g_i}{y_i - \bar{S}_i}\right) \quad (2.10)$$

where the superscript  $S$  denotes scatter-corrected line integrals, and the function  $f$  from the general form of Eq. (2.2) refers in this section specifically to scatter correction followed by log transformation. As in the previous section, negligible variation associated with  $g$  is assumed, and the error in the scatter fluence is assumed to be negligible compared to the quantum noise (i.e., a high-fidelity scatter fluence estimate as obtained by accurate Monte Carlo calculation).

The variance of the scatter-corrected line integrals can therefore be computed from Eq. (2.3c), and since  $\dot{f}(\bar{y}_i) = -1/(\bar{y}_i - \bar{S}_i)$ , the variance is:

$$\begin{aligned}\text{var}(l_i^s) &\approx [\dot{f}(\bar{y}_i)]^2 \text{var}(y_i) \\ &= \frac{1}{(\bar{y}_i - \bar{S}_i)^2} \text{var}(y_i)\end{aligned}\tag{2.11}$$

This form shows that the noise is scaled as a result of scatter correction, but the scale factor is different from that in Eq. (2.7). The ratio of the post-correction variance to the pre-correction variance is:

$$\frac{\text{var}(l_i^s)}{\text{var}(l_i)} \approx \frac{\bar{y}_i^2}{(\bar{y}_i - \bar{S}_i)^2} = \frac{(\bar{P}_i + \bar{S}_i)^2}{\bar{P}_i^2} = (1 + \text{SPR}_i)^2\tag{2.12}$$

where  $\bar{P}_i$  is mean primary fluence and scatter-to-primary ratio (SPR) is the scatter-to-primary ratio. This analytical result is consistent with previous work<sup>152</sup> but is derived as a special case of the general form in Sec. 2.2.1. In CBCT systems, where the SPR is often greater than unity<sup>162</sup> and can be as high as ~5-10 in highly attenuated regions of the projection (e.g., the skull base, as shown below), this implies a substantial modification from the conventional PWLS weights – e.g., a 9-fold increase in variance for a SPR of 2.

## 2.2.4 Effect of beam hardening correction on variance

Beam hardening correction typically includes a so-called water correction that compensates for the beam hardening error (e.g., cupping) introduced by soft tissue. When many bony structures are present in addition to soft tissue in a scanned object, a so-called bone correction is often used after water correction to compensate for bone-induced artifacts such as blooming or shading between bones.<sup>157,158</sup> Since ICH can present anywhere throughout the intracranial space (both deep in the parenchyma and immediately adjacent to the cranium), both corrections are needed for high-quality CBCT of the head. The beam hardening correction is modeled as a two-step process of water correction followed by bone correction<sup>157</sup>:

$$l_i^B = f(y_i) = l_i^{Bw} + l_i^{Bb} \quad (2.13a)$$

where the superscript  $B$  denotes beam hardening correction, and the superscripts  $B_w$  and  $B_b$  more specifically denote water and bone correction, respectively. The function  $f$  from the general form in Eq. (2.2) refers in this section specifically to log transformation followed by the two-step beam hardening correction. The water-corrected line integrals can be regarded as a 1-dimensional remapping of measured line integrals based on the calibration of the beam hardening response in water (e.g., from a measured calibration or an analytical model). This 1-dimensional remapping is denoted as  $f_w$  and approximated using polynomial functions of the log transformed data:

$$l_i^{Bw} = f_w(f_l(y_i)) \approx \sum_{u=0}^U \alpha_u [f_l(y_i)]^u = \sum_{u=0}^U \alpha_u \left[ \log\left(\frac{g_i}{y_i}\right) \right]^u \quad (2.13b)$$

where  $U$  and  $\alpha_u$  are the order and coefficients of the water correction polynomial function, respectively. The bone correction can be regarded as a 2-dimensional remapping of water-corrected line integrals and the line integrals of bony structures, which can also be approximated using polynomial functions. This 2-dimensional remapping is denoted as  $f_b$  and the bone correction is modeled based on Eq. (16) in the paper from Joseph and Spital<sup>157</sup> (polynomials not written out for simplicity):

$$l_i^{Bb} = l_i^{bone} \left[ \lambda_0 - f_b(f_w(f_l(y_i)), l_i^{bone}) \right] \quad (2.13c)$$

where  $l_i^{bone}$  denotes the line integrals of the bony structures, and  $\lambda_0$  represents the ratio of the bone mass attenuation coefficient to the water mass attenuation coefficient. The error associated with the calibration of the remapping functions, segmentation of the bony structures, and computation of the measurement-dependent gains are assumed to be small.

The variance of the beam-hardening-corrected line integrals can be derived from Eq. (2.3c):

$$\begin{aligned}
\text{var}(l_i^B) &\approx [\dot{f}(\bar{y}_i)]^2 \text{var}(y_i) \\
&= \left[ \dot{f}_w(f_l(\bar{y}_i)) \dot{f}_l(\bar{y}_i) - l_i^{bone} \frac{\partial f_b(f_w(f_l(\bar{y}_i)), l_i^{bone})}{\partial f_w(f_l(y_i))} \dot{f}_w(f_l(\bar{y}_i)) \dot{f}_l(\bar{y}_i) \right]^2 \text{var}(y_i)
\end{aligned} \tag{2.14a}$$

where  $\dot{f}_w(f_l(\bar{y}_i))$  represents the derivative of the function  $f_w$  with respect to  $f_l(y_i)$  evaluated at  $f_l(\bar{y}_i)$ ,  $\dot{f}_l(\bar{y}_i)$  represents the derivative of the function  $f_l$  evaluated at mean measurement  $\bar{y}_i$ , and  $\partial f_b(f_w(f_l(\bar{y}_i)), \cdot) / \partial f_w(f_l(y_i))$  represents the partial derivative of the function  $f_b$  with respect to the first operand  $f_w(f_l(y_i))$  evaluated at  $f_w(f_l(\bar{y}_i))$ . Rearranging the terms and replacing  $\dot{f}_l(\bar{y}_i)$  with  $1/\bar{y}_i^2$  in Eq. (2.14a) yields:

$$\text{var}(l_i^B) \approx [\dot{f}_w(f_l(\bar{y}_i))]^2 \left[ l_i^{bone} \frac{\partial f_b(f_w(f_l(\bar{y}_i)), l_i^{bone})}{\partial f_w(f_l(y_i))} - 1 \right]^2 \frac{1}{\bar{y}_i^2} \text{var}(y_i) \tag{2.14b}$$

showing that the variance is scaled by two terms corresponding to the variance change in water and bone corrections, respectively. For simplicity, these two terms are denoted as:

$$\begin{aligned}
\eta_w(\bar{y}_i) &= [\dot{f}_w(f_l(\bar{y}_i))]^2 \\
\eta_b(\bar{y}_i) &= \left[ l_i^{bone} \frac{\partial f_b(f_w(f_l(\bar{y}_i)), l_i^{bone})}{\partial f_w(f_l(y_i))} - 1 \right]^2
\end{aligned} \tag{2.14c}$$

where the functional argument  $\bar{y}_i$  in  $\eta_w(\bar{y}_i)$  and  $\eta_b(\bar{y}_i)$  denotes that the derivative of the function  $f_w$  is taken at  $f_l(\bar{y}_i)$  and the partial derivative of the function  $f_b$  is taken at  $f_w(f_l(\bar{y}_i))$ , respectively.

The ratio of the variance after and before beam hardening correction is therefore:

$$\frac{\text{var}(l_i^B)}{\text{var}(l_i)} \approx \eta_w(\bar{y}_i) \eta_b(\bar{y}_i) \tag{2.15}$$

These analytical results of the variance following water and bone beam hardening correction were derived as a special case of the general form in Sec. 2.2.1. However, the results can

also be applied to other scenarios of beam hardening correction in which only the water correction is applied simply by removing the  $\eta_b(\bar{y}_i)$  term in Eq. (2.14) and (2.15).

## 2.2.5 Effect of both scatter and beam hardening corrections: Modified PWLS weights

A PWLS image reconstruction method is proposed that accounts for the changes in variance resulting from artifact correction. Modified weights associated with the post-correction variance are derived to account for both scatter and beam hardening corrections as illustrated in Fig. 2.1. The post-correction variance is derived from the analysis of scatter and beam hardening in the previous two sections, yielding:

$$\text{var}(l_i^{SB}) \approx \frac{1}{(\bar{y}_i - \bar{S}_i)^2} \eta_w(\bar{y}_i - \bar{S}_i) \eta_b(\bar{y}_i - \bar{S}_i) \text{var}(y_i) \quad (2.16a)$$

where the superscript  $SB$  denotes both scatter and beam hardening corrections and

$$\begin{aligned} \eta_w(\bar{y}_i - \bar{S}_i) &= \left[ \dot{f}_w(f_l(\bar{y}_i - \bar{S}_i)) \right]^2 \\ \eta_b(\bar{y}_i - \bar{S}_i) &= \left[ l_i^{bone} \frac{\partial f_b(f_w(f_l(\bar{y}_i - \bar{S}_i)), l_i^{bone})}{\partial f_w(f_l(y_i - \bar{S}_i))} - 1 \right]^2 \end{aligned} \quad (2.16b)$$

The ratio of the variance after and before both corrections is therefore:

$$\frac{\text{var}(l_i^{SB})}{\text{var}(l_i)} \approx (1 + \text{SPR}_i)^2 \eta_w(\bar{y}_i - \bar{S}_i) \eta_b(\bar{y}_i - \bar{S}_i) \quad (2.17)$$

showing the change in variance as a result of scatter correction (by a factor  $(1 + \text{SPR}_i)^2$ ), water correction (by a factor  $\eta_w(\bar{y}_i - \bar{S}_i)$ ), and bone correction (by a factor  $\eta_b(\bar{y}_i - \bar{S}_i)$ ).

The modified PWLS weights can therefore be written as the inverse of the post-correction variance, replacing mean measurements  $\bar{y}_i$  with actual measurement values  $y_i$ :

$$\left[\mathbf{W}_*^{SB}\right]_i = \frac{1}{\text{var}(l_i^{SB})} \approx \frac{(\bar{y}_i - \bar{S}_i)^2}{\eta_w(\bar{y}_i - \bar{S}_i)\eta_b(\bar{y}_i - \bar{S}_i)\text{var}(y_i)} \approx \frac{(y_i - \bar{S}_i)^2}{\eta_w(y_i - \bar{S}_i)\eta_b(y_i - \bar{S}_i)y_i} \quad (2.18)$$

In the methods and results below, PWLS image reconstruction with conventional weights is denoted simply as PWLS, and PWLS image reconstruction with modified weights is denoted as PWLS\*.

## 2.2.6 Optimization approach

Because the objective function for PWLS\* is equivalent to the conventional PWLS formulation (with modification to the diagonal weights), PWLS\* can employ well established PWLS optimization algorithms. In this chapter, the separable quadratic surrogate with ordered subsets (OS-SQS) method<sup>163</sup> was used for solving PWLS\*, due in part to its suitability for parallelizable image updates (i.e., using parallel computation on GPU), with OS selected to further speed convergence. The OS-SQS algorithm used in this chapter for solving PWLS\* differs from the OS-SQS algorithm for solving penalized-likelihood reconstruction<sup>163</sup> on the following two points. First, instead of the log-likelihood data fit term for which the optimal curvatures change over the course of iterations, the optimal curvatures of the data fit term in PWLS\* are constants that can therefore be precomputed. Furthermore, the term corresponding to the optimal curvature of the data fit term ( $d_j$  in Table 2.1) in the update equation can also be precomputed; however, because this term is subset-dependent (and carries a large computation time to calculate for every subset), an approximation using all projections is precomputed (found to be a good approximation in the data reported below).

Table 2.1 presents pseudocode for solving PWLS\* using OS-SQS. The notation  $[\cdot]_+$  denotes the nonnegativity constraint,  $N_y$  and  $N_\mu$  are the size of the measurements and the image, respectively,  $\gamma_i$  is the  $i^{\text{th}}$  projection of an image of all ones,  $a_{ij}$  is the  $(i, j)$  element of the matrix  $\mathbf{A}$ ,  $n_{iter}$  is the maximum number of iterations,  $M$  is the number of subsets,  $\hat{l}_i$  is the projection of the

current image estimate  $\hat{\mu}$ , and  $S_m$  denotes all the projections in the  $m^{\text{th}}$  subset. In the regularization part,  $\Psi$  computes the first-order neighborhood difference,  $K$  is the number of neighboring voxels, and  $\dot{H}$  and  $\omega_H$  are the gradients and curvatures of the Huber penalty function  $H$ , respectively. The exact form for  $H$  is given in Sec. 1.4.8. The gradients  $\dot{H}$  are straightforward to compute and the curvatures  $\omega_H$  are defined as  $\omega_H(x) = \dot{H}(x)/x$ .

The computational complexity of PWLS\* (solved by OS-SQS) can be characterized by the number of projection operations needed (forward projections and backprojections), which is the dominant factor in computation time. Execution of PWLS\* as expressed in Table 2.1 requires  $2 \times n_{\text{iter}}$  projection operations on all the projections. The PWLS\* method was implemented in Matlab (The Mathworks, Natick MA), with the major part of the computation (i.e., projection and arithmetic operations) executed on GPU using CUDA-based libraries.

Table 2.1: Pseudocode for solving PWLS\* using OS-SQS.

---

Precompute optimal curvatures  $c_i(\hat{l}_i) = [\mathbf{W}_*]_i]_+ \quad \hat{l}_i \geq 0$

Precompute approximate  $d_j$  term  $d_j = \sum_{i=1}^{N_\mu} a_{ij} \gamma_i c_i(\hat{l}_i)$

**for** each iteration  $n = 1, \dots, n_{\text{iter}}$

**for** each subset  $m = 1, \dots, M$

$$\hat{l}_i = \sum_{j=1}^{N_\mu} a_{ij} \hat{\mu}_j \quad \dot{h}_i = w_i(l_i - \hat{l}_i) \quad \forall i \in S_m$$

**for**  $j = 1, \dots, N_\mu$

$$\dot{L}_j = M \sum_{i \in S_m} a_{ij} \dot{h}_i$$

$$\hat{\mu}_j^{\text{OS}} = \left[ \mu_j + \frac{\dot{L}_j - \beta \sum_{k=1}^K [\Psi]_{kj} \dot{H}([\Psi \hat{\mu}]_k)}{d_j + 2\beta \sum_{k=1}^K [\Psi]_{kj}^2 \omega_H([\Psi \hat{\mu}]_k)} \right]_+$$

**end**

**end**

**end**

---

Table 2.2: Summary of major parameters in the OS-SQS algorithm for PWLS\*.

Symbol	Description	Nominal Values or Range
$\Psi$	Sparsifying operator in image roughness penalty term	First-order neighborhood difference
$\beta$	Parameter controlling strength of image roughness penalty term	$10^1 \sim 10^3$
$\delta$	Size of quadratic region in Huber penalty function	5 HU
$n_{iter}$	Number of iterations	100
$M$	Number of subsets	20
$N_y$	Number of measurements	$668 \times 668 \times 720$
$N_\mu$	Number of voxels in reconstructed image	$412 \times 512 \times 512$

## 2.3 Experimental methods

### 2.3.1 Experiments on a CBCT test-bench and head phantom

Experiments were performed on a CBCT test-bench with a FPD (PaxScan 4343R, Varian, Palo Alto CA) as shown in Fig. 2.2(a). A custom anthropomorphic head phantom was scanned at 100 kVp, 0.4 mAs per projection with 720 projections ( $0.5^\circ$  angular steps), and a  $0.556 \times 0.556 \text{ mm}^2$  pixel size (after  $2 \times 2$  binning). The system used a 580 mm source-to-axis distance (SAD) and a 800 mm source-to-detector distance (SDD), resembling a typical configuration for compact head CBCT system.<sup>164</sup> As shown in Fig. 2.2(b-c), the head phantom was filled with a gelatin mixture carefully prepared to provide contrast equivalent to brain. The phantom was equipped with ventricle models prepared from wax with contrast equivalent to cerebrospinal fluid, and rows of plastic spheres with diameters ranging from 1.5 mm to 12 mm that were included to encompass a pertinent range of imaging tasks for the detection of intra-parenchymal hemorrhages. The resulting gelatin-plastic contrast closely simulated that of brain to fresh blood ( $\sim 50$  HU).<sup>142,165</sup>

The performance of the PWLS\* method was evaluated using a variety of CBCT scan conditions and correction methods. First, a narrow-beam scan was performed with longitudinal collimation reduced to a  $\sim 10$  mm field of view covering only the skull base region. The narrow-beam scan provided projection data with a low contribution of x-ray scatter. As such, these data were only corrected for beam hardening using the Joseph and Spital approach<sup>157</sup> mentioned above, referred to as the *beam-hardening-corrected* dataset. Second, a wide collimation scan was performed covering the entire cranium and incurring the substantial effects of both scatter and beam hardening. The projections were corrected for scatter using a high-fidelity Monte Carlo correction method<sup>149</sup> to form a *scatter-corrected* dataset, and then corrected for beam hardening to form a *fully-corrected* dataset. All projection data were offset-corrected and gain-normalized by mean dark and flood field calibrations and corrected for detector lag and veiling glare<sup>149</sup> prior to scatter and/or beam hardening corrections. Since the latter two effects are small in comparison to scatter and beam hardening, they were not explicitly considered in the current work.

For each dataset, images were reconstructed using three methods: FBP, PWLS, and PWLS\*. The same regularization parameters, number of iterations (100), and subsets (20) were used for both PWLS methods. Similarly, both PWLS methods used matched separable footprint projectors and backprojectors,<sup>98</sup> and FBP used voxel-driven interpolating backprojection. All images were reconstructed with  $0.5 \times 0.5 \times 0.5$  mm<sup>3</sup> voxel size on a workstation equipped with one Nvidia GeForce GTX TITAN graphics card. The main parameters in the OS-SQS algorithm for PWLS\* along with the nominal values or range used in the experiments are summarized in Table 2.2. The execution of OS-SQS over one subset in one iteration took about 10 seconds, and the total execution time over all 20 subsets and 100 iterations was about 5.5 hours. Since the current work mainly focused on evaluation of image quality for the proposed method (rather than computation speed), acceleration of the algorithm was not fully investigated in the current implementation and will be the subject of future work. For example, the projection/backprojection step was implemented using the separable footprint method. Replacing that with faster methods (such as a

combination of ray-driven Siddon<sup>117</sup> and voxel-driven interpolating method<sup>115</sup>) is expected to reduce time by a factor of  $\sim 5$  while achieving similar image quality, according to previous work by Wang *et al.*<sup>166</sup> Major reduction in execution time may also be achieved (beyond the scope of this chapter) by methods including acceleration techniques compatible with OS-SQS such as spatially non-uniform updates<sup>167</sup> and Nesterov's method.<sup>168</sup> For example, previous work<sup>166</sup> showed that Nesterov-accelerated penalized-likelihood reconstruction by OS-SQS with simplified forms of projection/backprojection reduced reconstruction time of C-arm CBCT data from  $\sim 100$  min to as little as  $\sim 2$  min.

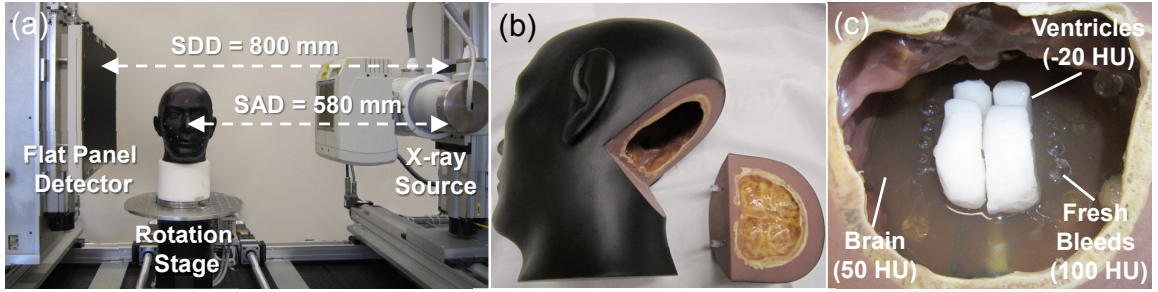


Figure 2.2: Experimental methods. (a) Benchtop CBCT system with the x-ray source and FPD in a geometry emulating a compact head scanner. (b) Anthropomorphic head phantom incorporating simulated intracranial hemorrhage. The occipital portion can be removed to access inside the interior of the cranium. (c) A photograph inside the head phantom during assembly, showing the gelatin mixture (brain), wax (ventricles), and plastic spheres (hemorrhage).

### 2.3.2 Evaluation of image quality

Both FBP and PWLS permit control of the tradeoffs between spatial resolution and noise through adjustment of algorithm parameters: for FBP, via the cutoff frequency of the apodization (smoothing) kernel; for PWLS, via the regularization parameter,  $\beta$ , and the Huber distance,  $\delta$ . For fair comparison, images reconstructed by different methods were matched in terms of either the spatial resolution or the image noise by adjusting the apodization kernel in FBP and the regularization parameter  $\beta$  in PWLS and PWLS\*. The Huber distance  $\delta$  was fixed and chosen to be the same between PWLS and PWLS\*, and the selection of its value is discussed in Sec. 2.4.2.

Spatial resolution was assessed using the method detailed in the paper by Wang *et al.*<sup>93</sup> in terms of the edge spread function (ESF) of a sphere of contrast equal to the structure of interest (i.e., fresh blood). The ESF was fit to a sigmoid parameterized by width,  $\varepsilon$ , as shown in Fig. 2.3(b). The ESF width was computed using all voxels within 60° fans centered on the sphere, and spatial resolution was characterized as the average  $\varepsilon$  computed over all the fans. The contrast of the sphere was given by a second fit parameter  $c$  ( $\text{mm}^{-1}$ ) and was also averaged over all the fans. Image noise was calculated as the standard deviation of voxel values in a region-of-interest (ROI) ( $19 \times 19$  voxels) in a homogeneous region of gelatin (brain) immediately adjacent to the sphere. Contrast-to-noise ratio (CNR) was computed using the contrast of the sphere and image noise mentioned above.

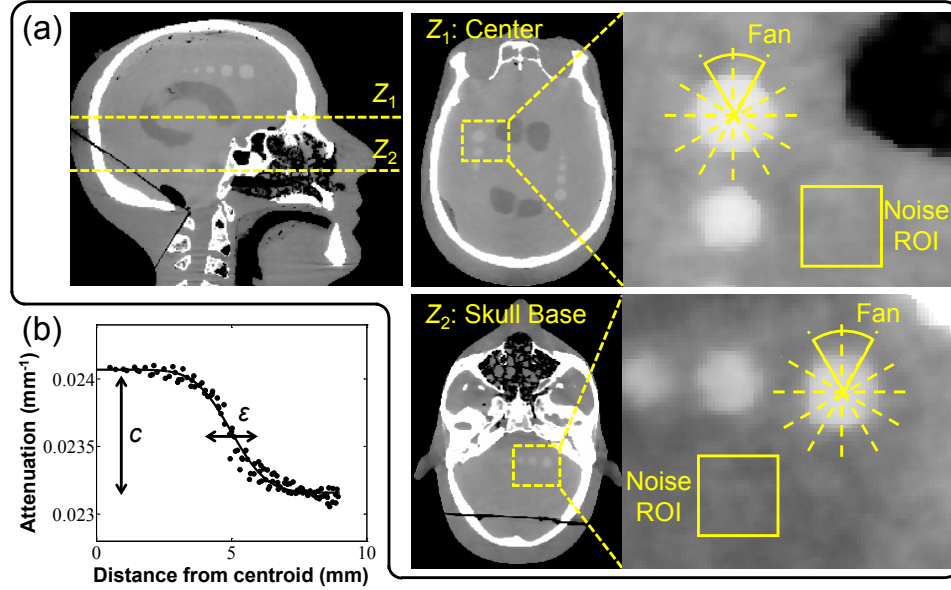


Figure 2.3: Illustration of image quality metrology. (a) Two axial slices in which the spatial resolution, noise, and contrast-to-noise ratio (CNR) were computed. The spatial resolution and contrast were computed by fitting an edge spread function (ESF) to voxel values in 60° fans centered on a sphere of contrast equivalent to blood. Noise was computed as the standard deviation of voxel values in a homogeneous ROI immediately adjacent to the sphere. (b) Example sigmoid fit to the measured ESF, from which spatial resolution was characterized in terms of the ESF width,  $\varepsilon$ , and contrast was given by the parameter  $c$ .

Figure 2.3(a) shows two axial slices ( $Z_1$  and  $Z_2$ ) in which the image quality metrics were evaluated. For the beam-hardening-corrected dataset, the metrics were computed in axial slices about the  $Z_2$  region near the skull base. For the scatter-corrected dataset and fully-corrected dataset,

the metrics were computed in axial slices about the  $Z_1$  region of central brain parenchyma. Note that for fair comparison, the spatial resolution in the longitudinal ( $z$ ) direction in FBP reconstructions was also matched to that in PWLS reconstructions by additional apodization in the  $z$  direction in the FBP smoothing kernel as in the paper by Wang *et al.*<sup>93</sup> Conversion of voxel values to HU assumed a constant value of water attenuation ( $0.0219 \text{ mm}^{-1}$ ).

## 2.4 Results

### 2.4.1 Effect of artifact corrections on variance

The change in variance associated with each step of the artifact correction was computed using the test-bench data as shown in Fig. 2.4. An increase in variance throughout the head was observed in every step of the artifact correction. For the beam hardening correction, such changes are seen to vary spatially throughout the head for the water ( $\eta_w$ ) correction (Fig. 2.4(a)) and was more highly concentrated in the bony regions in bone ( $\eta_b$ ) correction (Fig. 2.4(b)), consistent with the particular correction applied in each case. The combined beam hardening correction ( $\eta_w \eta_b$ ) gives a maximum change in variance by  $\sim 30\%$  (Fig. 2.4(c)). For scatter correction, the change in variance also depended strongly on location (Fig. 2.4(d)) with a maximum change as high as two orders of magnitude in dense bone regions, where scatter fractions are highest. The total variance change after applying all corrections (Fig. 2.4(e)) was dominated mainly by the scatter correction part, indicating scatter correction to be the dominant factor affecting the variance in this case. The substantial changes in variance throughout the head suggests the importance of the proposed modified weighting terms to compensate for such variations.

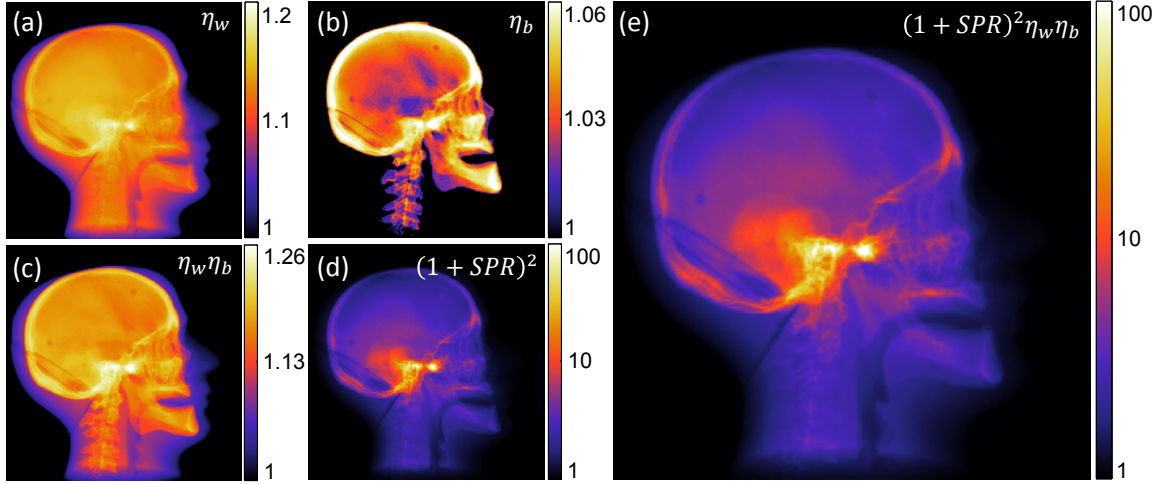


Figure 2.4: Example illustration (in a lateral projection view) of the change in variance associated with each step of artifact correction. The change in variance is shown for corrections associated with: (a) beam hardening correction for water ( $\eta_w$ ); (b) beam hardening correction for bone ( $\eta_b$ ); (c) total beam hardening correction ( $\eta_w \eta_b$ ); (d) scatter correction  $(1+SPR)^2$ ; and (e) all scatter and beam hardening corrections.

### 2.4.2 Selection of image regularization parameters

Fair comparison of image quality for different reconstruction methods requires a justification of parameter settings suitable to the imaging task as well as evaluation of noise and spatial resolution on equal footing. The PWLS\* and PWLS methods used a Huber penalty function to reduce noise in a manner governed by the range of the quadratic neighborhood  $\delta$ . Previous work showed that a small value of  $\delta$  (relative to the contrast of structures of interest) should be selected to provide edge preservation and noise reduction,<sup>93,111</sup> but selection of  $\delta$  too small tends to over-regularize the image and lead to unrealistic, piecewise-constant images. To find a reasonable value of  $\delta$ , the PWLS\* images were reconstructed using a fully-corrected dataset over a range of  $\delta$  values (1, 3, 5, and 10 HU). A region-of-interest from images resulting from PWLS\* reconstructions at various values of  $\delta$  are shown in Fig. 2.5(a) with spatial resolution matched at a level of  $\varepsilon = 0.65$  mm. The image with  $\delta = 10$  HU exhibits a high degree of noise with little benefit from the linear region of the Huber penalty, and noise reduces for smaller values of  $\delta$ . However, reduction of  $\delta$  to 1 HU leads to an

unrealistic patchy texture and decreased contrast, suggesting over-regularization. Figures 2.5(b-c) summarize the performance in terms of the noise and CNR over a broad range of spatial resolution for different  $\delta$  values. Consistent with the images in Fig. 2.5(a), the quantitative analysis shows that reducing  $\delta$  from 10 HU to 3 HU reduces the noise and increases the CNR at any level of spatial resolution, whereas at  $\delta = 1$  HU the CNR exhibits a ‘kink’ in the noise-resolution tradeoff, suggesting over-regularization and contrast degradation. Together, these results suggest a reasonable value of  $\delta$  in the range  $\sim 3$ -5 HU. Considering the narrow range of suitable values (3~5 HU) for  $\delta$ , its value was fixed for both PWLS and PWLS\* at a level that avoided under-regularization or over-regularization. In the results reported below,  $\delta = 5$  HU was selected as a conservative choice balancing noise reduction and edge preservation for both the PWLS\* and PWLS methods and all three datasets.

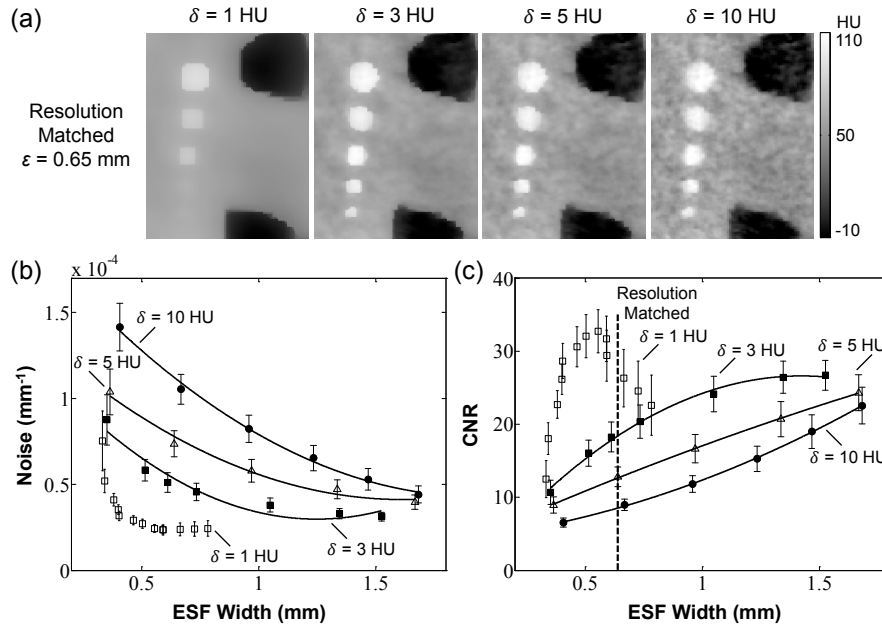


Figure 2.5: Selection of regularization parameters with respect to noise-resolution tradeoffs. (a) PWLS\* images in a ROI in the  $Z_1$  axial slice using different  $\delta$  values. Images were reconstructed using the fully-corrected dataset with spatial resolution matched at  $\varepsilon = 0.65$  mm. (b-c) Noise-resolution tradeoff and CNR-resolution tradeoff for PWLS\* using different  $\delta$  values. The error bars are based on the standard deviation of the noise in 6 neighboring axial slices. The solid curves are second-order polynomial fits intended merely as a guide to trends evident in the underlying data points. A value of  $\delta = 5$  HU was selected as a conservative choice giving a reasonable level of noise reduction and edge preservation without introducing an unnatural appearance of patchy over-regularization.

### 2.4.3 Image reconstruction with beam hardening correction

The narrow-beam collimation case strongly reduced x-ray scatter effects (SPR less than  $\sim 10\%$ ) and provided data that largely isolated beam hardening effects. FBP, PWLS, and PWLS\* reconstructions were performed following water and bone beam hardening correction to quantify and visualize the effects of the correction algorithm on image variance and the potential improvements in image quality with modified PWLS weights. Figure 2.6 shows reconstructions for each case within a ROI in the axial slice  $Z_2$ , where beam hardening effects were severe due to the presence of thick, dense bone in the skull base. FBP reconstruction without beam hardening correction (Fig. 2.6(a)) exhibits fairly severe artifacts and a nominal level of quantum noise  $\sigma = 9.3$  HU, whereas the same data with beam hardening correction (denoted FBP<sup>B</sup> in Fig. 2.6(b)) shows a substantial reduction in artifacts but a  $\sim 13\%$  increase in noise. PWLS reconstruction with conventional weights provides a strong ( $\sim 43\%$ ) reduction in noise (at matched spatial resolution) as shown in Fig. 2.6(c). A further ( $\sim 12\%$ ) reduction in noise is obtained using modified weights as shown in Fig. 2.6(d), demonstrating the benefits of the proposed method in preserving the fidelity of artifact corrections and reducing noise by nearly a factor of 2 compared to FBP<sup>B</sup>.

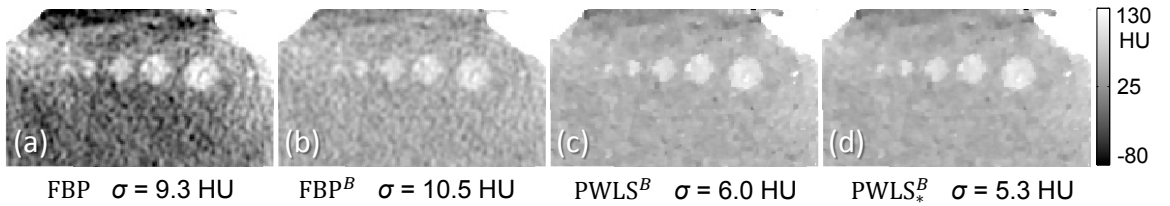


Figure 2.6: Reconstructions with beam-hardening correction. (a) FBP reconstruction with no corrections. (b) FBP reconstruction of beam-hardening-corrected data. (c) PWLS reconstruction of beam-hardening-corrected data with conventional weights. (d) PWLS\* reconstruction of beam-hardening-corrected data with modified weights. The proposed method (PWLS<sub>\*</sub>) maintains the quality of artifact correction and gives a factor of  $\sim 2$  improvement in noise compared to FBP<sup>B</sup>. (Resolution was matched at  $\varepsilon = 0.70$  mm for fair comparison.)

#### 2.4.4 Image reconstruction with scatter correction

Full field-of-view (wide-collimation) projection data were reconstructed using FBP, PWLS, and PWLS\* following scatter correction by a Monte Carlo method.<sup>149</sup> Figure 2.7 shows reconstructions for each case in a ROI in the  $Z_1$  axial slice at matched spatial resolution ( $\varepsilon = 0.50$  mm, top row) and at matched contrast-to-noise ratio (CNR = 13, bottom row). In Fig. 2.7(a) (matched spatial resolution images), FBP reconstruction without scatter correction exhibits severe cupping artifacts, reduced contrast, and inaccurate HU (but relatively low noise). FBP reconstruction of the scatter-corrected data (denoted FBP<sup>S</sup>) shows a strong reduction in artifacts but an increase in noise by ~72%. This dramatic increase in noise is consistent with the large change in variance associated with scatter correction shown in Fig. 2.4. PWLS reconstruction of the scatter-corrected data (using conventional weights, denoted PWLS<sup>S</sup>) reduced the noise by ~41%, and use of the modified weights further reduced noise by an additional ~26%, denoted PWLS<sup>S</sup>\*. In Fig. 2.7(b) (matched CNR images), complementary improvement in spatial resolution is seen for the proposed method, with the ESF width improving from  $\varepsilon = 1.44$  mm for FBP<sup>S</sup> to 1.25 mm for PWLS<sup>S</sup> with conventional weights to and to 0.85 mm for PWLS<sup>S</sup>\* with modified weights.

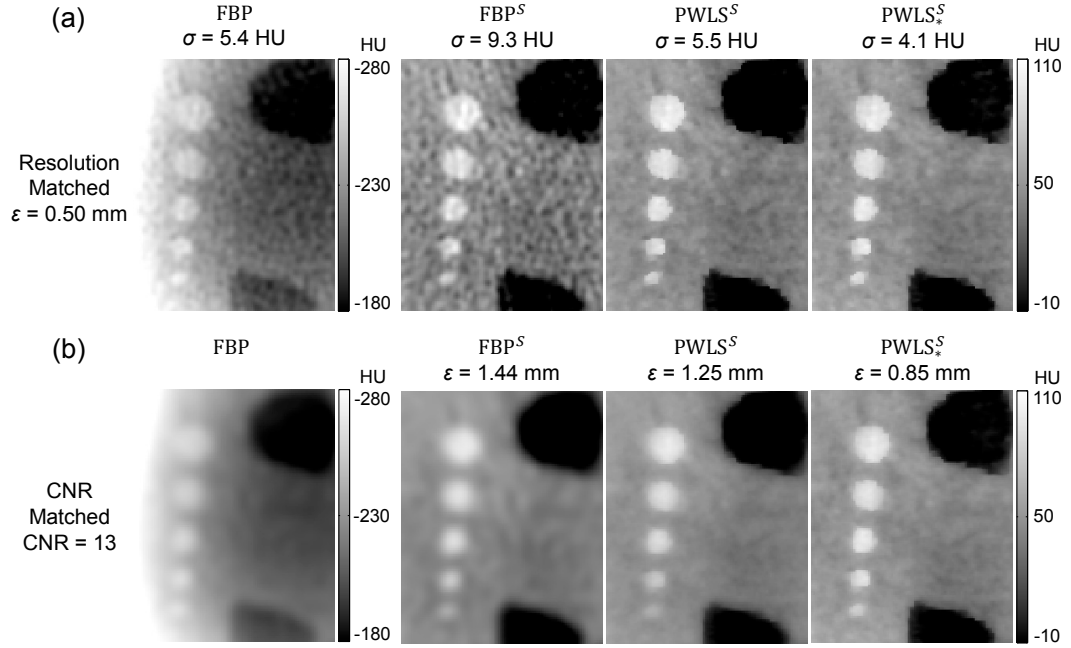


Figure 2.7: Reconstructions with scatter correction - top row at matched spatial resolution and bottom row at matched CNR. From left to right: FBP reconstruction of uncorrected and scatter-corrected projection data; PWLS of scatter-corrected data with conventional weights; and PWLS\* reconstructions of scatter-corrected data with modified weights.

Figure 2.8 quantifies the noise-resolution tradeoffs among the three reconstruction methods with scatter correction. As expected, both the PWLS and PWLS\* methods show improved noise (and CNR) compared to FBP at any level of spatial resolution. The proposed PWLS\* method shows a further improvement in noise and CNR compared to PWLS with conventional weights - for example, a 34% improvement in CNR at a spatial resolution of  $\epsilon = 0.50$  mm.

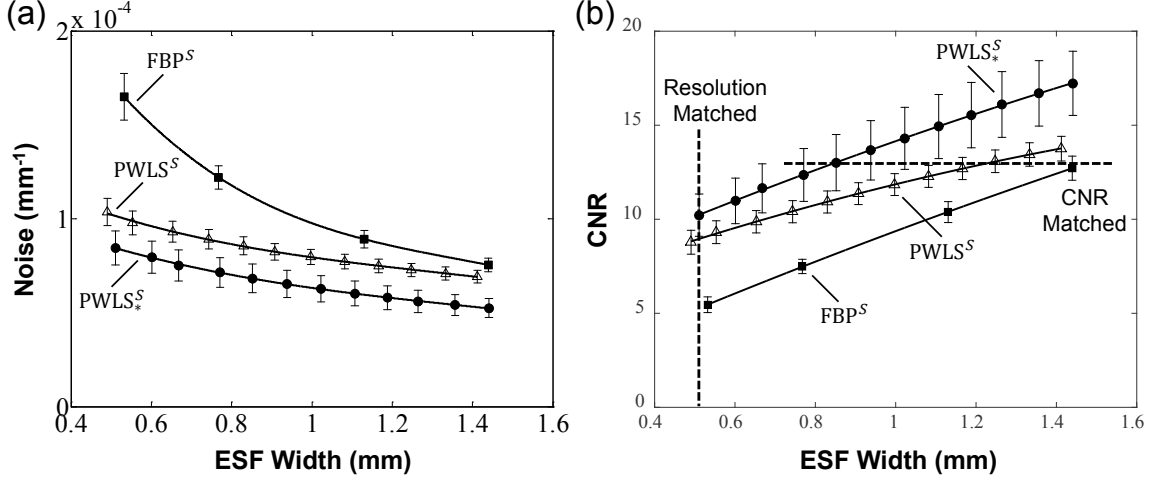


Figure 2.8: Noise-resolution tradeoffs for the FBP, PWLS, and PWLS\* reconstruction methods operating on scatter-corrected projection data. The levels of matched spatial resolution and CNR used in Fig. 2.7 are marked by the dashed lines in (b).

#### 2.4.5 Image reconstruction with scatter and beam hardening corrections

Finally, the combined effects of scatter and beam hardening corrections were evaluated in FBP, PWLS, and PWLS\* reconstructions. Similar to Fig. 2.7, the results in Fig. 2.9 show images from full field (wide collimation) scans in a ROI in the  $Z_1$  axial slice. Reconstruction of the fully corrected (scatter and beam hardening) projection data are denoted by the superscript  $SB$ . Considering the top row of Fig. 2.9 (images at matched spatial resolution), a progressive reduction in noise is seen for the FBP<sup>SB</sup>, PWLS<sup>SB</sup>, and PWLS\*<sup>SB</sup> reconstructions, with the last providing a ~40% reduction in noise compared to the first. Considering the bottom row of Fig. 2.9 (images at matched CNR), the PWLS\* method also exhibits the highest spatial resolution ( $\varepsilon = 0.50$  mm). Moreover, the PWLS\* images show reliable detection of simulated intracranial hemorrhages as small as 3 mm diameter. The corresponding noise-resolution tradeoffs among the three methods are quantified in Fig. 2.10, with PWLS\* demonstrating an improvement in each respect. Note also the increased CNR of all three methods in Fig. 2.10 compared to Fig. 2.8, owing to the additional correction (beam hardening).

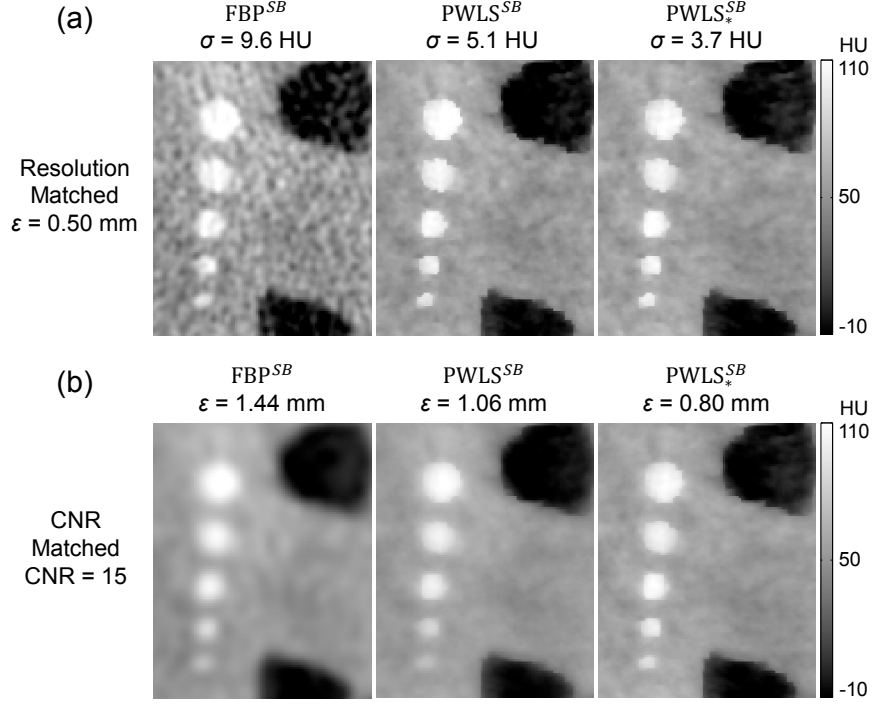


Figure 2.9: Reconstructions with scatter and beam hardening corrections - top row at matched spatial resolution and bottom row at matched CNR. As in Fig. 2.7, from left to right: FBP, PWLS, and PWLS\* reconstructions of fully corrected projection data.

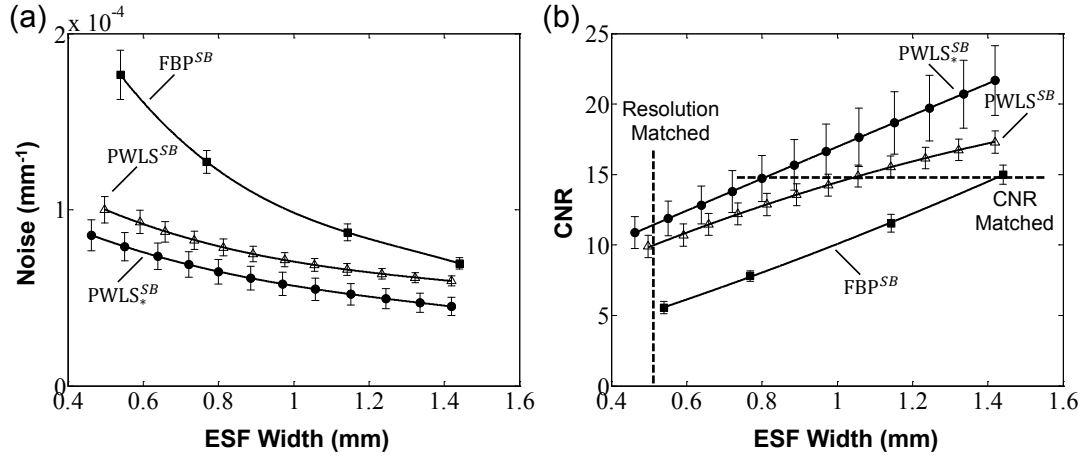


Figure 2.10: Noise-resolution tradeoffs for the FBP, PWLS, and PWLS\* reconstruction methods operating on fully-corrected (scatter and beam hardening) projection data. The levels of matched spatial resolution and CNR used in Fig. 2.9 are marked by the dashed lines in (b).

Visual assessment is further illustrated in Fig. 2.11, showing FBP, PWLS, and PWLS\* reconstructions of the fully-corrected data in regions about the skull base, where scatter and beam hardening effects are strongest. Axial, coronal, and sagittal slices are shown with spatial resolution

matched at  $\varepsilon = 0.40$  mm. Uncorrected images (not shown) are severely degraded by artifact and are not suitable for detection of ICH in this region. The fully corrected FBP<sup>SB</sup> reconstruction yields strong reduction of artifact but exhibits a high level of noise, particularly in the region between the temporal bones. Interestingly, the PWLS<sup>SB</sup> reconstruction with conventional weights does not provide a reduction in noise due to gross mismatch between the presumed noise model with the fully corrected data. The proposed PWLS\* reconstruction with modified weights substantially reduced the noise at a fixed level of spatial resolution, which may improve detection of low contrast intracranial hemorrhage. These results in a challenging region demonstrated the importance of incorporating an accurate noise model in MBIR. The bias in the PWLS<sub>\*</sub><sup>SB</sup> reconstruction surrounding the spheres was most likely caused by under or over correction of image artifacts in this challenging region. Such bias may be reduced by methods including more accurate bone segmentation in beam hardening correction and performing the “two-step” beam hardening correction multiple times.

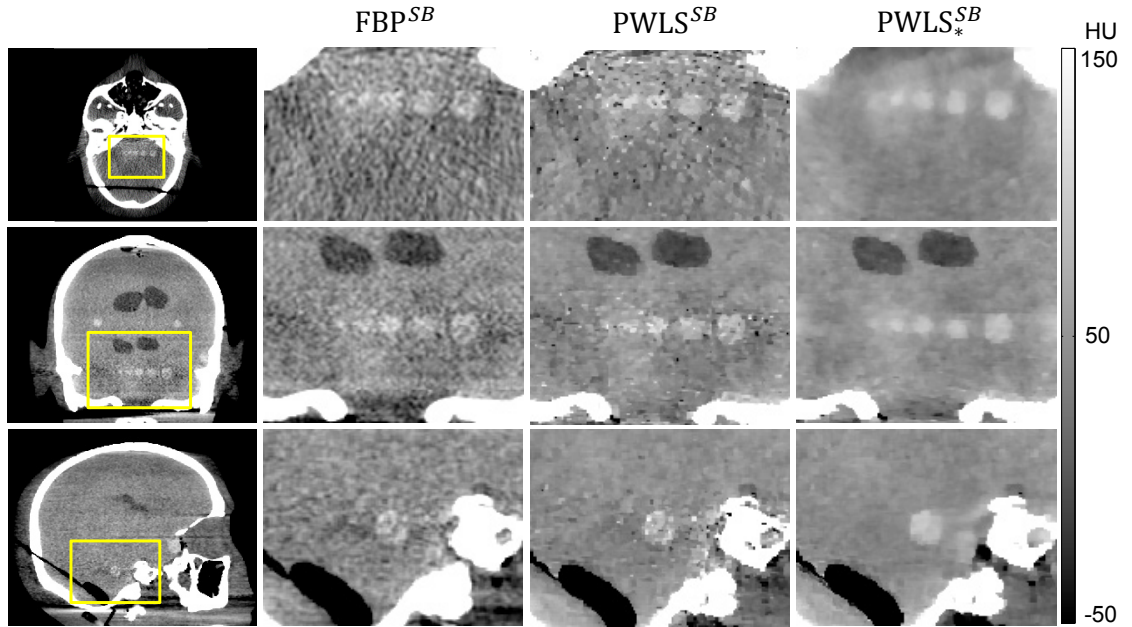


Figure 2.11: Axial (top row), coronal (middle row), and sagittal (bottom row) CBCT images of the head phantom with scatter and beam hardening corrections, reconstructed by FBP, PWLS, and PWLS\*. The skull base presents a challenging region for which proper account of scatter and beam hardening corrections is essential to high-quality reconstruction. The spatial resolution in each case was matched ( $\varepsilon = 0.40$  mm) at the largest sphere in the axial slice  $Z_1$ .

## 2.5. Conclusions and discussion

The chapter has reported a novel PWLS image reconstruction method that incorporates a general framework for accurately accommodating modified noise models for artifact-corrected CBCT data. A specific scenario was considered where the two dominant artifact correction methods essential to high-quality CBCT - scatter and beam hardening corrections – are applied. The resulting reconstruction method (denoted PWLS\*) utilizes modified weights to compensate for noise amplification imparted by each step of the artifact correction. Experiments included physical data acquired on a FPD-CBCT test-bench using an anthropomorphic head phantom emulating intraparenchymal hemorrhage. A conservative level of regularization in the Huber penalty was selected to improve the conspicuity of simulated hemorrhages while not resulting in an unnaturally patchy, over-regularized image. The proposed PWLS\* method demonstrated superior noise-resolution tradeoffs in comparison to traditional methods, including FBP and PWLS with conventional statistical weights (and noise models).

The work suggests a number of interesting points meriting future investigation. The first is that the current model of variance change focuses on Poisson noise propagated through each correction. While this is a valid starting point, assuming accurate scatter and beam hardening estimates and gain normalization, the model could be extended to consider other sources of noise, including error in the Monte Carlo scatter estimate, uncertainties in the segmentation of bony structures in the beam hardening correction, and incomplete correction of detector gain variations. For example, in the Monte Carlo estimation of the scatter fluence, noise could be introduced in steps that require random sampling from a known distribution (e.g., photon energy, direction, path length before interaction, etc.), and the effects could be exaggerated when only a small number of photons are simulated.<sup>82</sup> Incorporating such additional sources of variation is the subject of future work.

A second area for future study is to account for other sources of image artifacts in FPD-CBCT, such as the effects of image lag and low-frequency glare / off-focal radiation. These effects are included within the comprehensive artifact correction framework of Sisniega *et al.*,<sup>149</sup> but such effects were not included in the current work. Since lag and veiling glare corrections also introduce variance changes (although measurably small compared to scatter and beam hardening effects), the noise changes associated with these corrections could also be included in further modification of the PWLS weights - for example, derived from the known relationship between noise and image lag<sup>68,169</sup> and low-frequency blur. Note that the assumption of independent measurements no longer holds in the case of lag and veiling glare corrections, due to the existence of temporal (lag) or spatial (veiling glare) correlations between measurements. Therefore, Eq. (2.2) and (2.3) need to be re-written for consideration of correlations when dealing with these corrections, and a generalized reconstruction model as in the work by Stayman *et al.*<sup>170</sup> and by Tilley *et al.*<sup>171</sup> may be required.

A third topic for additional investigation is the recognition and challenge associated with spatially varying spatial resolution effects in MBIR<sup>132</sup> including the PWLS\* method detailed in this chapter, which could potentially degrade detectability of ICH. For example, in a PWLS\* image with a fixed regularization parameter,  $\beta$ , and for a given size and contrast of sphere (simulated hemorrhage), lower spatial resolution characteristics were observed in the central region of the head, while higher spatial resolution was evident adjacent to the cranium. This problem could be at least partly addressed by methods that encourage space-invariant spatial resolution, such as the certainty-based method.<sup>132</sup> Preliminary results showed that the use of certainty-based method achieved a more uniform spatial resolution across the image, thereby enabling similar noise-resolution tradeoffs between the simulated hemorrhages in different regions of the head. Moreover, intentional use of spatially varying regularization may be beneficial to the imaging task(s) - for example, sharper resolution (via reduced penalty strength) in regions of the cranium for detection of fracture, simultaneous with stronger noise reduction (via higher penalty strength) within the brain parenchyma for detection of intracranial hemorrhage. Among the possible methods for

achieving such a spatially varying noise-resolution characteristic is a spatially varying  $\beta$  map<sup>136</sup> optimized for CBCT imaging of the head. This topic is pursued in depth in Chapter 3, involving a spatially varying penalty designed to maximize task-based imaging performance.

Another point worth further investigation is the assessment of image quality with respect to specific clinical tasks. In this chapter, simple imaging performance metrics were used to quantify spatial resolution, contrast, noise, and CNR. These metrics provide useful preliminary insight on the improvements obtained with the proposed reconstruction method, and future work will include assessment with respect to task-based measures such as detectability index, various observer models, and human observer studies. For example, detection of a low-contrast, low-frequency lesion (analogous to the clinical task of intracranial hemorrhage detection) and/or a higher-contrast, high-frequency abnormality (analogous to the clinical task of fracture detection) can be analyzed via statistical decision theory hypothesis testing in terms of task-based detectability index and various model observers.<sup>62,136,172–175</sup> Application of such methods in a manner that accounts for the complexities of nonlinear reconstruction methods is an area of active research, and such work will necessarily account not only for the clinical task but also complexities associated with nonlinear, nonstationary characteristics of the 3D image reconstruction. The results reported in this chapter suggest imaging performance consistent with the clinical task of visualizing a 3 mm intracranial hemorrhage (e.g., CNR = 11.9 in images with ESF width = 0.50 mm, also evident in Fig. 2.9). This may be sufficient for detection of a broad spectrum of acute injuries (subject to validation in future clinical studies) and suggest the possibility for further improvement allowing detection of mm-scale micro-hemorrhage in diffuse axonal injury after concussion.<sup>176</sup>

Finally, certain application scenarios present an opportunity to leverage a patient-specific head image from a previous scan to improve image quality and/or reduce dose in a subsequent scan. Examples include longitudinal monitoring of brain hemorrhage in the intensive care unit (ICU) (where acquisition of multiple head CT scans over the course of ICU monitoring is common) and populations at high risk of head injury in sports and military theatres. Such patient-specific prior

images can be incorporated into the PWLS\* reconstruction in the form of prior image regularization<sup>111,177</sup> to maximize the conspicuity of low-contrast hemorrhages and increase the sensitivity to subtle anatomical changes. The methods<sup>111,177</sup> also jointly register the patient-specific prior image to the current anatomy in the course of image reconstruction so that the corresponding anatomical structures are well aligned for correct prior image regularization. The joint registration might be performed in a rigid fashion (which is often a good approximation of the motion of the head) or in a deformable fashion<sup>177</sup> if soft tissue deformation is present in the head (e.g., midline shift, ventricular compression). This topic is pursued in depth in Chapters 5-6, where novel registration and regularization methods are incorporated in the Prior Image Registration, Penalized Likelihood Estimation (PIRPLE) method.

## Chapter 3

### 3. Task-Based Image Reconstruction

#### 3.1 Introduction

##### 3.1.1 Motivation

A medical image is always produced to accomplish a particular clinical task (or tasks), which may be generally or specifically defined. For example, x-ray screening mammography images are obtained to detect suspicious lesions (e.g., masses or microcalcifications) that may be indicative of breast cancer. Similarly, CT angiography of the head is performed to help detect and diagnose disruptions of the neurovasculature, as in ischemic or hemorrhagic stroke. In image-guided interventions, fluoroscopy or intraoperative CBCT are acquired often with the purpose to localize the position of the surgical target and/or adjacent healthy tissue with respect to interventional tools. Thus, the performance of the imaging system is most meaningfully described with respect to the intended task, as noted by Barrett in 1990<sup>178</sup>:

*“...A scientific or medical image is always produced for some specific purpose or task, and the only meaningful measure of its quality is how well it fulfills that purpose. An objective approach to assessment of image quality must therefore start with a specification of the task and then determine quantitatively how well the task is performed.”*

Task-based assessment of medical imaging performance has provided a basis for the design and optimization of a variety of medical imaging systems, including mammography,<sup>74</sup> tomosynthesis,<sup>179</sup> multi-detector CT,<sup>180,181</sup> CBCT in breast, musculoskeletal, and head,<sup>73,182–184</sup> and image-guided interventions.<sup>185</sup> However, less efforts have been made on the incorporation of task-

based assessment of imaging performance into the process of CT image reconstruction. Statistical image reconstruction in CT often involves the use of regularization techniques to enforce desired image quality properties in the reconstructed image. However, regularization is often formulated in general terms to encourage smoothness and/or sharpness (e.g., a linear, quadratic, or Huber penalty) without explicit formulation of the task. This chapter proposes a statistical reconstruction method that incorporates a model for task-based imaging performance into the selection of regularization parameters to maximize task-based image quality. Performance is tested and evaluated in the context of CBCT of the head.

As described in Chapter 2, point-of-care imaging of acute ICH can improve the diagnosis and monitoring of acute ICH and reduce the risk associated with patient transport from the critical care unit to the CT scanner suite. CBCT is potentially well suited to point-of-care imaging, but its imaging performance tends to be challenged for low-contrast, soft-tissue visualization, as in the detection of acute ICH. A prototype head CBCT scanner was recently developed, using a model for task-based imaging performance as a guide to system design.<sup>183,184</sup> An artifact correction framework was also developed to mitigate artifacts arising from x-ray scatter, beam hardening, detector lag, and veiling glare in CBCT of the head.<sup>149</sup> The previous chapter described a PWLS reconstruction algorithm with statistical weights modified to account for the change in variance following artifact corrections in CBCT of the head. The results suggest the possibility for point-of-care imaging of acute ICH using CBCT, but also introduce new image quality challenges and opportunities.

First, task-based assessment of imaging performance was not explicitly formulated in the PWLS reconstruction method described in the previous chapter but can be potentially incorporated into and benefit the image reconstruction process. The primary task in the diagnosis and monitoring of acute ICH is to discriminate a low-contrast lesion (acute ICH) from a relatively uniform background (brain). Early development of the CBCT prototype mentioned above used such task-based analysis to optimize the imaging configuration and technique factors with respect to the

detectability index ( $d'$ ).<sup>186</sup> As shown below, such analysis can similarly benefit the process of 3D image reconstruction, taking task-based imaging performance as the objective function for optimization in statistical MBIR. Specifically, the flexibility of statistical reconstruction can be leveraged to derive a penalty that maximizes ICH detectability.

A second consideration is that images reconstructed by statistical reconstruction typically exhibit nonuniform spatial resolution and noise characteristics. For example, Fig. 3.1 shows an image of an anthropomorphic head phantom reconstructed using a basic PWLS method with a simple (constant, spatially uniform) quadratic penalty function. The local spatial resolution (e.g., the width of the edge spread function, ESF, denoted  $\varepsilon$  in Fig. 3.1) and noise (e.g., the standard deviation in voxel values, denoted  $\sigma$  in Fig. 3.1) and the conspicuity of the simulated ICH vary in different regions of the head. Such non-uniformity carries a variety of implications. First, analysis of noise, resolution, etc. must recognize assumptions and limitations of a local approximation – e.g., as in the work by Tward and Siewerdsen<sup>62</sup> and by Pineda *et al.*,<sup>187</sup> and the effect of such non-uniformity on human observer performance is an area of ongoing investigation. Furthermore, and most pertinent to the work reported below, a penalty designed to maximize task-based detectability at one location in the head will not necessarily maximize detectability at another location.

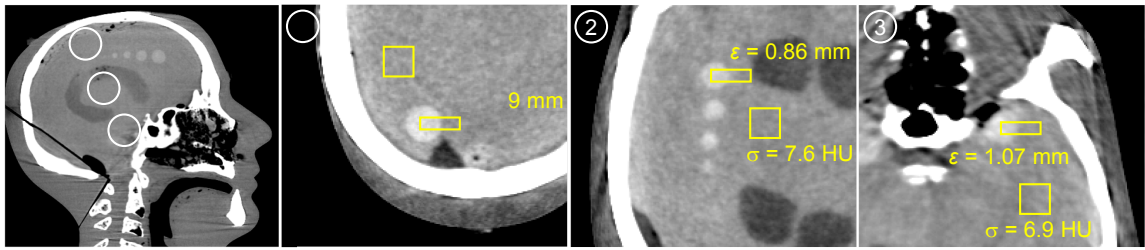


Figure 3.1: Illustration of nonuniform spatial resolution (ESF width,  $\varepsilon$ ) and noise (standard deviation,  $\sigma$ ) at three regions in an anthropomorphic head phantom (1 - near the cranium, 2 - at the center of the cranial vault, and 3 - near the skull base) in an image reconstructed by PWLS with a quadratic penalty. The phantom was scanned on a CBCT test-bench at a dose of 24 mGy. Grayscale window for axial images 1-3 is [-60 160] HU.

Such considerations motivated the work detailed below to derive a spatially varying penalty that enforces constant  $d'$  throughout the image – or more importantly, to maximize local

$d'$  at every point in the image. A number of previous reports also investigate the potential benefit of a spatially varying penalty. Fessler and Rogers<sup>132</sup> proposed a penalty containing a spatially varying “certainty” term to achieve uniform spatial resolution – specifically, uniform point spread function (PSF) in PL reconstruction. Yang *et al.*<sup>135</sup> optimized directional weights in the penalty for PL reconstruction to maximize task-based detectability for the detection of breast lesion in 3D PET. Gang *et al.*<sup>136</sup> optimized a parameter that weights the regularization term in PL reconstruction to maximize detectability for a variety of detection tasks in axial CT.

The work in this chapter extends task-based regularization design to CBCT and focuses on improving task-based imaging performance for the detection of acute ICH in CBCT of the head. This chapter is structured as follows. First, the theoretical prediction framework for spatial resolution (local PSF and modulation transfer function, MTF) and noise (local covariance and noise-power spectrum, NPS) in PWLS reconstruction is described and validated in simulation studies involving a realistic 3D digital head phantom. Second, a task-based regularization design framework is introduced along with techniques to accelerate the design process in 3D CBCT. Finally, the imaging performance of the proposed task-based penalty is evaluated in both simulation studies and test-bench experiments in comparison to a conventional (constant, space-invariant) penalty and a spatially-varying “certainty-based” penalty encouraging uniform spatial resolution. The proposed regularization was investigated for both a mid-frequency task emulating detection of a low-contrast ICH as well as tasks emphasizing other frequency content (low- and high-frequency).

### **3.1.2 Acknowledgements and unique contributions**

The methods and results reported in this chapter were reported in conference proceedings and journal articles as follows:

- (1) H. Dang, J. W. Stayman, J. Xu, A. Sisniega, W. Zbijewski, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, J. H. Siewerdsen, "Regularization design for high-quality cone-beam CT of intracranial hemorrhage using statistical reconstruction," *SPIE Medical Imaging*, San Diego, CA, Vol. 9783, 97832Y (2016).
- (2) H. Dang, J. W. Stayman, J. Xu, A. Sisniega, W. Zbijewski, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, J. H. Siewerdsen, "Task-based regularization design for detection of intracranial hemorrhage in cone-beam CT," *The 4th International Conference on Image Formation in X-Ray Computed Tomography*, Bamberg, Germany, 557-560 (2016).
- (3) H. Dang, J. W. Stayman, J. Xu, A. Sisniega, W. Zbijewski, M. Mow, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, and J. H. Siewerdsen, "Task-based statistical reconstruction for high-quality cone-beam CT," *Physics in Medicine and Biology*, submitted (2017).

with permission from the publishers for reproduction of content in this dissertation. The author's primary contributions in this work include: development of a task-based regularization design framework; efficient implementation of the reported framework in CBCT; simulation and physical experimentation; and quantitative analysis. The author would like to gratefully acknowledge the contributions of coauthors, including: Dr. Jennifer Xu for providing initial code on modeling of task-based imaging performance; Dr. Alejandro Sisniega for assisting physical experiments and artifact correction; and Dr. Wojciech Zbijewski for providing helpful suggestions through the course of this project. Valuable discussion with Dr. Grace J. Gang on task-based regularization design is also gratefully acknowledged. The work was supported by research collaboration with Carestream Health (Rochester NY).

## 3.2 Methods

### 3.2.1 Statistical reconstruction for head imaging

This chapter considers the same forward model and PWLS objective function as in Chapter 2. The PWLS objective function is:

$$\hat{\mu} = \arg \min_{\mu} \frac{1}{2} \|\mathbf{A}\mu - l\|_{\mathbf{W}}^2 + R(\mu) \quad (3.1)$$

as defined previously. Recall that  $\mathbf{W}$  is a diagonal weighting matrix whose diagonal elements are statistical weights computed as the inverse of the variance of the measurements propagated through data transformation. In CBCT of the head, the statistical weights are modified as described in the previous chapter to account for the changes in variance associated with x-ray scatter and beam-hardening corrections.

The regularization term  $R(\mu)$  in the PWLS objective penalizes intensity differences between neighboring voxels and enforces local smoothness in the reconstructed image. A commonly used form for the regularization was described in Chapter 1 as:

$$R_C(\mu) = \frac{1}{2} \beta \sum_j \sum_{k \in N_j} w_{jk} \varphi(\mu_j - \mu_k) \quad (3.2)$$

where the difference between voxel  $\mu_j$  and its neighboring voxel  $\mu_k$  is penalized by a function  $\varphi$  and weighted by a directional weight  $w_{jk}$ . Conventionally, the penalty function and directional weights are chosen to be the same for all the voxels in the image, and a scalar (constant, spatially uniform) regularization parameter  $\beta$  is used to control the overall strength of regularization. This regularization method is referred to as the “conventional penalty” (denoted by subscript  $C$ ) in this chapter. Such regularization typically results in nonuniform spatial resolution and noise characteristics in the reconstructed image.

An alternative regularization method proposed by Fessler and Rogers<sup>132</sup> seeks to achieve uniform spatial resolution (specifically, uniform PSF) throughout the image. This method introduces a spatially varying quantity  $\kappa_j$  representing the certainty (or fidelity) of all the rays that intersect the  $j^{\text{th}}$  voxel in the image. The “certainty” quantity  $\kappa_j$  can be computed as:

$$\kappa_j = \sqrt{\sum_i a_{ij}^2 \mathbf{w}_i / \sum_i a_{ij}^2} \quad (3.3)$$

where  $a_{ij}$  denotes the  $(i, j)^{\text{th}}$  element of the matrix  $\mathbf{A}$ . The regularization term can then be written as:

$$R_R(\mu) = \frac{1}{2} \beta \sum_j \sum_{k \in N_j} \kappa_j \kappa_k w_{jk} \varphi(\mu_j - \mu_k) \quad (3.4)$$

where the penalty between each pair of neighboring voxels is further weighted by the “certainty” quantity. This regularization method is referred to as the “certainty-based” or “uniform resolution” penalty (denoted by subscript  $R$ ). In this chapter, the certainty-based penalty is considered not as a competing method to the task-based approach proposed below – after all, each seeks to optimize with respect to a particular objective – but rather as a point of reference with respect to an alternative form of spatially varying penalty.

To design the regularization term in such a way as to maximize task-based detectability at any location within the object, one can optimize the penalty function ( $\varphi$ ), directional weight ( $w_{jk}$ ), and/or regularization parameter ( $\beta$ ) with respect to a specific location. The work in this chapter focuses on determination of a spatially varying regularization parameter ( $\beta_j$ ) to maximize local task-based detectability at every location within the object while keeping the penalty function and directional weights the same for all locations. The design of the directional weights and penalty function to further improve task-based detectability are the subject of possible future work. This regularization method is referred to as a “task-based penalty” (denoted by subscript  $T$ ), written as:

$$R_T(\mu) = \frac{1}{2} \sum_j \beta_j \sum_{k \in N_j} w_{jk} \varphi(\mu_j - \mu_k) \quad (3.5)$$

Note that the regularization parameter  $\beta_j$  is now inside the outer summation and dependent on the location  $j$ , unlike the conventional penalty.

For all three regularization methods in the current work, a quadratic penalty function was selected, and the directional weights were set to 1 for first-order neighbors (e.g., 6 first-order neighbors for a voxel in a 3D image) and 0 for higher-order neighbors. Future work could consider a non-quadratic (e.g., Huber) penalty, recognizing the need for modified performance prediction (described in Sec. 3.2.2 for the quadratic penalty). Similarly, directional weights could be modified to increase / decrease smoothing in certain directions to potentially improve detectability.

### 3.2.2 Task-based performance prediction

To derive a task-based penalty, one needs to estimate the local detectability index ( $d'_j$ ), which in turn requires prediction of local spatial resolution (MTF<sub>j</sub>) and noise (NPS<sub>j</sub>) characteristics. Previous work<sup>132</sup> shows that the effects of a spatially varying regularization term ( $\beta_j$ ) on image quality can be considered local if the spatially varying term is spatially smooth. Thus, although  $\beta_j$  will be spatially varying in the resulting penalty, one can assume that  $\beta_j$  is constant in the design stage. In other words, one may assume a conventional (constant) penalty when estimating local spatial resolution, noise, and task-based detectability in a PWLS image.

Fessler *et al.*<sup>132</sup> derived the PSF estimate of PL reconstruction with a conventional (constant) quadratic penalty using the Implicit Function Theorem and first-order Taylor expansion. We apply the same prediction model to PWLS reconstruction as a simplified case of the PL estimate. The PSF estimate at the  $j^{\text{th}}$  voxel can be written as:

$$PSF_j(\hat{\mu}) = [\mathbf{F}(\tilde{\mu}) + \beta \mathbf{R}(\tilde{\mu})]^{-1} \mathbf{F}(\mu^{\text{true}}) e_j \quad (3.6)$$

$$\mathbf{F}(\mu) = \mathbf{A}^T \mathbf{D} [\bar{\mathbf{y}}(\mu)] \mathbf{A} \quad (3.7)$$

where  $\tilde{\mu}$  denotes PWLS reconstruction from noiseless measurements,  $\mu^{true}$  is the truth image,  $\bar{y}(\mu)$  is the forward projection of the image estimate  $\mu$ , and  $e_j$  is a unit vector (Kronecker delta) equal to 1 at the  $j^{\text{th}}$  element (and 0 elsewhere). The Fisher information matrix  $\mathbf{F}(\mu)$  usually has nonuniform diagonal elements, resulting in nonuniform noise-resolution characteristics in the image reconstruction. The term  $\mathbf{R}(\tilde{\mu})$  denotes the Hessian of the regularization term, which is independent of the input  $\mu$  when a quadratic penalty function is used. When  $\tilde{\mu}$  and  $\mu^{true}$  are not available (e.g., estimating the PSF from real data), one can substitute  $\bar{y}(\tilde{\mu})$  and  $\bar{y}(\mu^{true})$  with measured projection data. This is typically a good approximation for x-ray fluence in the diagnostic range, because the measured projection data are sandwiched between forward projection and backprojection [Eq. (3.7)], which greatly reduces the effect of noise on the estimation.<sup>132</sup>

Because the PSF is estimated in a relatively uniform region (i.e., the brain parenchyma), one can assume it to be locally invariant within a small neighborhood  $N$  of the location of interest. The matrix representation of the local PSF (i.e., the “system matrix”) can then be approximated as circulant with shifted copies of the PSF as its column entries.<sup>188</sup> Thus, the MTF can be computed as the magnitude of the discrete Fourier transform (DFT) of the PSF normalized at zero frequency<sup>136</sup>:

$$H_j(f) = FT\{PSF_{j,N}(\hat{\mu})\} \quad (3.8)$$

$$MTF_j(f) = |H_j(f)| / |H_j(0)| \quad (3.9)$$

where  $PSF_{j,N}(\hat{\mu})$  denotes the PSF estimate for the  $j^{\text{th}}$  voxel in a small neighborhood  $N$ , and  $H_j(f)$  denotes the DFT of the PSF.

The covariance estimate for PWLS reconstruction with a conventional penalty (constant  $\beta$ ) can be derived in a similar manner as the PSF estimate for a quadratic penalty. According to the definition of covariance matrix, the covariance of the  $j^{\text{th}}$  voxel with all other voxels in the image is a column in the covariance matrix, written as:

$$Cov_j(\hat{\mu}) = [\mathbf{F}(\tilde{\mu}) + \beta\mathbf{R}(\tilde{\mu})]^{-1} \mathbf{F}(\mu^{true}) [\mathbf{F}(\tilde{\mu}) + \beta\mathbf{R}(\tilde{\mu})]^{-1} e_j \quad (3.10)$$

Similar to the PSF, one can assume that the covariance of a given voxel (with other voxels) in a small, relatively uniform neighborhood  $N$  is the same as the covariance of any other voxel (with other voxels) in the neighborhood, resulting in an approximately circulant covariance matrix. Thus, the local NPS is described by the magnitude of the DFT of a column of the covariance matrix, which is far more tractable than the DFT of the entire covariance matrix. The local NPS about voxel  $j$  is then<sup>136</sup>:

$$NPS_j(f) = \left| FT \{ Cov_{j,N}(\hat{\mu}) \} \right| \quad (3.11)$$

where  $Cov_{j,N}(\hat{\mu})$  denotes the covariance of the  $j^{\text{th}}$  voxel with all other voxels within a small neighborhood  $N$ .

The formulations in Eqs. (3.6) – (3.11) reveal the dependence of spatial resolution and noise on the object ( $\hat{\mu}$ ) and location in the object ( $j$ ). It is important to note that the dependence on the object is realized only through its projections; therefore, estimation of the  $MTF_j$  and  $NPS_j$  does not require knowledge of the true object nor a reconstruction of the object - estimated instead directly from the projection data. In the studies below, the PSF and covariance [Eqs. (3.6) and (3.10)] were estimated using a linear conjugate gradient (CG) algorithm. Specifically, the CG algorithm was applied once to invert the matrix  $[F(\tilde{\mu}) + \beta R(\tilde{\mu})]$  in Eq. (3.6) and was applied twice to invert the two matrices  $[F(\tilde{\mu}) + \beta R(\tilde{\mu})]$  in Eq. (3.10).

Based on predictions of the local MTF and NPS, one may compute the local task-based detectability index ( $d'_j$ ) for PWLS reconstruction. The detectability index relates the MTF and NPS to a spatial-frequency-dependent task function via an observer model, such as the Fisher-Hotelling (prewhitening) or non-prewhitening (NPW) models.<sup>186</sup> While many observer models can be formulated, a NPW matched filter observer model is used in the current work. This model does not include “anthropomorphic” parameters relating to a human observer (e.g., a mid-frequency “eye filter” and internal noise) and instead encapsulates the intrinsic signal and noise characteristics of the image with respect to an observer that cannot decorrelate noise. The NPW model has

demonstrated reasonable agreement with human observer performance for simple tasks in tomosynthesis and CBCT.<sup>64</sup> This observer model was also used in previous work on design of a CBCT head scanner using task-based imaging performance.<sup>183</sup> The local detectability index for a NPW observer is given by:

$$d_j'^2(\beta_j) = \frac{\left\{ \iiint \left[ MTF_j(\beta_j) \cdot W_{Task} \right]^2 df_x df_y df_z \right\}^2}{\iiint NPS_j(\beta_j) \left[ MTF_j(\beta_j) \cdot W_{Task} \right]^2 df_x df_y df_z} \quad (3.12)$$

Note that  $d_j'$ ,  $MTF_j$ , and  $NPS_j$  are all defined with respect to the  $j^{\text{th}}$  voxel and are functions of the local regularization parameter  $\beta_j$ . The task function,  $W_{Task}$  (detailed below), may be similarly taken to vary in space, but in the current work was held fixed – i.e., the same task function at all locations in the image. The MTF, NPS, and  $W_{Task}$  are all three-dimensional Fourier-domain functions of  $(f_x, f_y, f_z)$ , and the integrals in Eq. (3.12) are over the 3D Fourier domain bounded by the Nyquist frequency of the image reconstruction.

A task function is considered that describes detection of a small, low-contrast lesion emulating ICH, with spatial frequency content modeled as a difference of two Gaussians to emphasize mid-frequency content. This can be interpreted as discrimination of two Gaussian structures or detection of the stimulus edge in the spatial domain.<sup>183</sup> The task function can be expressed as:

$$W_{Task} = C \left[ \exp\left(-\frac{|f|^2}{2\sigma_1^2}\right) - \exp\left(-\frac{|f|^2}{2\sigma_2^2}\right) \right] \quad (3.13)$$

where  $C$  denotes the contrast of ICH with respect to the brain (50 HU), and  $\sigma_1$  and  $\sigma_2$  are the widths of the two Gaussians. Figure 3.2(b) shows the mid-frequency task function  $W_{Mid}$  used primarily in this work, with contrast  $C = 50$  HU,  $\sigma_1 = 0.35 \text{ mm}^{-1}$ , and  $\sigma_2 = 0.25 \text{ mm}^{-1}$ . This task function peaks at approximately  $0.5 \text{ mm}^{-1}$ , corresponding to a characteristic feature size of  $\sim 1 \text{ mm}$ . Two additional task functions are considered to further evaluate the sensitivity of the proposed method to  $W_{Task}$ ,

including: a low-frequency task described by a single Gaussian with  $\sigma = 0.2 \text{ mm}^{-1}$  [denoted  $W_{Low}$  and shown in Fig. 3.2(a)]; and a high-frequency task described by Eq. (3.13) with  $\sigma_1 = 0.70 \text{ mm}^{-1}$  and  $\sigma_2 = 0.65 \text{ mm}^{-1}$  [denoted  $W_{High}$  in Fig. 3.2(c)]. To better allow inter-comparison among the three tasks, the magnitude of each task function was scaled (via  $C$ ) such that the signal power – i.e., the integral of  $(W_{Task})^2$  over the Nyquist region was the same.

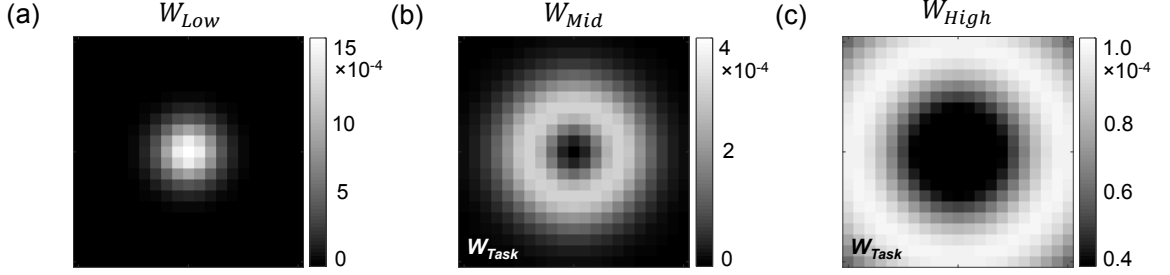


Figure 3.2: Three task functions investigated in the current work: (a) low-frequency, (b) mid-frequency, and (c) high-frequency. The mid-frequency task was modeled as a difference of two Gaussians with 50 HU contrast, corresponding to a low-contrast feature (acute ICH) with characteristic length of  $\sim 1 \text{ mm}$ .

### 3.2.3 Task-based regularization

An objective function can be formulated based on the local task-based detectability index to solve for the local regularization parameter  $\beta_j$  that maximizes  $d_j'^2$ :

$$\hat{\beta}_j = \underset{\beta_j}{\operatorname{argmax}} d_j'^2(\beta_j) \quad (3.14)$$

The optimal  $\beta_j$  is solved by evaluating  $d_j'$  for different  $\beta_j$  values with regular spacing and choosing the  $\beta_j$  that yields maximum  $d_j'$ . Future work could consider a direct analytical solution of the maximization (setting the gradient of the objective function to zero); however, in the current work, Eq. (3.14) was solved numerically.

Directly repeating the maximization at every spatial location yields a spatially varying  $\beta_j$  map that maximizes  $d'$  everywhere in the image. Considering the large number of voxels typically in a

CBCT image (e.g.,  $N_\mu = 500^3$ ) and positing that the  $\beta_j$  map will be slowly varying throughout the fairly uniform region of the brain, a more computationally efficient approximation was used to accelerate the design process, as follows:

- 1) First, the optimal  $\beta$  was computed on a downsampled grid internal to the cranium and interpolated at intermediate voxels using radial basis functions. This step reduced the number of  $\beta$  computations from the number of voxels in the image (e.g.,  $N_\mu = 500^3$ ) to the number of voxels on the downsampled grid (e.g.,  $N_\mu^{Grid} = N_\mu / (DS_1)^3 = 20^3$  assuming a downsampling factor of  $DS_1 = 25$  in each dimension).
- 2) Second, recognizing that the local PSF and covariance reduce to zero at voxels sufficiently far from the voxel of interest, unity impulses are placed at multiple locations (instead of just one location) in the input  $e_j$  in Eqs. (3.6) and (3.10). This allows one to predict local PSF and covariance for multiple locations simultaneously. For example, if one places unity impulses on a subgrid with a spacing of  $DS_2$  voxels, it only takes one simultaneous prediction to compute the local PSF and covariance for each voxel on this subgrid. As a result, the downsampled grid from the first acceleration step can be divided into  $K = (DS_2 / DS_1)^3$  subgrids (e.g.,  $K = (50/25)^3 = 8$ , assuming  $DS_2 = 50$ ), which requires only  $K$  simultaneous predictions.

The two acceleration steps reduce the number of predictions needed from the number of voxels in the image (e.g.,  $N_\mu = 500^3$ ) to the number of subgrids (e.g.,  $K = 8$ ). A pseudocode outline of the accelerated design framework is shown in Table 3.1.

Table 3.1: Spatially varying regularization design for maximal task-based performance.

---

```

Input precomputed  $\mathbf{R}$ 
for each subgrid  $k = 1$  to  $K$ 
    Place  $N_k$  unity impulses in  $e_k$  with uniform spacing
    Predict local PSF and covariance for  $N_k$  locations simultaneously using Eqs. (3.6) and (3.10)
    for each voxel  $j$  on the  $k^{th}$  subgrid
        Compute local  $MTF_j$ ,  $NPS_j$ , and  $d'_j$  for different  $\beta_j$  using Eqs. (3.9), (3.11), and (3.12)
        Estimate  $\beta_j$  that maximizes  $d'_j$ 
    end
end
for each voxel not on the downsampled grid
    Interpolate  $\beta_j$  based on the optimal  $\beta_j$  on the downsampled grid
end
return a  $\beta$  map

```

---

### 3.2.4 CBCT simulation studies

The task-based penalty approach was first evaluated in simulation studies in comparison to both the conventional penalty and the uniform resolution penalty. An anthropomorphic head phantom was scanned using a MDCT scanner (SOMATOM Definition, Siemens Healthineers, Erlangen, Germany) to generate the 3D digital head phantom illustrated in Fig. 3.3(a). A high-dose scan protocol (120 kVp, 500 mAs,  $0.5 \times 0.5 \times 0.6$  mm<sup>3</sup> voxels, H30s reconstruction filter) was used to reduce noise and provide easy segmentation for producing the digital phantom. Soft-tissue structures were segmented and set to a constant value of 40 HU, and bone tissues retained the natural reconstructed attenuation values. The system geometry of the CBCT head scanner prototype<sup>183</sup> was used, including 550 mm SAD, 1000 mm SDD, and  $0.556 \times 0.556$  mm<sup>2</sup> detector pixel size. Image simulation involved 720 projections of the phantom over 360° with  $2 \times 10^5$  photons per detector pixel. Both noiseless and Poisson-distributed x-ray projections were simulated. The 3D image reconstruction grid contained  $390 \times 485 \times 498$  voxels with  $0.5 \times 0.5 \times 0.5$  mm<sup>3</sup> voxel size.

Local PSF was predicted using the CG algorithm in Eq. (3.6) (with 100 iterations to reach convergence), and local covariance was predicted similarly for Eq. (3.10). The simultaneous prediction was performed  $K = 8$  times, each time on a subgrid with a spacing of  $DS_2 = 50$  voxels.

The optimal  $\beta_j$  was identified for each of the voxels on the 8 subgrids, from which a spatially varying  $\beta$  map was interpolated using radial basis functions.<sup>189</sup> The neighborhood used in computing the DFT of the local PSF and covariance was  $21 \times 21 \times 21$  voxels. PWLS images were reconstructed as described in Sec. 2.2.6 by initializing with a filtered backprojection image and applying 100 iterations of separable quadratic surrogate (SQS) updates with 20 ordered subsets<sup>163</sup> to reach convergence.

The local MTF and NPS were also measured from reconstructed images and were used to evaluate the accuracy of the predictions. The local MTF was measured following Eq. (10) in the paper by Fessler and Rogers<sup>132</sup> by subtracting two PWLS reconstructions of noiseless projection data with and without an impulse at the location of interest. The local NPS was measured from a large ensemble ( $n = 200$ ) of PWLS reconstructions with different realizations of Poisson noise – specifically, calculated following Eq. (26) in the paper by Gang *et al.*<sup>136</sup> as the sample average of squared Fourier transform of a small neighborhood in the difference image between two PWLS images with different noise realizations.

Both the prediction and image reconstruction methods were implemented in Matlab (The Mathworks, Natick MA), with projection operations implemented using matched separable footprint method<sup>98</sup> and executed using CUDA libraries on GPU. The execution of the CG and OS-SQS algorithms for all 720 projections in one iteration required about the same time ( $\sim 3.6$  min), because both algorithms required one forward projection and one backprojection in one iteration (which were the dominant factor for computation time). Further reduction in computation time (beyond the scope of this chapter) may be achieved using methods such as CG preconditioning<sup>190</sup> to speed the PSF estimation and acceleration techniques such as spatially non-uniform updates<sup>167</sup> and Nesterov's method<sup>168,191</sup> to speed image reconstruction.

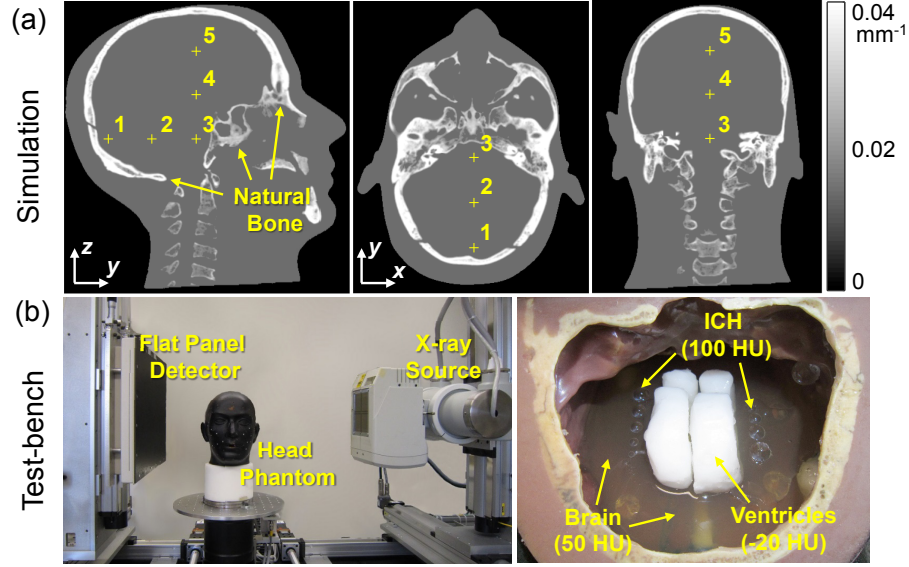


Figure 3.3: Head phantom and CBCT test-bench. (a) Sagittal, axial, and coronal slices of the 3D digital head phantom used in the simulation studies, which contains realistic bone attenuation and uniform soft tissue. Five locations of interest are denoted 1-5. (b) Photographs of the CBCT test-bench and an anthropomorphic head phantom filled with materials emulating brain, ventricles, and acute ICH.

### 3.2.5 Experiments on a CBCT test-bench

The proposed task-based reconstruction technique was also evaluated in experiments on a CBCT test-bench equipped with a flat-panel detector (PaxScan 4343R, Varian, Palo Alto CA) as shown in Fig. 3.3(b). A custom anthropomorphic head phantom containing a natural skull in soft-tissue equivalent plastic (Rando™, The Phantom Laboratory, Greenwich NY) was filled with a gelatin mixture, ventricle models prepared from wax, and plastic spheres of different diameters to provide x-ray attenuation approximating brain (50 HU), cerebrospinal fluid (-20 HU), and fresh blood (100 HU). The resulting contrast of brain to fresh blood (~50 HU) was consistent with previous reports of acute ICH presentation in CT.<sup>142,165</sup> System geometry similar to that in simulation studies and the head scanner prototype was employed: SAD = 580 mm, SDD = 800 mm, and detector readout with 0.556 mm isotropic pixel size. The phantom was scanned at 100 kVp and 0.4 mAs per projection for 720 projections over 360°, yielding a total dose of 24.8 mGy (measured at the center

of a 16 cm CTDI phantom placed at isocenter) which was somewhat lower than that from a typical adult head CT [ $\sim 44.2$  mGy reported by Huda *et al.*<sup>192</sup>].

Projection data were corrected for x-ray scatter using a Monte Carlo scatter correction method<sup>149</sup> and for beam hardening using the Joseph and Spital method.<sup>157</sup> The statistical weights in PWLS were modified accordingly to account for changes in variance following artifact correction following the method described in Sec. 2.2.5. The projections were also corrected for detector lag and veiling glare as in the work by Sisniega *et al.*,<sup>149</sup> but these corrections were not explicitly considered in the statistical weights. The image grid contained  $412 \times 512 \times 512$  voxels with  $0.5 \times 0.5 \times 0.5$  mm<sup>3</sup> voxel size. The task-based regularization design and PWLS reconstructions were performed using the same methods and parameter settings as in the simulation studies. In a PWLS image, spatial resolution was assessed in terms of the ESF of a simulated ICH sphere in an axial slice. A sigmoid function parameterized by width  $\varepsilon$  was fit to all the voxels within  $60^\circ$  fan-shaped sectors centered on the simulated ICH, and spatial resolution was quantified as the average  $\varepsilon$  computed over each sector.<sup>93</sup> Image noise was quantified as the standard deviation of voxel values within a small region of interest of  $19 \times 19$  voxels in a uniform region of the brain adjacent to the simulated ICH.

### 3.3 Results

#### 3.3.1 Local MTF and NPS

Previous work by Gang *et al.*<sup>136</sup> studied the shift-variance and anisotropy of 2D local MTF and NPS in penalized likelihood reconstruction for fan-beam CT. The 3D local MTF and NPS were analyzed similarly in this chapter for CBCT of the head as summarized in Fig. 3.4, which shows the local MTF and NPS in the  $(f_x, f_y)$  and  $(f_y, f_z)$  planes at five locations in the head as denoted in Fig. 3.3(a). The reconstruction method was PWLS with the conventional (constant) regularization

parameter  $\beta = 10^{6.4}$ . The local MTF and NPS exhibit shift-variance and anisotropy as expected. In particular, the local MTF is narrower (i.e., reduced spatial resolution) and local NPS is elevated (especially at low frequencies) near the skull base (location 3) compared to regions adjacent to the cranium (locations 1 and 5). In addition to higher MTF, the peripheral locations (1 and 3) exhibit more anisotropic local NPS, owing to the strong difference in attenuation for rays traversing, for example, in PA views versus the LAT views. Such characteristics result from the interplay between the object / attenuation length and data-dependent smoothing applied by PWLS: rays traversing location 3 (near the center of the head and surrounded by thick bone structures such as the temporal bones) are more attenuated (noisier) than rays traversing peripheral locations 1 and 5; therefore, the conventional PWLS algorithm applies greater smoothing to data corresponding to location 3 than locations 1 and 5.

Figure 3.4 also compares the measured local MTF and NPS with that predicted by Eqs. (3.6-3.9) and (3.10-3.11), respectively. Each case demonstrates reasonably good agreement in terms of both spatial dependence and anisotropy. Such agreement is similarly observed when the location of interest is far from the central slice (location 4 and 5), where the  $(f_y, f_z)$  domains exhibit a “null cone” of unsampled frequencies due to the cone-beam and circular orbit geometry.

Reasonable agreement between measurement and prediction was also observed for other  $\beta$  values (not shown for brevity) over the range of  $10^{5.0} \sim 10^{8.0}$ , which is the relevant range for  $d'$  optimization investigated below.

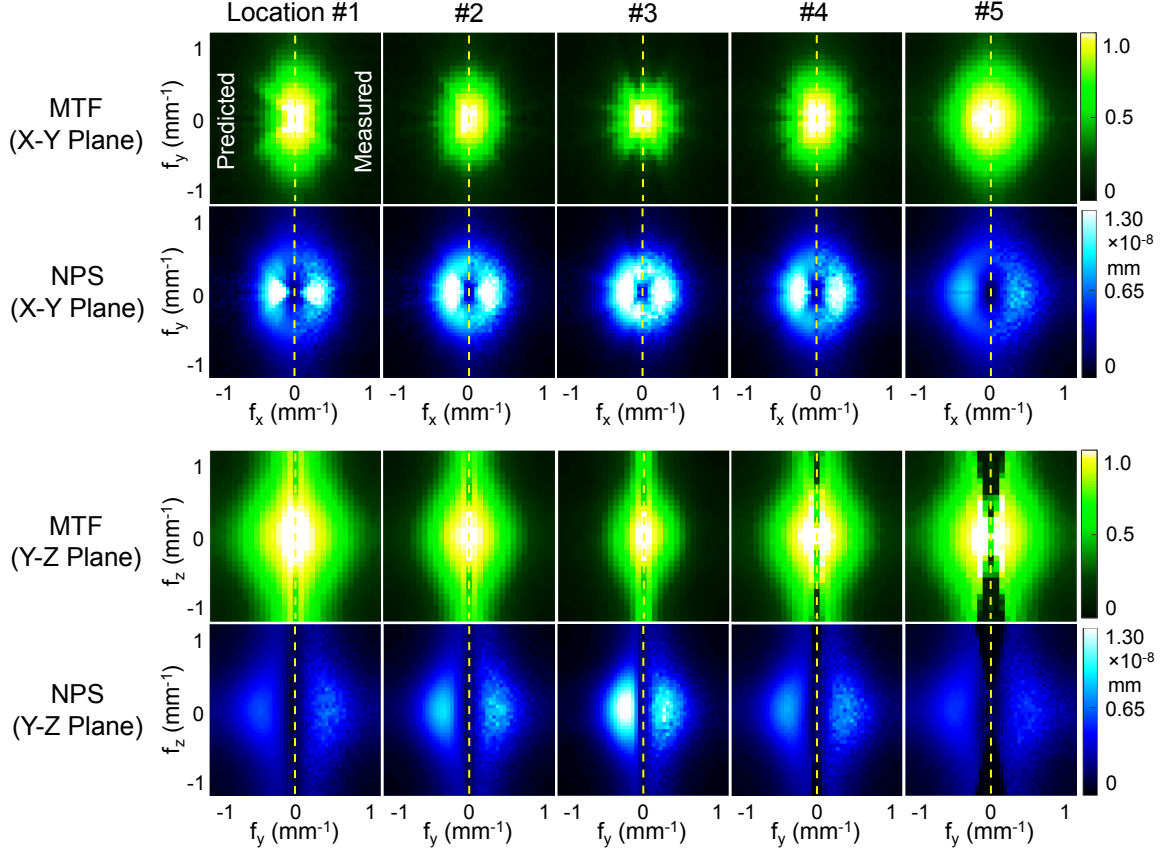


Figure 3.4: Local MTF and NPS in 3D images of the head reconstructed using PWLS with a conventional penalty (constant  $\beta = 10^{6.4}$ ). Each exhibits shift-variance and anisotropy at locations defined in Fig. 3.3(a). The top two rows show “axial” MTF( $f_x, f_y$ ) and NPS( $f_x, f_y$ ), and the bottom two rows show “sagittal” MTF( $f_y, f_z$ ) and NPS( $f_y, f_z$ ), the latter demonstrating the null cone of unsampled frequencies in regions (e.g., location 5) far from the central slice.

### 3.3.2 Task-based regularization (mid-frequency detection task)

Task-based regularization design was performed in simulation studies for the mid-frequency detection task  $W_{Mid}$  defined in Section 3.2.2. Figure 3.5 shows the resulting  $\beta$  map for the task-based penalty (i.e., that which maximizes  $d'$  at each location in the head) in comparison to the conventional constant penalty and the certainty-based uniform resolution penalty. For the conventional penalty, a scalar  $\beta$  of  $10^{6.4}$  was chosen to yield the highest average  $d'$  over all locations in the head - in this way, representing the best choice constant value. For the uniform resolution penalty, the  $\beta$  map was a product of a scalar  $\beta$  and spatially varying certainty  $\kappa_j^2$ , with the scalar  $\beta$

chosen to provide uniform PSF width of 0.95 mm (measured as the full width at half maximum (FWHM) of the local PSF averaged over all radial directions). The task-based design results in a  $\beta$  map exhibiting a similar overall structure to the uniform resolution penalty: each exhibits lower  $\beta$  near the skull base and higher  $\beta$  near the cranium compared to the conventional penalty; however, the  $\beta$  value in the task-based penalty exhibits a broader range (over an order of magnitude) than the uniform resolution penalty and – for the mid-frequency task considered here – presents lower  $\beta$  (i.e., reduced smoothing) in the interior of the skull.

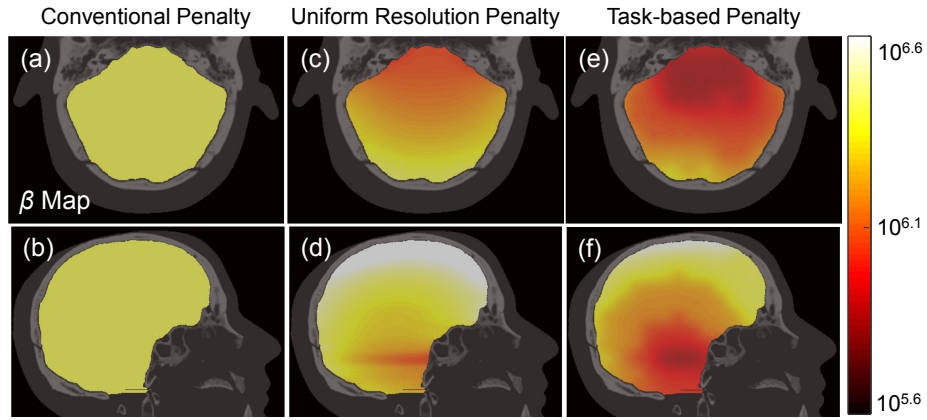


Figure 3.5: Maps of the regularization parameter  $\beta$  for the three penalty methods. (a-b) The conventional penalty employs a constant scalar value  $\beta$  ( $= 10^{6.4}$ , chosen to yield the highest  $d'$  averaged over all regions of the brain). (c-d) The certainty-based uniform resolution penalty involves the product of a scalar  $\beta$  and a spatially-varying certainty term  $\kappa_j^2$  (chosen to yield uniform PSF width of 0.95 mm). (e-f) The task-based penalty adjusts the  $\beta$  map to maximize  $d'$  everywhere in the brain (chosen here for a mid-frequency task,  $W_{Mid}$ ).

Figure 3.6 shows the resulting spatial resolution and noise characteristics in the head from the three penalty methods in Fig. 3.5. The conventional penalty results in higher spatial resolution near the cranium and lower near the skull base, consistent with the results of local PSF in Fig. 3.4. The uniform resolution penalty leads to more uniform spatial resolution (PSF width  $\sim 0.95$  mm) than the conventional penalty but results in stronger spatial variation in noise (lower near the cranium and higher near the skull base). The task-based penalty results in nonuniform spatial resolution and noise, which is not surprising considering that the penalty is designed to maximize  $d'$  at each location, not to improve uniformity in resolution or noise.

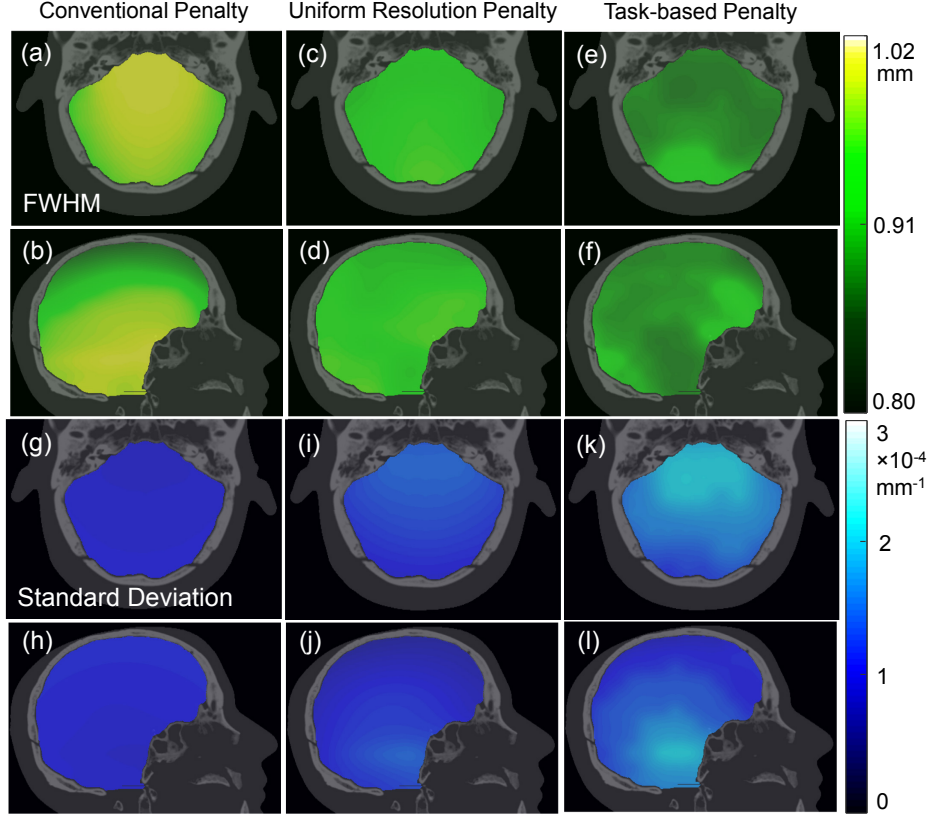


Figure 3.6: Spatial resolution and noise for the three penalty methods. (a-f) Spatial resolution is described by the FWHM of the local PSF, and (g-l) noise is given by the local standard deviation in voxel values. The conventional penalty yields (a-b) nonuniform spatial resolution and (g-h) fairly uniform noise. The certainty-based penalty yields (c-d) uniform spatial resolution and (i-j) slightly stronger nonstationarity in the noise. The task-based penalty yields spatially varying resolution and noise characteristics that adjust in a manner to maximize local  $d'$  (chosen here for the mid-frequency task,  $W_{Mid}$ ).

Figure 3.7 shows the resulting  $d'$  map for the three penalty methods, all of which exhibit strong spatial variation in  $d'$  - highest near the cranium and reduced in the interior of the cranial vault. Compared to the “best” conventional penalty (i.e., highest average  $d'$ ), the uniform resolution penalty increased  $d'$  by up to 10% in certain regions of the brain (near the skull base); however, the uniform resolution penalty actually reduced  $d'$  in other regions (near the periphery). These results are somewhat expected, since the uniform resolution penalty is designed not to maximize  $d'$  but to achieve uniform spatial resolution.

The task-based penalty exhibits the highest  $d'$  value at all locations – equal to or exceeding that of the other penalties by up to 12%; moreover, it does not exhibit the slight reduction in  $d'$  near the periphery. Given the objective function for each penalty method, this result is somewhat expected – i.e., that the task-based penalty exhibits the highest  $d'$ , since that is what it was designed to do. The extent to which this corresponds to an increase in visual image quality depends on the extent to which the observer model (in this case, NPW) and task function (in this case,  $W_{Mid}$ ) provide a realistic quantification of image quality. This point is investigated further in Sec. 3.3.4, with future work to include alternative observer models and more in depth human observer studies.

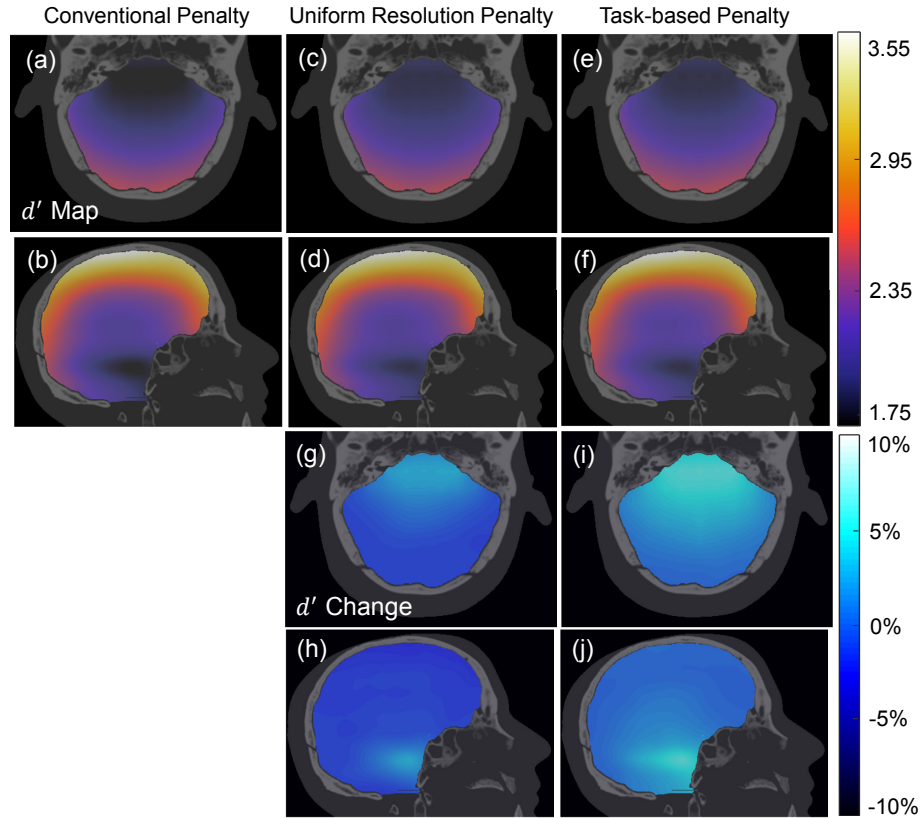


Figure 3.7: Local detectability for the three penalty methods. The top two rows (a-f) show maps of  $d'$  (mid-frequency detection task) for the (a-b) conventional constant penalty; (c-d) uniform resolution penalty; and (e-f) task-based penalty. The bottom two rows (g-j) show the ratio of detectability index for the (g-h) uniform resolution penalty and (i-j) task-based penalty to that of the conventional penalty. The uniform resolution penalty increases  $d'$  near the skull base by  $\sim 10\%$  but degrades  $d'$  near the periphery by  $\sim 5\%$ . The task-based penalty improves  $d'$  by up to  $\sim 12\%$  near the skull base and meets or exceeds that of other methods at every location in the head.

### 3.3.3 Task-based regularization (various tasks)

The task-based regularization method was further investigated for tasks emphasizing different frequency content, including a low-frequency task ( $W_{Low}$ , Fig. 3.2(a)), a mid-frequency task ( $W_{Mid}$ , Fig. 3.2(b), investigated in the previous section), and a high-frequency task ( $W_{High}$ , Fig. 3.2(c)). Figure 3.8 summarizes  $d'$  as a function of  $\beta$  for each task (denoted  $d'_{Low}$ ,  $d'_{Mid}$ , and  $d'_{High}$ ) at locations near the skull base (location 3), in the center of the brain (location 4), and near the cranium (location 5). Recall that the three task functions were separately scaled to maintain constant signal power and thereby support intercomparison in terms of spatial-frequency response characteristics (rather than simply contrast).

For all tasks and locations considered, the function  $d'(\beta)$  was concave and exhibited a clear optimum in  $\beta$  (within the range  $\beta = 10^{5.0} \sim 10^{8.0}$ ), suggesting the possibility of directly solving for the optimal  $\beta$  using optimization algorithms. For all three locations,  $d'$  is seen to be higher for the low-frequency task and lower for the high-frequency task, consistent with the more challenging nature of a high-frequency task for a system with low-mid bandpass characteristics. For  $W_{Low}$ , the task-based penalty improved  $d'$  by 5.8% near the skull base (location 3) and 0.5% near the cranium (location 5) compared to the conventional (constant) penalty. Moreover, for  $W_{High}$ , the task-based approach improved  $d'$  by 10.4% near the skull base and 1.7% near the cranium. Gains in detectability were slightly less for the uniform resolution penalty and, in some cases, slightly worse than the conventional penalty (e.g., a 0.9% reduction in  $d'$  for  $W_{High}$  at location 5).

The optimal  $\beta$  is seen to depend somewhat on the imaging task, which could be problematic considering that one may desire to accomplish many tasks within a given image (as opposed to reconstructing a separate image for each task - which is possible, but may be impractical for more than  $\sim 2$  tasks). For example, the low-frequency task at location 3 was optimized with  $\beta_j = 10^{6.8}$ , however, this  $\beta_j$  value results in  $d'_{High}$  that is  $\sim 47\%$  lower than its maximum value (realized with

$\beta_j = 10^{5.2}$ ). Conversely, selecting  $\beta_j$  to maximize  $d'_{High}$  reduces  $d'_{Low}$  by only 12.6% compared to its maximum value. Analogous to common practice in which “sharp” and “smooth” images are reconstructed for visualization of “bone” and “soft” tissue features, one could reconstruct images with task-based penalties corresponding to  $W_{High}$  and  $W_{Low}$ , respectively. Alternatively, one could consider  $W_{Mid}$  as a nominal choice, optimizing mid-frequency detection with less tradeoff for high- and low-frequency tasks. For example, selecting  $\beta_j$  to maximize  $d'_{Mid}$  corresponds to  $d'_{High}$  and  $d'_{Low}$  values that are reduced by 9.5% and 6.1% from their maximum values, respectively. Previous work by Yang *et al.*<sup>135</sup> also found that a penalty designed to maximize  $d'$  for small lesions did not adversely affect the detection of larger lesions in 3D PET of the breast.

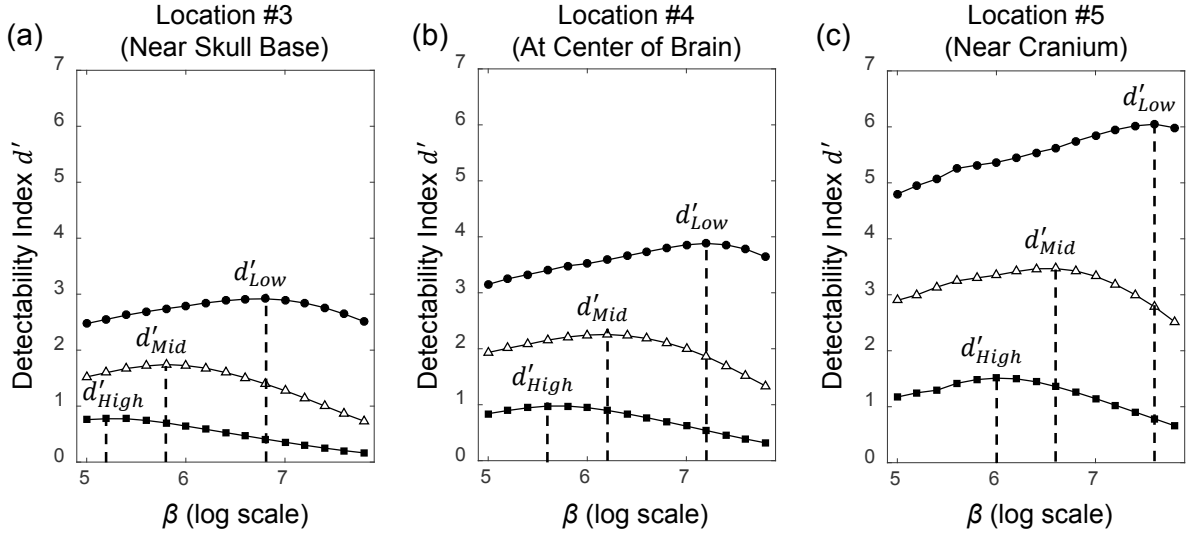


Figure 3.8: Detectability index  $d'$  computed as a function of regularization parameter  $\beta$  at locations (a) near the skull base, (b) in the center of the brain, and (c) near the cranium. Each plot shows  $d'$  computed for a low-frequency task (denoted as  $d'_{Low}$ ), a mid-frequency task ( $d'_{Mid}$ ), and a high-frequency task ( $d'_{High}$ ).

### 3.3.4 Image reconstructions

The simulation studies described in Sec. 3.2.4 included a digital anthropomorphic head model in which spheres of 2 mm diameter and 50 HU contrast were added to emulate acute ICH. Three such

lesions were added to a region in the deep interior of the brain near the skull base (location 3 in Fig. 3.3) and to a region at the periphery adjacent to the cranium (location 5). Projections were simulated with Poisson noise as described in Sec. 3.2.4 and reconstructed by PWLS with each of the three penalty methods. Figure 3.9 summarizes the results. As shown in Fig. 3.9(a,b), the conventional penalty (again chosen with constant  $\beta$  value to maximize average  $d'$ ) exhibited good visualization of ICH near the cranium but strongly smoothed the data near the skull base, resulting in an arguably over-smoothed appearance. As shown in Fig. 3.9(c,d), the uniform resolution penalty yielded a more uniform overall image appearance (viz., in terms of spatial resolution) and somewhat improved the conspicuity of the lesions near the skull base ( $d'$  improved by 7.2%); however, conspicuity near the cranium was slightly reduced ( $d'$  reduced by 2.3%) – though still conspicuous for the lesions shown. As demonstrated in Fig. 3.9(e,f), the task-based penalty improved conspicuity of the lesions near the skull base ( $d'$  increased by 12.3%) without reduction near the cranium.

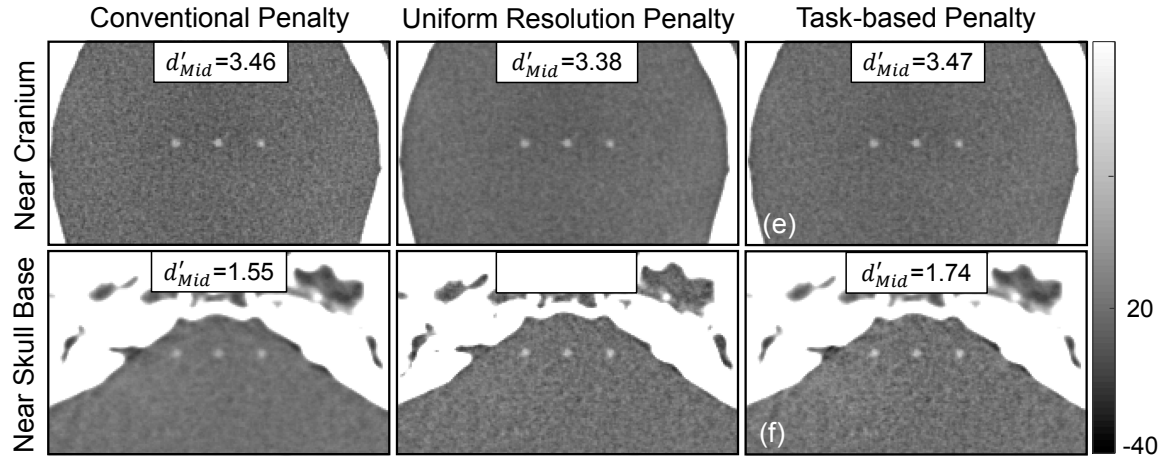


Figure 3.9: Visualization of simulated ICH lesions using three regularization methods. The conventional penalty ( $\beta=10^{6.4}$ ) exhibited a somewhat over-smoothed image near the skull base. The uniform resolution penalty yielded a more uniform image appearance (uniform PSF) and improved conspicuity near the skull base, with slight tradeoff near the cranium. The task-based penalty further improved conspicuity at each location, particularly near the skull base.

The task-based regularization design was finally applied to data acquired on the CBCT test-bench. The local PSF and covariance were estimated by substituting  $\bar{y}(\bar{\mu})$  and  $\bar{y}(\mu^{true})$  with

the measured projection data as described in Sec. 3.2.2. The task-based penalty was designed for the mid-frequency task in Fig. 3.2(b). Figure 3.10 shows PWLS reconstructions of a simulated ICH of 50 HU contrast and 12 mm diameter both near the cranium and near the skull base. The large simulated lesion is clearly detectable in all cases, but differences in conspicuity of the lesion extent (i.e., detection of its edge) can be appreciated among the three regularization approaches. As shown in Fig. 3.10(a,b), the conventional penalty (constant  $\beta = 10^{5.6}$  to maximize average  $d'$ ) exhibited strongly varying spatial resolution and noise characteristics in the two regions and over-smoothed the image near the skull base. As shown in Fig. 3.10(c,d) and consistent with the simulation studies of Fig. 3.9, the uniform resolution penalty yielded qualitatively improved conspicuity of the ICH lesion near the skull base ( $d'$  improved by 16.7%) but slightly reduced conspicuity near the cranium ( $d'$  reduced by 2.9%). Finally, as shown in Fig. 3.10(e,f), the task-based design improved conspicuity of the lesion (in particular, identification of the lesion edge) both near the skull base and periphery.

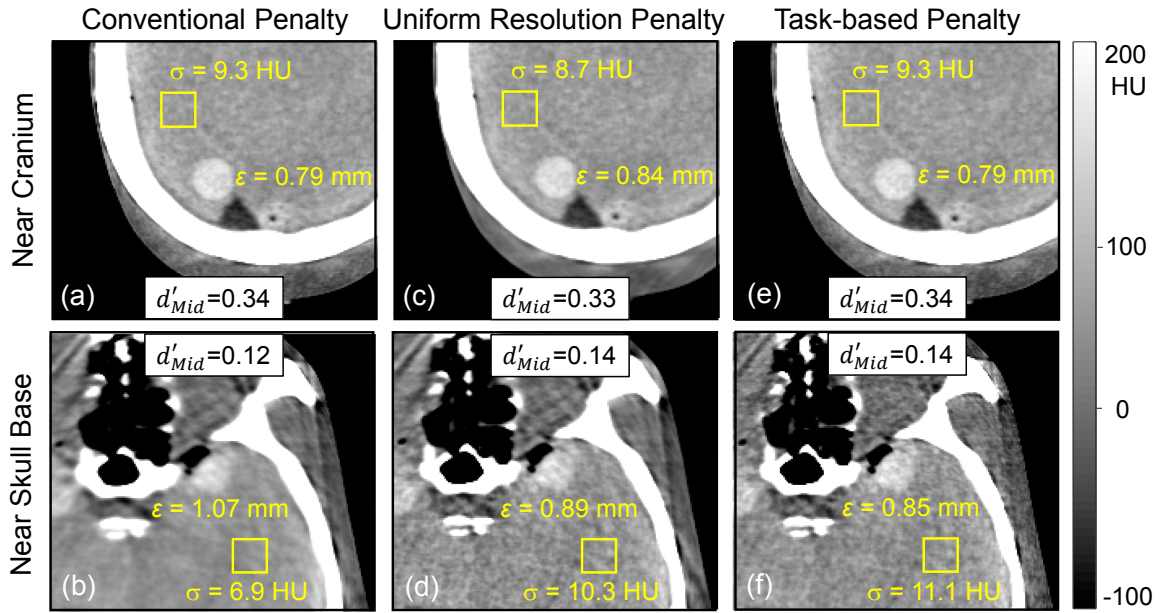


Figure 3.10: PWLS reconstructions of a head phantom containing a simulated ICH of 12 mm diameter and 50 HU contrast in regions near the cranium (a, c, e) and skull base (b, d, f). (a,b) Conventional penalty ( $\beta=10^{5.6}$ ). (c,d) Uniform resolution penalty. (e,f) Task-based penalty.

### 3.4 Conclusions and discussion

A regularization approach for MBIR was formulated that explicitly incorporates task-based imaging performance as the objective function, yielding a spatially varying penalty that maximizes task-based detectability index ( $d'$ ) at every location in the image. Theoretical predictions of local spatial resolution and noise were shown to agree with measurements in a realistic head phantom. The prediction framework for local MTF and NPS was leveraged to compute local detectability and adjust regularization strength ( $\beta$ ) throughout the image to maximize local detectability. The method was applied to imaging of small, low-contrast ICH lesions in CBCT of the head.

Simulations and test-bench experiments showed that a conventional (constant) penalty exhibits a fairly strong degree of variation in  $d'$  throughout the interior of the cranial vault, and while a certainty-based penalty achieved uniform PSF throughout the image, each exhibited a reduction in  $d'$  compared to the proposed task-based penalty (up to ~15% for certain tasks and locations). The improvement for the task-based penalty was strongest in areas of high attenuation (near the skull base) where the other two methods tended to over-smooth the data. The proposed method presents a promising means to improve task-based imaging performance in MBIR and could support the development of point-of-care CBCT systems for high-quality imaging of acute ICH in brain injury.

The design of regularization techniques in MBIR to improve task-based imaging performance represents an active area of research in recent years. Qi<sup>193</sup> derived fast computation of task-based detectability for lesion detection in PET, which was used in subsequent studies to improve lesion detectability compared to a conventional quadratic penalty at a known location in 2D<sup>194</sup> and at all possible locations in 3D.<sup>135</sup> Yendiki *et al.*<sup>195,196</sup> studied the degree of improvement in lesion detectability using regularization design for both location-known and location-unknown tasks and a number of commonly used observer models. Gang *et al.*<sup>136</sup> studied nonuniform spatial

resolution and noise characteristics in PL reconstruction and designed spatially varying penalties for a number of detection tasks in axial CT. The work in this chapter extended task-based regularization design to CBCT and investigated the detection of ICH lesions in CBCT of the head.

The current work is not without assumptions and limitations. First, the calculation of the Fourier metrics (i.e., MTF and NPS) assume that the system is linear and shift-invariant (LSI) and the noise is wide-sense stationary within a small neighborhood. The extent to which the “local” LSI assumption holds depends on a number of factors, including the degree of heterogeneity within the neighborhood and the size of the neighborhood. This work investigates a scenario in which the tissue in the region of interest (namely, the brain) is fairly uniform, so the “locality” assumption tends to hold well, as supported by the reasonable agreement observed between the prediction and measurement in Sec. 3.3.1. However, the extent to which this assumption holds deserves further investigation in other scenarios, for example, at the interface of bone and soft tissue and in the presence of highly attenuating objects (e.g., shunt, coil, or other surgical tools).

Furthermore, the current work employed a particular type of observer model (i.e., NPW), and other observer models may certainly be considered. The NPW observer model assumes the observer cannot decorrelate noise in the image and does not include various “anthropomorphic” characteristics of the observer, such as a bandpass eye filter or internal noise. This model has demonstrated good agreement with human observer performance for simple tasks in tomosynthesis and CBCT.<sup>64</sup> The prewhitening (PW) observer model, on the other hand, assumes that the observer can decorrelate noise, and previous work<sup>136</sup> showed that detectability based on the PW observer model exhibited weak dependence on the regularization parameter. In addition, the channelized Hotelling observer (CHO)<sup>197</sup> is another commonly used observer model that has demonstrated good correlation with human performance.<sup>198</sup> Human observer studies in future work will help guide the selection of observer model and appropriateness of simplifying assumptions.

Moreover, the current work employs a prediction framework for local MTF and NPS that is appropriate to a quadratic penalty function. Alternative penalties (e.g., Huber penalty or total-

variation penalty) can provide desirable edge preservation characteristics, but require a modified performance prediction framework that is beyond the scope of the current work.

To support translation to practical use, the proposed regularization design needs to be completed within a time period consistent with clinical requirements, since it adds to the already computationally intense MBIR process. The acceleration methods proposed in this chapter reduced the required number of predictions (of PSF and covariance) to as few as the number of subgrids (e.g.,  $2^3$ ). In the current work, each prediction required  $\sim 9$  hours to reach a converged solution (Matlab and CUDA calls executed on a workstation equipped with one GeForce GTX TITAN graphics card (Nvidia, Santa Clara CA)). A large reduction in computation time can be potentially achieved by replacing current linear CG algorithm with Fourier approximations.<sup>199</sup>

Statistical reconstruction that explicitly incorporates a formulation of the imaging task and optimizes the regularization approach with respect to that task presents a promising approach for “task-based image reconstruction”. One advantage of this approach involves the extraction of information from the projection data itself (rather than from the image reconstruction) to maximize imaging performance, as opposed to adjusting regularization generally (often somewhat heuristically) to encourage smoothness / sharpness in the image. Another advantage is that the approach is essentially a software-based approach (as opposed to hardware-based approaches, such as fluence modulation) so it allows one to optimize regularization (post-acquisition) for as many tasks as desired. As discussed in Sec. 3.3.3, separate “smooth” and “sharp” images can be reconstructed (optimal to  $W_{Low}$  and  $W_{High}$ , respectively) in support of accomplishing multiple tasks, or a nominal medium ( $W_{Mid}$ ) could be selected to minimize the tradeoff in all tasks considered.

The current work focuses on regularization design for detection of low-contrast ICH lesions in CBCT of the head. A natural extension of this work involves regularization design for other tasks, such as detection of high-contrast bone fractures. Moreover, the design framework can potentially be extended to consider more than one task in a single image reconstruction, for example, designing a penalty in the interior of the cranial vault for detection of acute ICH

(according to a task function " $W_{ICH}$ ") and a separate penalty in the skull for improved detection of bone fracture (according to a different task function, " $W_{Fracture}$ ") - within the *same* image. The task-based image reconstruction approach could also be applied to other anatomical sites and/or imaging techniques, such as detection of low-contrast structures in CBCT-guided interventions or the detection of suspicious nodules in low-dose CT screening of the lungs. Moreover, this work focuses on the optimization of the regularization parameter ( $\beta_j$ ), and further improvement may be gained by optimizing the directional weights and/or penalty function. Such scenarios are subjects of future work for which task-based image reconstruction could present a promising means to improve image quality in MBIR.

## **Chapter 4**

# **4. Multi-Resolution Reconstruction to Mitigate Image Truncation Effects**

## **4.1 Introduction**

### **4.1.1 Motivation**

As detailed in Chapters 2-3, novel reconstruction methods demonstrate strong promise for improved image quality in CBCT, supporting translation of the technology to applications beyond conventional limitations in contrast, noise, and radiation dose. A common practical challenge that can confound imaging performance is lateral truncation of the projection data by anatomy and/or patient support mechanisms outside the FOV of the detector. Such effects are common in clinical application. For example, in CBCT of adult body sites, such as image-guided interventions targeting the thorax, abdomen, or pelvis, the patient anatomy and/or operating table almost always exceeds the FOV. Such effects present a challenge for conventional FBP<sup>200–203</sup> and are particularly problematic for MBIR, because MBIR attempts to solve for an image estimate that best matches all of the measurements and is therefore particularly sensitive to data truncation.

For the specific application of head imaging treated in Chapters 2-3, a head holder is typically used to support the head and minimize motion during the scan – typically a U-shaped carbon-fiber support as illustrated in Fig. 4.1(a). Such a head support can be partially truncated in the projection data (even for the fairly large –  $43 \times 43 \text{ cm}^2$  – detector employed on the prototype scanner<sup>204</sup>), and the amount of truncation varies depending on the separation (e.g., a pillow)

between the head and the support. In this respect, the head holder is truncated in the axial plane. In the  $z$  direction, the head holder may or may not be longitudinally truncated at its superior extent (top of the head), though it is certainly truncated at the inferior extent (below the neck); such longitudinal truncation is not studied in this chapter. Axial truncation introduces artifacts in the reconstructed image as shown in Fig. 4.1, giving rise to nonuniformity that could hinder ICH detection.

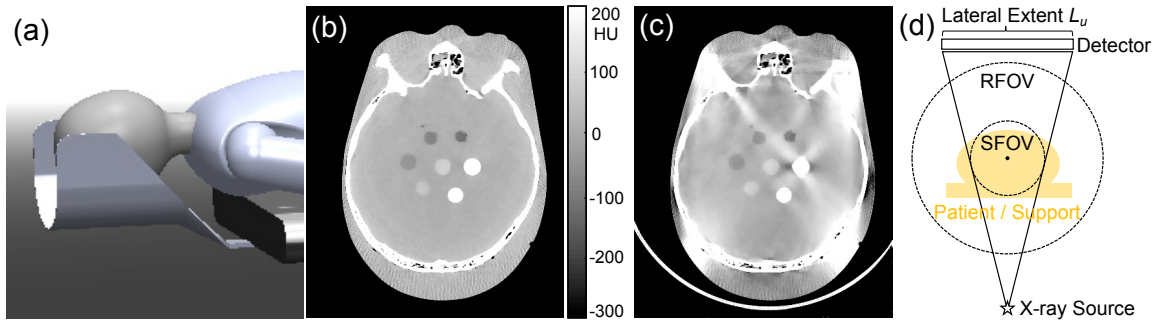


Figure 4.1: Artifacts caused by lateral truncation of the head support in CBCT of the head. (a) CAD drawing of a patient with head supported by a carbon-fiber head holder during a CBCT scan. (b) PWLS image of an anthropomorphic head phantom without a head holder. The circular inserts within the central region of the cranium span a range of contrast including that of ICH. (c) PWLS image of the same, with a U-shaped carbon-fiber head holder in place during the scan (evident beneath the posterior of the head). (d) Illustration of RFOV and SFOV for circular orbit CBCT.

A variety of strategies to mitigate truncation effects have been investigated. For example, lateral extrapolation of the projection data prior to MBIR has been proposed, including symmetric mirroring,<sup>200</sup> approximation as a scalable water cylinder,<sup>201</sup> elliptical fitting,<sup>202</sup> and using scout images to constrain anatomical boundaries.<sup>203</sup> These methods have demonstrated reduction of truncation effects to varying degrees but usually assume the main source of truncation is the patient, and the missing projection data are treated as a continuous extension of the projection of the patient at the edge of the detector. These assumptions may not hold well when the truncation is primarily due to patient support – e.g., the head holder as in Fig. 4.1.

The axial scan FOV (denoted SFOV) is distinguished from the axial reconstruction FOV (denoted RFOV) as illustrated in Fig. 4.1(d). The SFOV is the region of support within which data

are fully sampled and a complete reconstruction can be obtained. The diameter of the SFOV is defined as the lateral extent of the detector ( $L_u$ ) divided by the geometric magnification (Mag) for a circular isocentric orbit as typical in CBCT. The RFOV is a region that includes the SFOV and may be defined with arbitrary extent in the image reconstruction process. Sampling of structures outside the SFOV is incomplete and yields a “tomosynthesis-like” reconstruction in the RFOV.

Alternative to the truncation correction methods mentioned above, truncation effects can be mitigated by increasing the RFOV beyond the SFOV, thereby reducing bias within the SFOV. One advantage of this method is that it does not require additional processing of the projection data (e.g., extrapolation). However, simply increasing the RFOV increases the computational cost of MBIR. In cases where the truncated object is relatively far from the patient (e.g., an obese patient, a wide operating table, or a thick pillow inserted between the patient and the head holder), a RFOV much larger than the SFOV may be needed to mitigate truncation effects, posing a significant burden to computation time and memory.

A multi-resolution reconstruction approach is proposed in this chapter to mitigate truncation effects, which extends the RFOV without major increase in computational burden. Specifically, an image volume is defined to contain two regions: 1) a fine interior region containing the region of interest (i.e., the head) with voxel size appropriate to the diagnostic task; and 2) a coarse outer region that can be extended as much as needed to mitigate truncation, with coarser voxel size to reduce computational load. Multi-resolution MBIR has been studied previously in a 2D digital phantom by Hamelin *et al.*<sup>205</sup> and applied to ROI reconstruction of high-resolution bone morphology by Cao *et al.*<sup>206</sup> In this chapter, the multi-resolution approach is incorporated into the PWLS framework developed for high-quality CBCT of the head in Chapter 2. Accordingly, a previously reported scatter correction method<sup>149</sup> was also modified to account for the presence of the head holder. The method was evaluated in CBCT scans of an anthropomorphic head phantom with varying degrees of realistic data truncation by a carbon-fiber head support. The method is also

compatible with the task-based reconstruction framework described by Chapter 3 and is important to the translation of the head scanner prototype to clinical studies.

#### **4.1.2 Acknowledgements and unique contributions**

The methods and results reported below were reported in conference proceedings and journal articles as follows:

- (1) H. Dang, J. W. Stayman, A. Sisniega, W. Zbijewski, J. Xu, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, and J. H. Siewerdsen, "Multi-resolution statistical image reconstruction for mitigation of truncation effects: application to cone-beam CT of the head," *Physics in Medicine and Biology*, 62 (2), 539-559 (2017).

with permission from the publisher for reproduction of content in this dissertation. The author's primary contributions in this work were as follows: development and implementation of the reported algorithm; physical experimentation; and quantitative evaluation of the reported algorithm. The author gratefully acknowledges the contributions of coauthors, including: Dr. Wojciech Zbijewski for his expertise and advice on multi-resolution statistical reconstruction; Dr. Alejandro Sisniega for providing code and advice for scatter correction; and Dr. Nafi Aygun for providing the head holder used in physical experiments. The author also would like to acknowledge Mr. Alex Martin for assistance in the measurement of radiation dose and Mr. Qian Cao for valuable discussion on multi-resolution statistical reconstruction. The work was supported by research collaboration with Carestream Health (Rochester NY) and by the National Institute of Health Grant R01-EB-018896.

## 4.2 Methods

### 4.2.1 Multi-resolution PWLS for high-quality head imaging

Conventional PWLS methods<sup>139</sup> usually model an image volume  $\mu$  as a 3D region containing voxels with a fixed voxel size. In this chapter, an image volume  $\mu$  is modeled as a combination of an inner 3D rectangular region with a fine voxel size (referred to as the “fine region” or  $\mu_F$ ) and an outer 3D rectangular shell with a coarser voxel size (referred to as the “coarse region” or  $\mu_C$ ). Figure 4.2 illustrates the two regions in the multi-resolution method in imaging of the head. The fine region is defined to cover the SFOV of the CBCT system, while the coarse region is defined to cover objects that are outside the SFOV and subject to truncation. The combination of both regions defines the RFOV. For CBCT of the head, the anatomy is entirely within the SFOV (i.e., the fine region), while the head holder spans the fine and/or coarse regions of the RFOV. The resulting boundary between the fine and coarse regions is outside the cranium (in air, presumably not of diagnostic interest), so downsampling / upsampling voxels in the other region is not considered when calculating neighboring voxel differences in the subsequent image reconstruction. In the current work, an implementation specifically with two voxel sizes (coarse and fine) is investigated, but the term “multi-resolution” (c.f., “dual resolution”) is used for consistency with previous work<sup>207,208</sup> and for generality in anticipation of future work in which voxel size is more continuously varied from a fine value in the SFOV to progressively coarser values outside the SFOV.

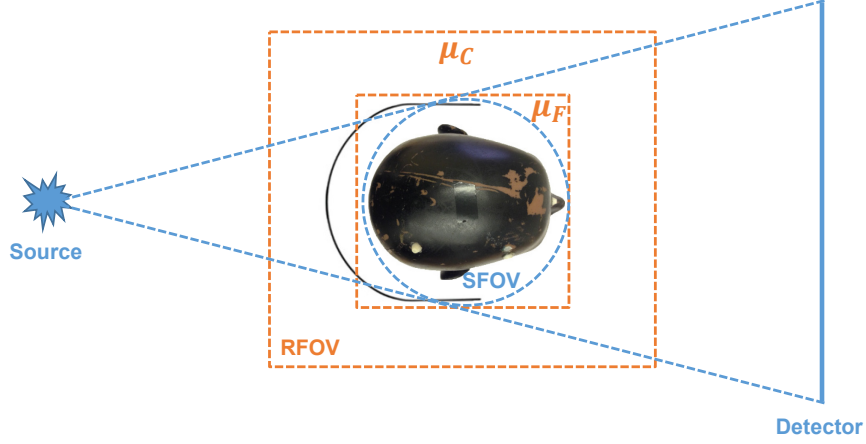


Figure 4.2: Illustration of fine and coarse regions in multi-resolution reconstruction (only x-y plane shown here). The dashed circle denotes the SFOV. The fine region ( $\mu_F$ ) is a 3D rectangle that contains the head, and the coarse region ( $\mu_C$ ) is the space outside the fine region that contains the head holder (depicted as a black, U-shaped arc posterior to the head). The volume encompassing both the fine and coarse regions is the RFOV.

Following Cao *et al.*,<sup>206</sup> one can write the forward model for multi-resolution PWLS reconstruction as follows, assuming independent measurements:

$$\bar{y} = \mathbf{D}(g) \exp(-\tilde{\mathbf{A}}\mu) = \mathbf{D}(g) \exp\left(-\begin{bmatrix} \mathbf{A}_F & \mathbf{A}_C \end{bmatrix} \begin{bmatrix} \mu_F \\ \mu_C \end{bmatrix}\right) \quad (4.1)$$

where the mean measurements are modeled by  $\bar{y}$  (a  $N_y \times 1$  vector),  $g$  is a  $N_y \times 1$  vector of measurement-dependent gains, and  $\mathbf{D}(\cdot)$  is an operator that places a vector on the main diagonal of a matrix. The notation  $\tilde{\mathbf{A}}$  denotes a system matrix representing the linear projection operation (and  $\tilde{\mathbf{A}}^T$  denotes the matched backprojection operation), which consists a  $N_y \times N_{\mu_F}$  system matrix  $\mathbf{A}_F$  for the fine region  $\mu_F$  and a  $N_y \times N_{\mu_C}$  system matrix  $\mathbf{A}_C$  for the coarse region  $\mu_C$ . The resulting line integral estimate  $\tilde{\mathbf{A}}\mu$  is thus a sum of the line integral estimate from the fine region (i.e.,  $\mathbf{A}_F\mu_F$ ) and that from the coarse region (i.e.,  $\mathbf{A}_C\mu_C$ ).

The objective function for multi-resolution PWLS reconstruction can be written:

$$\hat{\mu} = \arg \min_{\mu \geq 0} \frac{1}{2} \|\tilde{\mathbf{A}}\mu - l\|_{\mathbf{w}}^2 + \beta_F R_F(\mu) + \beta_C R_C(\mu) \quad (4.2)$$

where  $l$  denotes a vector of line integrals, and  $\mathbf{W}$  is a diagonal weighting matrix with the  $i^{th}$  diagonal element  $\mathbf{W}_i$  representing the fidelity of the  $i^{th}$  measurement. The terms  $R_F$  ( $R_C$ ) and  $\beta_F$  ( $\beta_C$ ) are the regularization term and regularization parameter for the fine region (the coarse region).

The two regularization terms enforce image smoothness in the fine and coarse regions respectively, which can be defined as:

$$R_F(\mu) = \frac{1}{2} \sum_{k \in K_F} H([\Psi_F \mu_F]_k) \quad (4.3)$$

$$R_C(\mu) = \frac{1}{2} \sum_{k \in K_C} H([\Psi_C \mu_C]_k) \quad (4.4)$$

where  $K_F$  ( $K_C$ ) denotes the number of neighboring voxels in the fine region (coarse region),  $\Psi_F$  ( $\Psi_C$ ) is an operator that computes first-order neighborhood differences in the fine region (coarse region), and  $H(\cdot)$  is Huber penalty function<sup>131</sup> which is quadratic within a neighborhood of  $[-\delta, \delta]$  and linear for larger differences as in the work by Wang *et al.*<sup>93</sup> Separate regularization terms for the fine and coarse regions allow independent control of the regularization strength. Calculation of neighborhood differences for voxels near the boundary between fine and coarse regions downsamples (or upsamples) neighboring voxels in the other region. This downsampling / upsampling operation is especially important when the boundary contains anatomy of interest (e.g., bone morphology in the work by Cao *et al.*<sup>206</sup>); however, in the scenario considered here, the boundary is outside the cranium (in air), so downsampling / upsampling at the boundary was not considered in the current work.

In the data fidelity term of the multi-resolution PWLS objective, the line integrals in  $l$  are typically derived from raw measurements  $y$  through a number of steps. Such steps include a log transformation to convert from the measurement domain to the line integral domain, and in many situations, also include correction for artifacts and/or processing to reduce noise in the measurements.<sup>159</sup> Such steps can potentially lead to changes in the noise characteristics of the

measurements, which need to be accommodated into the PWLS weighting terms. The work in Chapter 2 has modeled the processing of the measured data as a generic function  $f$  as:

$$l_i = f(y_i) \quad (4.5)$$

and derived the variance following data processing using first-order Taylor expansion of  $f$ :

$$\text{var}(l_i) \approx [\dot{f}(\bar{y}_i)]^2 \text{var}(y_i) \quad (4.6)$$

where  $\dot{f}$  denotes the derivative. In CBCT of the head, scatter and beam hardening corrections represent the dominant corrections in the artifact correction framework corrections.<sup>149</sup> The function  $f$  in this case thereby corresponds to scatter correction in the measurement domain, followed by log transformation, and then beam hardening correction in the line integral domain. The variance following this particular function  $f$  using Eq. (4.6) implies modification of the statistical weights as in Chapter 2:

$$\mathbf{W}_i = \frac{1}{\text{var}(l_i)} \approx y_i \cdot \left( \frac{y_i - \bar{S}_i}{y_i} \right)^2 \cdot \frac{1}{\eta_w(y_i - \bar{S}_i)} \cdot \frac{1}{\eta_b(y_i - \bar{S}_i)} \quad (4.7)$$

The term  $y_i$  corresponds to the weights used in conventional PWLS methods that model data processing simply as a log transformation.<sup>139</sup> The term  $((y_i - \bar{S}_i)/y_i)^2$  corresponds to the variance changes following scatter correction, where  $\bar{S}_i$  denotes the mean scatter for the  $i^{\text{th}}$  measurement. For the Joseph-Spital beam-hardening correction method,<sup>157</sup> the terms  $1/\eta_w(y_i - \bar{S}_i)$  and  $1/\eta_b(y_i - \bar{S}_i)$  correspond to the variance changes following water correction and bone correction, respectively, defined in Eq. (2.16b) in Chapter 2. The statistical weights in Eq. (4.7) were used in all PWLS reconstructions in this chapter.

Table 4.1: Pseudocode for solving the multi-resolution PWLS reconstruction using OS-SQS.

---

```

Precompute optimal curvatures  $c_i(\hat{l}_i) = [\mathbf{W}_i]_+$   $\hat{l}_i \geq 0$ 

Precompute approximate  $d$  term for the fine region  $d_{F_j} = \sum_{i=1}^{N_y} a_{F_{ij}} \gamma_i c_i(\hat{l}_i)$ 

Precompute approximate  $d$  term for the coarse region  $d_{C_j} = \sum_{i=1}^{N_y} a_{C_{ij}} \gamma_i c_i(\hat{l}_i)$ 

for each iteration  $n = 1, \dots, n_{iter}$ 
  for each subset  $m = 1, \dots, M$ 
    
$$\hat{l}_i = \sum_{j=1}^{N_{\mu_F}} a_{F_{ij}} \hat{\mu}_{F_j} + \sum_{j=1}^{N_{\mu_C}} a_{C_{ij}} \hat{\mu}_{C_j} \quad \forall i \in S_m$$

    
$$\dot{h}_i = w_i (l_i - \hat{l}_i) \quad \forall i \in S_m$$

    for voxel in the fine region  $j = 1, \dots, N_{\mu_F}$ 
      
$$\dot{L}_{F_j} = M \sum_{i \in S_m} a_{F_{ij}} \dot{h}_i, \quad \hat{\mu}_{F_j} = \left[ \hat{\mu}_{F_j} + \frac{\dot{L}_{F_j} - \beta_F \sum_{k=1}^{K_F} [\Psi_F]_{kj} \dot{H}([\Psi_F \hat{\mu}_F]_k)}{d_{F_j} + 2\beta_F \sum_{k=1}^{K_F} [\Psi_F]_{kj}^2 \omega([\Psi_F \hat{\mu}_F]_k)} \right]_+$$

    end
    for voxel in the coarse region  $j = 1, \dots, N_{\mu_C}$ 
      
$$\dot{L}_{C_j} = M \sum_{i \in S_m} a_{C_{ij}} \dot{h}_i, \quad \hat{\mu}_{C_j} = \left[ \hat{\mu}_{C_j} + \frac{\dot{L}_{C_j} - \beta_C \sum_{k=1}^{K_C} [\Psi_C]_{kj} \dot{H}([\Psi_C \hat{\mu}_C]_k)}{d_{C_j} + 2\beta_C \sum_{k=1}^{K_C} [\Psi_C]_{kj}^2 \omega([\Psi_C \hat{\mu}_C]_k)} \right]_+$$

    end
  end
end

```

---

## 4.2.2 Optimization approach for multi-resolution PWLS

The multi-resolution PWLS objective in Eq. (4.2) was solved using the OS-SQS algorithm.<sup>163</sup> The OS-SQS algorithm facilitates fast convergence not only via ordered subsets (nominally 10 subsets from 360 projection data; see below) but also via parallelizable image updates allowing parallel implementation on GPU. The work in Chapter 2 adapted OS-SQS to the single-resolution PWLS objective with modified statistical weights. For the multi-resolution case, for every subset of projections in every iteration, the image update was computed and applied to the fine and coarse

regions separately. Moreover, since the optimal curvature  $c_i$  of the data fidelity term is constant in PWLS objective, the term  $d$  in the image update can be precomputed ( $d_{F_j}$  and  $d_{C_j}$  for the fine and coarse regions, respectively).

Table 4.1 shows pseudocode for the OS-SQS solution of the multi-resolution PWLS objective. The pseudocode is similar to that in the work by Cao *et al.*,<sup>208</sup> updated with respect to notation and detector pixel model. The notation  $[\cdot]_+$  denotes the nonnegativity constraint,  $\gamma_i$  is the  $i^{\text{th}}$  projection of an image of all ones,  $a_{F_{ij}}$  and  $a_{C_{ij}}$  are the  $(i, j)^{\text{th}}$  element of the matrix  $\mathbf{A}_F$  and  $\mathbf{A}_C$  respectively,  $n_{iter}$  is the maximum number of iterations,  $M$  is the number of subsets,  $\hat{l}_i$  is the sum of the projection of the current image estimate  $\hat{\mu}_F$  and  $\hat{\mu}_C$ , and  $S_m$  denotes all the projections in the  $m^{\text{th}}$  subset. In the regularization part, taking the fine region as an example,  $K_F$  is the number of neighboring voxels in the fine region, and  $\dot{H}$  and  $\omega$  are the gradients and curvatures of the Huber penalty function  $H$ , respectively. While the pseudocode in the work by Cao *et al.*<sup>208</sup> used a small detector pixel size for the projection of a high-resolution region-of-interest and a large detector pixel size for the rest of the projection data, the pseudocode here used a single detector pixel size.

### 4.2.3 Experimental studies

The method was tested in phantom experiments performed on the CBCT test-bench shown in Fig. 4.3(a). The bench includes an x-ray source (RAD13, Dunlee, Aurora IL) and flat-panel detector (PaxScan 4343R, Varian, Palo Alto CA) in geometry equivalent to that of the prototype head scanner<sup>183</sup>: 550 mm SAD and 1000 mm SDD. Scans were acquired at 90 kVp, 0.8 mAs per projection, with 360 projections ( $1^\circ$  angular steps), and a  $0.556 \times 0.556 \text{ mm}^2$  pixel size (after  $2 \times 2$  binning). The radiation dose was measured using a Farmer chamber in an extended length CTDI phantom of 16 mm diameter, weighting the central ( $D_o$ ) and mean peripheral ( $D_p$ ) dose according to  $D_w = (1/3)D_o + (2/3)D_p$ . The dose measured with no head holder in place was 26.8 mGy. Adding

the carbon-fiber head holder reduced the dose slightly to 25.8 mGy, which was independent of Locations 1, 2, 3 (Fig. 4.3b) within 1%. This dose is comparable to that for scan protocols used in clinical studies using the prototype head scanner, 22.8 mGy.<sup>204</sup> The anthropomorphic phantom and head holder emulated a typical clinical setup in which the head was fully covered by the ( $23.7 \times 23.7 \times 23.7 \text{ cm}^3$ ) SFOV, but the head holder was truncated to varying extent. The head phantom (The Phantom Laboratory, Greenwich NY) included a natural skull and tissue-equivalent plastic (Rando™). The head holder (Siemens AG, Forchheim, Germany) was a carbon fiber ( $\sim 150 \text{ HU}$ ) unit identical to that used on routine head CT exams. The phantom was scanned with the head holder placed at three locations as illustrated in Fig. 4.3, increasing in anterior-posterior distance from the head in increments of 2.54 cm in a manner that emulated a broad range of clinically realistic setup (e.g., varying amount of padding beneath the head). A scan was also acquired without the head holder to provide a truncation-free dataset.

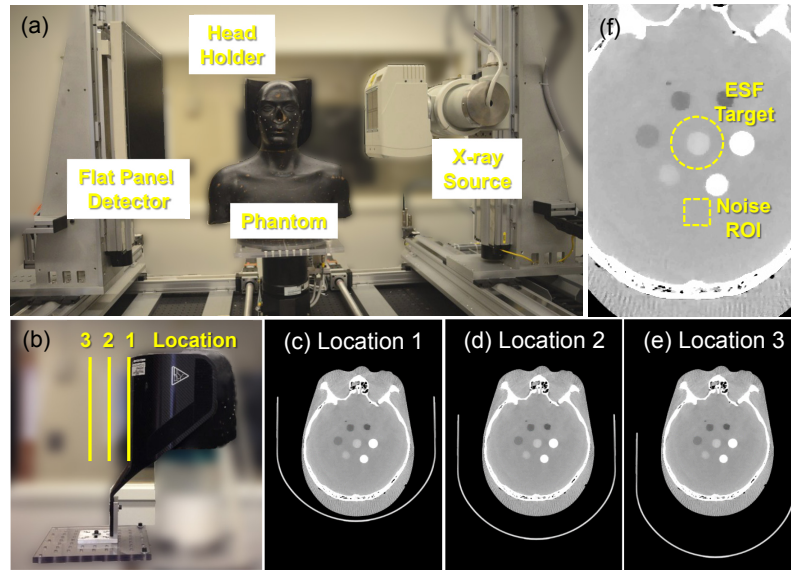


Figure 4.3: Experimental setup. (a) Photograph of the CBCT test-bench, head phantom, and head holder. (b) Illustration of three locations at which the head holder was positioned during the experiments. (c-e) Axial images superimposed with a representation of the head holder at each location. (f) Axial image illustrating structures and ROI used for image quality assessment. The central circular insert was used to compute spatial resolution (edge spread function) and contrast, and the nearby rectangular ROI was used to compute noise.

All projection data were first offset-corrected and gain-normalized by mean dark and flood field calibrations. Scatter correction involved a fast Monte Carlo method integrated with beam hardening correction using the Joseph and Spital approach. Previous work<sup>149</sup> validated the scatter and beam hardening correction without a head holder. A head holder model was added to the Monte Carlo scatter simulation to estimate the scatter from the head holder in addition to the scatter from the head. The head holder model (3D map of attenuation coefficient) was obtained from a separate CT scan of the head holder using a diagnostic CT scanner (SOMATOM Definition, Siemens Healthineers, Erlangen, Germany) with a SFOV sufficient to cover the entire head holder (i.e., without truncation). This separate scan yielded an accurate attenuation map of the head holder. In the Monte Carlo scatter simulation, the head holder was added to the system geometry based on its position as evident in the projection data. In the current work, the position of the head holder model was manually adjusted for each scan, but was subsequently automated by detecting the long edges of the holder in the scan data and computing a rigid 3D-2D registration.

Projection data were reconstructed using both conventional single-resolution PWLS and the proposed multi-resolution PWLS method. Both methods used matched separable footprint projectors and backprojectors<sup>98</sup> and 10 ordered subsets. A total of 50 iterations was found sufficient for convergence for both PWLS methods. The voxel size for single-resolution PWLS was  $0.5 \times 0.5 \times 0.5 \text{ mm}^3$ . For multi-resolution PWLS, the voxel size for the fine region was also  $0.5 \times 0.5 \times 0.5 \text{ mm}^3$ , and the voxel size for the coarse region was varied as described in the next section. A RFOV of  $400 \times 480 \times 480$  voxels (at isotropic 0.5 mm voxel size) was sufficient to cover the SFOV and was defined as the *basic* RFOV. For multi-resolution PWLS, the fine region was set to the basic RFOV, and the coarse region was varied as described in the next section (equivalent to varying the relative ratio of areas between the fine and coarse regions). This study investigates how extension of the coarse region of reconstruction outside the head (and the head holder) allows more accurate reconstruction of attenuation coefficient within the SFOV, essentially distributing bias from axial

truncation outside regions of interest. In the current work, the boundary between the fine and coarse regions is a fixed value determined by the system geometry (simply equal to the SFOV) and is not a parameter that needs to be manually defined. The water attenuation coefficient was  $0.0216 \text{ mm}^{-1}$ , and the Huber parameter  $\delta$  was set to  $10^{-4} \text{ mm}^{-1}$ , which enforced a degree of edge-preservation for features such as ICH and ventricles without causing an overly patchy appearance to the images, as shown in Chapter 2.

#### 4.2.4 Multi-resolution PWLS: parameter selection

Key parameters affecting the performance of multi-resolution PWLS were investigated, including regularization strength in the fine and coarse regions ( $\beta_F$  and  $\beta_C$ ), the voxel size in fine and coarse regions (related by the downsampling factor), and the size of the RFOV. For example, previous work<sup>206</sup> in extremity orthopaedic imaging showed that using a coarse region voxel size four times larger than the fine region voxel size yielded accurate ROI reconstruction.

(1) Regularization parameter. The parameter  $\beta_F$  controls the noise-resolution tradeoff in the fine region in a similar manner to  $\beta$  in single-resolution PWLS.<sup>93</sup> The parameter  $\beta_C$ , however, affects the fine region indirectly, and its effect on image quality was investigated as a function of downsampling factor, RFOV, and location of the head holder.

(2) Downsampling factor (DS). The ratio of the voxel size in the coarse region to that in the fine region defined the DS, which is expected to control the amount of speedup in multi-resolution PWLS. In the studies presented below, multi-resolution PWLS reconstructions were performed with the fine voxel size fixed at 0.5 mm, and DS was varied from 1 to 40.

(3) Reconstruction field-of-view. Extending the RFOV is expected to reduce truncation effects but increase reconstruction time. In the work reported below, multi-resolution PWLS images were reconstructed for RFOV ranging from the basic SFOV to a much larger RFOV, and the impact on image quality and reconstruction time were evaluated.

### 4.2.5 Imaging performance and computational complexity

The accuracy of image reconstruction was defined as the root mean square difference (RMSD) from a “truth” image, restricted to a region of the image within the cranium (i.e., in the brain). The “truth” image was defined as a single-resolution PWLS image reconstructed from the “no-holder” dataset (i.e., free of truncation effects). Spatial resolution was also assessed as in the work by Wang *et al.*<sup>93</sup> in terms of the width  $\varepsilon$  (mm) of the ESF of a low-contrast sphere within the brain [see Fig. 4.3(f)]. Contrast and CNR were evaluated with respect to a 50 HU sphere and nearby uniform ROI [see Fig. 4.3(f)].

The computational complexity of both single-resolution and multi-resolution PWLS methods are primarily determined by the total number of projection operations (including forward projection and backprojection). For single-resolution PWLS, one iteration of OS-SQS algorithm as shown in Table 2.1 in Chapter 2 requires two projection operations (one forward projection and one backprojection) for the entire RFOV, which can be written as:

$$T_{single} = (2T_{RFOV}) \cdot M \quad (4.8)$$

where  $T_{RFOV}$  denotes the time for one projection operation (for one forward projection and one backprojection) for the entire RFOV, and  $M$  is the number of subsets. For multi-resolution PWLS, one iteration of the OS-SQS algorithm requires two projection operations for both the fine and coarse regions, giving:

$$T_{multi} = (2T_F + 2T_C) \cdot M \quad (4.9)$$

where  $T_F$  and  $T_C$  denote the time for one projection operation for the fine and coarse region, respectively. Assuming the same RFOV, multi-resolution PWLS is expected to require less computation time than single-resolution PWLS, since projection operations at the fine voxel size are performed only for the fine region ( $2T_F$ ) for multi-resolution PWLS, but are performed for the entire RFOV ( $2T_{RFOV}$ ) for single-resolution PWLS. Although multi-resolution PWLS requires two

additional projection operations for the coarse region, the time associated with these two operations ( $2T_C$ ) is expected to be small.

Both PWLS methods were implemented in Matlab (The Mathworks, Natick MA), with projection operations executed on GPU using CUDA-based libraries. All image reconstructions were performed on a workstation equipped with a GeForce GTX TITAN (Nvidia, Santa Clara CA) graphics card.

## 4.3 Results

### 4.3.1 Scatter correction with a head holder model

The previously developed Monte Carlo scatter correction method was modified to include a model of the head holder, with results summarized in Fig. 4.4. The head holder was truncated in the three scans at Location 1 to 3 in Fig. 4.4. The study here focuses on the evaluation of scatter artifacts (not truncation artifacts), so the results show FBP reconstructions for simplicity, which appear to be somewhat less sensitive to truncation than PWLS reconstructions. As shown in Fig. 4.4(a-c), ignoring the head holder in the scatter correction model resulted in residual artifacts – evident primarily as shading, streaks, and overall underestimation in HU as shown in Fig. 4.4(d-f). In comparison, the results in Fig. 4.4(g-i) show that including the head holder model in scatter correction yielded images with improved uniformity for all three locations of the head holder. The “truth” image (FBP image with the head holder removed during the scan) is shown in Fig. 4.4(m). Including the head holder in the scatter correction method reduced the RMSD from the “truth” image from  $15.1 \times 10^{-4} \text{ mm}^{-1}$  to  $10.2 \times 10^{-4} \text{ mm}^{-1}$  at location 1,  $13.5 \times 10^{-4} \text{ mm}^{-1}$  to  $9.4 \times 10^{-4} \text{ mm}^{-1}$  at location 2, and  $11.7 \times 10^{-4} \text{ mm}^{-1}$  to  $9.1 \times 10^{-4} \text{ mm}^{-1}$  at location 3. In subsequent results reported below, the head holder model was always included in the scatter correction.

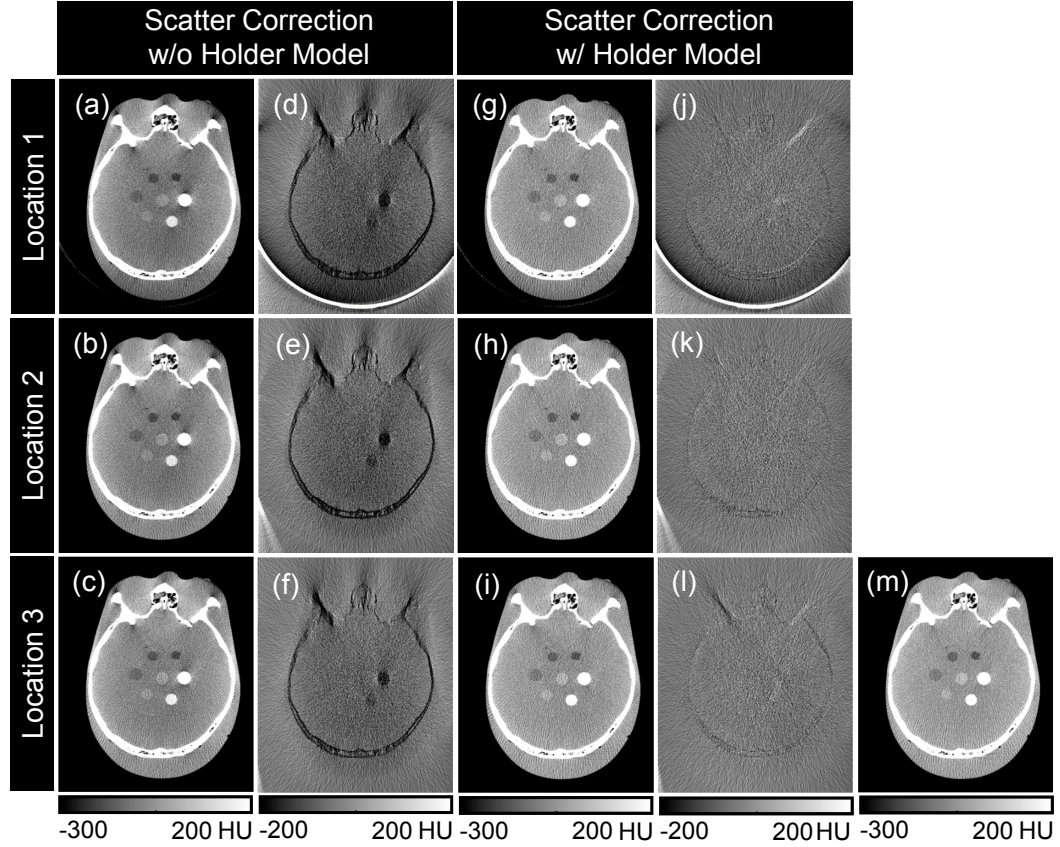


Figure 4.4: Scatter correction (a-f) without and (g-l) with the head holder included in the Monte Carlo model. (a-c) FBP reconstructions without a head holder model exhibit shading and streaks in the (d-f) difference images from (m) “truth”. Including the head holder in scatter correction reduces such residual errors as shown in (g-i) and difference images (j-l).

### 4.3.2 Single-resolution PWLS

The influence of truncation on the image quality of single-resolution PWLS reconstructions was first investigated. First, the nominal  $\beta$  value suitable for CBCT of the head was selected using the “no-holder” dataset. Figure 4.5(a) plots the ESF width and CNR measured as a function of  $\beta$ , showing a steep increase in CNR for  $\beta > 10^2$ , owing to the Huber penalty as shown in previous work.<sup>93</sup> A nominal value of  $\beta = 10^{2.4}$  was selected as balancing noise reduction and edge preservation without overly patchy image appearance, giving CNR = 25.4 and ESF width = 0.77

mm for the ROIs shown in Fig. 4.3. The resulting image [Fig. 4.5(b)] was taken as “reference” / “truth” in subsequent results.

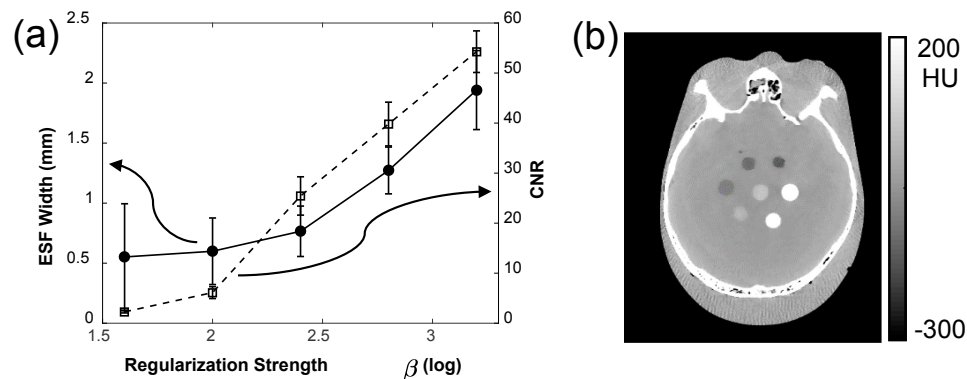


Figure 4.5: Nominal parameter selection for single-resolution PWLS. (a) ESF and CNR as a function of the regularization strength  $\beta$  (in the absence of truncation). (b) Axial slice of a single-resolution PWLS image using  $\beta = 10^{2.4}$ , exhibiting a reasonable balance between ESF and CNR and taken as the “truth” image for subsequent PWLS reconstructions.

Next, single-resolution PWLS reconstructions were computed with the head holder at three locations as shown in Fig. 4.6. Severe artifacts – including both positive and negative bias – are evident throughout the head, attributable to truncation by the head holder (not to x-ray scatter, which was corrected with the head holder model as summarized in the previous section and Fig. 4.4). The magnitude of truncation artifacts is seen to depend on the position of the head holder with respect to the head – i.e., somewhat stronger artifacts for location 1, and reduced for location 3. These artifacts appear to be associated with the truncated anterior edges of the head holder, giving rise to shading and streaks in the anterior part of the head as shown in the difference image of Fig. 4.6. The streaks are strongest for location 1 (where the edges are closest to the head) and reduced as the head holder was positioned toward the posterior of the head (location 3). Because the streaks appear to arise from the edges of the holder, they are most severe for location 1, even though location 3 involves a greater bulk of material attenuation farther from the SFOV. The RMSD from “truth” was  $11.0 \times 10^{-4} \text{ mm}^{-1}$  at location 1,  $10.2 \times 10^{-4} \text{ mm}^{-1}$  at location 2, and  $8.2 \times 10^{-4} \text{ mm}^{-1}$  at location 3.

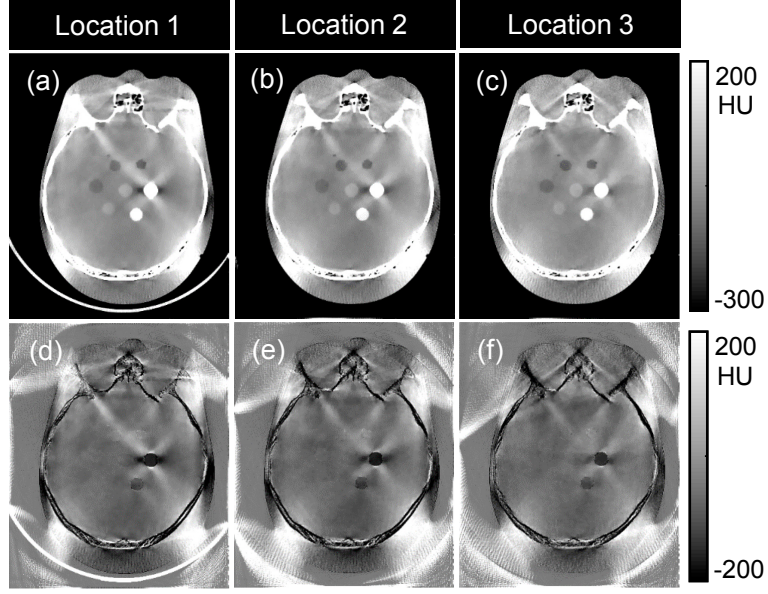


Figure 4.6: (a-c) Single-resolution PWLS with a carbon-fiber head holder positioned at three locations posterior to the head. RFOV is  $(400 \times 480 \times 480)$  voxels). (d-f) Difference images between (a-c) and the “truth” image of Fig. 4.5(b).

### 4.3.3 Multi-resolution PWLS

The sections below report a systematic evaluation of the performance of multi-resolution PWLS in the presence of truncation. First, a very large RFOV was chosen ( $1000 \times 1000 \times 1000$  voxels, with 0.5 mm isotropic voxel size), and the effects of regularization parameter  $\beta_C$  (Sec. 4.3.3.1) and downsampling factor DS (Sec. 4.3.3.2) were studied. Based on that analysis, nominal values of  $\beta_C$  and DS were selected, and the dependence of image quality on RFOV (Sec. 4.3.3.3) was investigated.

#### 4.3.3.1 Regularization parameters

Figure 4.7 plots the RMSD of multi-resolution PWLS as a function of  $\beta_C$  and DS for the three locations of the head holder. The value of  $\beta_F$  was fixed at  $10^{2.4}$  for each case. The reconstruction accuracy exhibited low dependence on  $\beta_C$  as long as  $\beta_C$  was below an upper limit in regularization

strength, but quickly reduced when  $\beta_C$  exceeded this limit. This was observed for all locations of the head holder and all DS levels. The rapid degradation in reconstruction accuracy beyond a regularization limit was also observed in previous work.<sup>206</sup> Note that for any level of DS, the regularization limit was the same for different locations of the head holder, suggesting that in practice the exact location of the head holder does not affect selection of  $\beta_C$ .

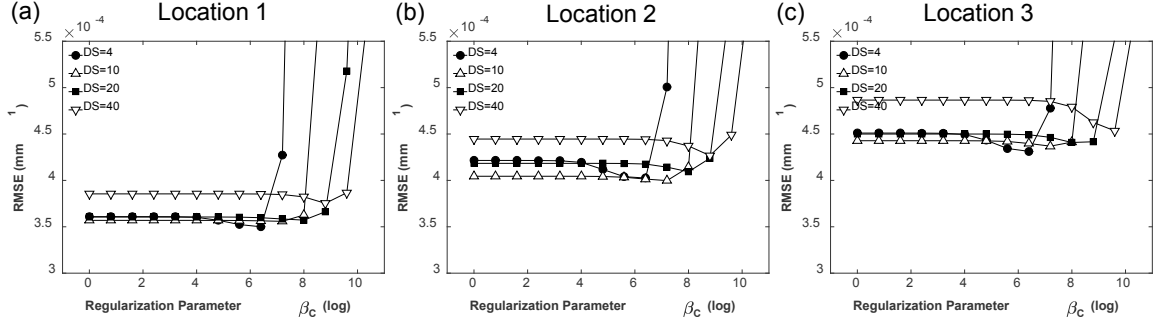


Figure 4.7: Accuracy of multi-resolution PWLS reconstructions as a function of coarse region regularization strength  $\beta_C$  and downsampling factor (DS) at three locations.

Figure 4.8 shows the fine (a-c) and coarse (d-f) regions of the multi-resolution PWLS images for three values of  $\beta_C$ . In each figure (a-c), the left half is the PWLS image, and the right half is the difference from “truth”. For the cases in Fig. 4.8, the head holder was at location 2, and the DS was set to 4. As evident in Fig. 4.8(a) and 4.8(b), PWLS exhibited fairly accurate reconstruction for a broad range of  $\beta_C$  below or near the regularization limit ( $\sim 10^{6.4}$ ), but performance degraded markedly for  $\beta_C$  above the limit (Fig. 4.8(c)). Figures 4.8(d-f) show the amount of smoothing in the coarse region (outside cyan box) for the three  $\beta_C$  values. It can be seen that the use of  $\beta_C$  beyond the regularization limit resulted in over-smoothing and low intensity error throughout the air region, which was a possible cause of the reduced accuracy in the fine region.

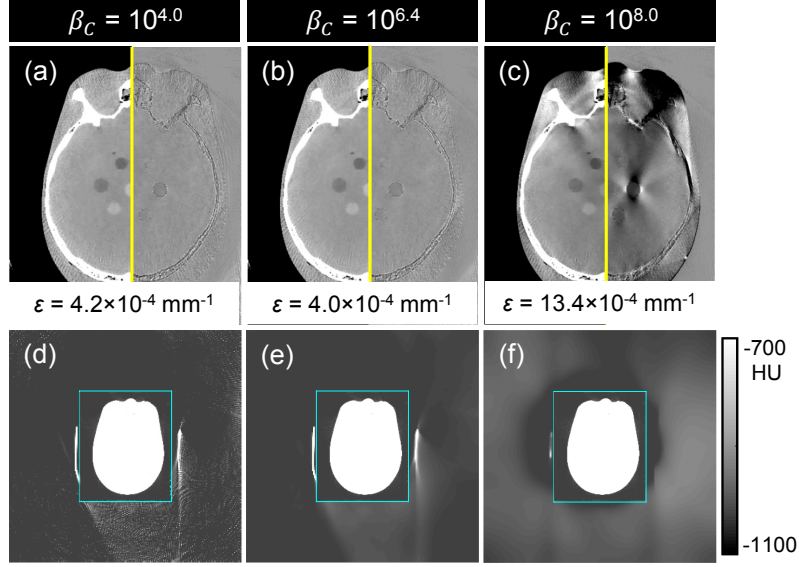


Figure 4.8: Multi-resolution PWLS reconstruction for various choices of coarse region regularization strength,  $\beta_C$ . Images (a-c) show the fine region, with the left half showing the PWLS image (grayscale window: [-300, 200] HU) and the right half showing the difference from truth (grayscale window: [-1100 -700] HU). Images (d-f) show the coarse region (outside cyan box). The head holder was at location 2, the DS was 4, and the RFOV was  $1000^3$  voxels with 0.5 mm isotropic voxel size.

#### 4.3.3.2 Downsampling factor

Figure 4.9 shows that the reconstruction accuracy was robust as DS was increased from 4 to 20, beyond which moderate degradation (e.g., at DS = 40) was observed. In the cases shown, the head holder was at location 3, and the  $\beta_C$  value was selected to achieve the lowest RMSD for each DS (fairly insensitive to selection as shown in Fig. 4.7). A small increase in streak artifacts is evident with increasing DS. Figures 4.9(e-h) show the coarse regions for various DS levels. Taking DS = 40 as an example, the coarse region involves a very large voxel size ( $20 \times 20 \times 20 \text{ mm}^3$ ), which led to coarse reconstruction of the head holder and likely led to the streaks observed in the fine region. Despite the small reduction in accuracy as DS increased, the reconstruction accuracy at all four DS levels was still much better than the single-resolution PWLS image (Fig. 4.6).

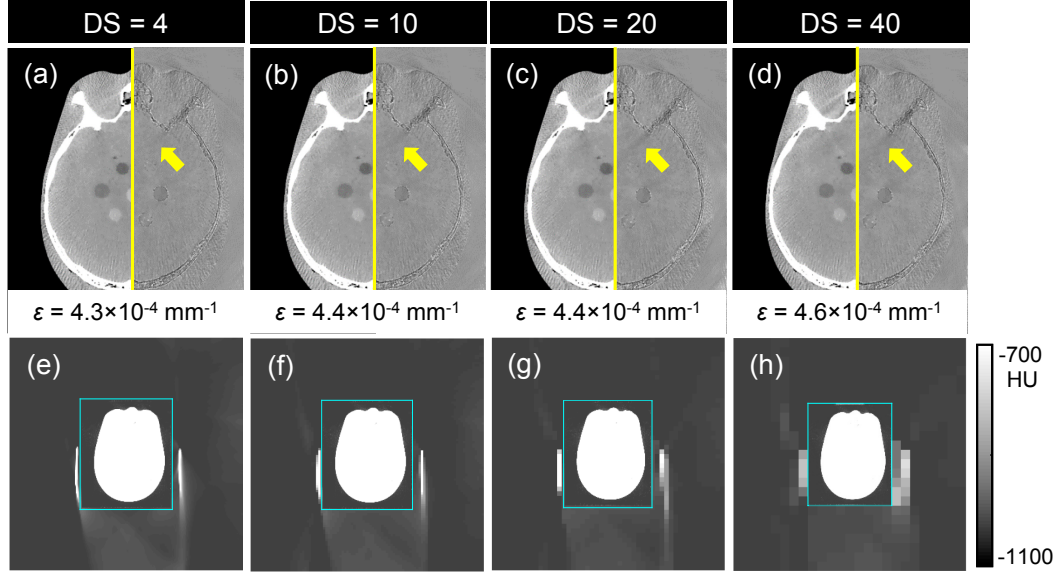


Figure 4.9: Multi-resolution PWLS reconstruction for various choices of coarse region voxel size, characterized by DS. Images (a-c) show the fine region, with the left / right presentation of the PWLS (grayscale window: [-300, 200] HU) and difference image (grayscale window: [-1100 -700] HU) as in Figure 4.8. The head holder was at location 3, the RFOV was  $1000^3$  voxels (with fine region voxel size = 0.5 mm isotropic), and  $\beta_C$  was chosen to minimize RMSD for each DS.

#### 4.3.3.3 Reconstruction field-of-view

The RFOV of multi-resolution PWLS was varied as summarized in Fig. 4.10, which effectively varies the size of the coarse region while keeping the size of the fine region equal to the basic RFOV. Specifically, the RFOV was varied from ( $400 \times 480 \times 480$  voxels) to  $\sim 10$  times as large ( $1000^3$  voxels). In the cases shown in Fig. 4.10, the head holder was at location 2, the DS was 4, and  $\beta_C$  was selected to minimize RMSD for each RFOV. The noise-resolution tradeoff in the fine region was found to exhibit small changes for varying the RFOV (which was not observed when varying  $\beta_C$  and DS), so the  $\beta_F$  value was selected to give the same noise-resolution performance for each RFOV. Images reconstructed using any of the three RFOV in Fig. 4.10 exhibited similarly high reconstruction accuracy compared to the basic RFOV (single-resolution PWLS) shown in Fig. 4.6. This again shows the benefit of increasing RFOV to mitigate truncation effects and suggests that one could freely choose a RFOV ( $> \sim 600 \times 600 \times 600$  voxels) to mitigate truncation effects.

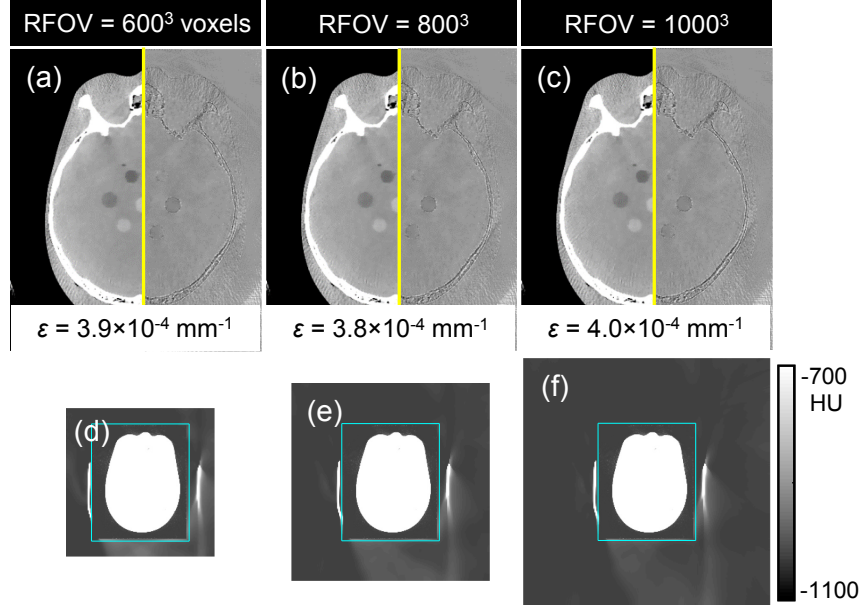


Figure 4.10: Multi-resolution PWLS reconstruction for various choices of RFOV. Images (a-c) show the fine region, with the left / right presentation of the PWLS (grayscale window: [-300, 200] HU) and difference image (grayscale window: [-1100 -700] HU) as in Fig. 4.8 and 4.9. The head holder was at location 2, the DS was set to 4, and  $\beta_F$  was adjusted slightly to maintain constant noise-resolution performance: (a)  $\beta_F = 10^{2.35}$ , (b)  $\beta_F = 10^{2.38}$ , and (c)  $\beta_F = 10^{2.40}$ . The parameter  $\beta_C$  was chosen to minimize RMSD for each RFOV (after selecting  $\beta_F$ ).

#### 4.3.4 Computation time

The computation time between single-resolution and multi-resolution PWLS reconstruction is summarized in Fig. 4.11 in terms of the measured time per projection operation (averaged over one forward projection and one backprojection) and memory usage as a function of RFOV (for isotropic voxel size of 0.5 mm). As shown in Fig. 4.11(a), both time and memory usage increase dramatically if the RFOV increases from the basic RFOV (denoted by the dashed line), suggesting that simply increasing the RFOV in single-resolution PWLS is computationally expensive and likely impractical. Figure 4.11(b) plots the measured time per iteration as a function of RFOV for various PWLS reconstruction methods. Single-resolution PWLS exhibited a steep increase in time per iteration with larger RFOV, consistent with the steep increase in Fig. 4.11(a). By comparison, the time per iteration was much reduced for the multi-resolution approach at DS = 2 and was further

reduced at  $DS = 4$ . The time per iteration was reduced from the single-resolution approach ( $DS = 4$ ) by 40.7% for a RFOV of  $600^3$  voxels, 83.0% for a RFOV of  $800^3$  voxels, and over 95% for a RFOV of  $1000^3$  voxels. Moreover, the time per iteration became almost independent of RFOV when  $DS$  increased to 4 or larger in the multi-resolution approach. This suggests that one could increase the RFOV as much as needed to mitigate truncation effects in the multi-resolution approach without corresponding increase in computational complexity. The results clearly demonstrate the advantages of multi-resolution reconstruction, especially in situations where a large RFOV is needed. Because an increase in  $DS$  larger than 4 gradually reduced the reconstruction accuracy as shown in Sec. 4.3.3.2 (but does not correspondingly reduce the computation time),  $DS = 4$  was selected as the nominal / optimal  $DS$ .

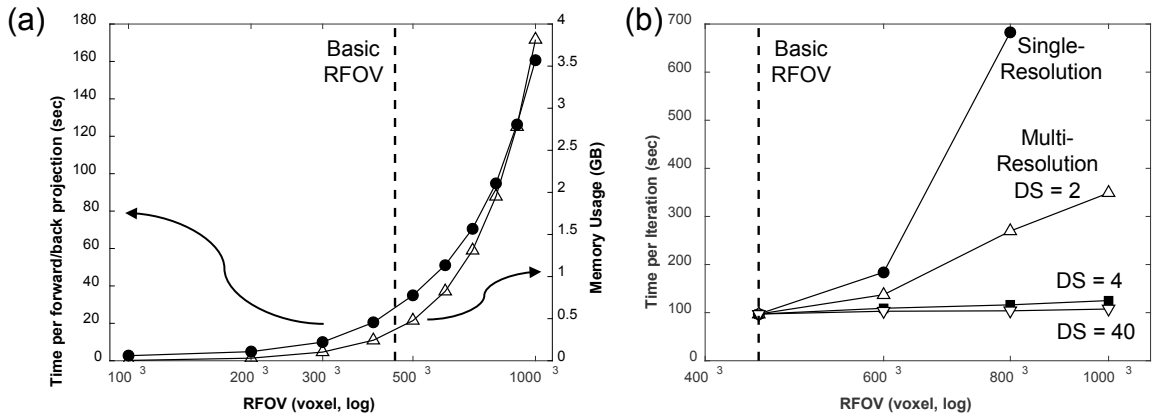


Figure 4.11: (a) Measured computation time (averaged over one forward projection and one backprojection) and measured memory usage as a function of reconstruction field-of-view (RFOV). (b) Time per iteration (i.e., for all subsets) as a function of RFOV for single-resolution and multi-resolution PWLS reconstruction at different  $DS$  levels. Substantial speedup can be seen compared to the single-resolution approach at  $DS = 2$  and to multi-resolution approach at  $DS = 4$ .

### 4.3.5 Comparison of reconstruction methods

Figure 4.12 shows a single-resolution PWLS reconstruction using (a-b) the basic RFOV and (c-d) an extended RFOV in comparison to (e-f) multi-resolution PWLS reconstruction (with the same extended RFOV). In this case, the head holder was at location 2. The multi-resolution PWLS

reconstruction parameters were: (1)  $\beta_C = 10^{6.4}$ , which is near the upper range of stable regularization identified in Sec. 4.3.3.1; (2)  $DS = 4$ , as indicated in Secs. 4.3.3.2 and 4.3.4; and (3) an extended RFOV of  $600^3$  voxels (assuming isotropic voxel size of 0.5 mm), which the results of Sec. 4.3.3.3 identify as the smallest RFOV providing good mitigation of truncation artifacts. Single-resolution PWLS using the basic RFOV exhibits severe artifacts due to truncation, whereas single-resolution PWLS with an extended RFOV substantially reduced such artifacts, but doubled computation time. Multi-resolution PWLS using the same extended RFOV exhibited visually and quantifiably similar reduction of truncation effects and only increased computation time by 12% (109 sec/iter vs. 97 sec/iter) compared to reconstruction with the basic RFOV.

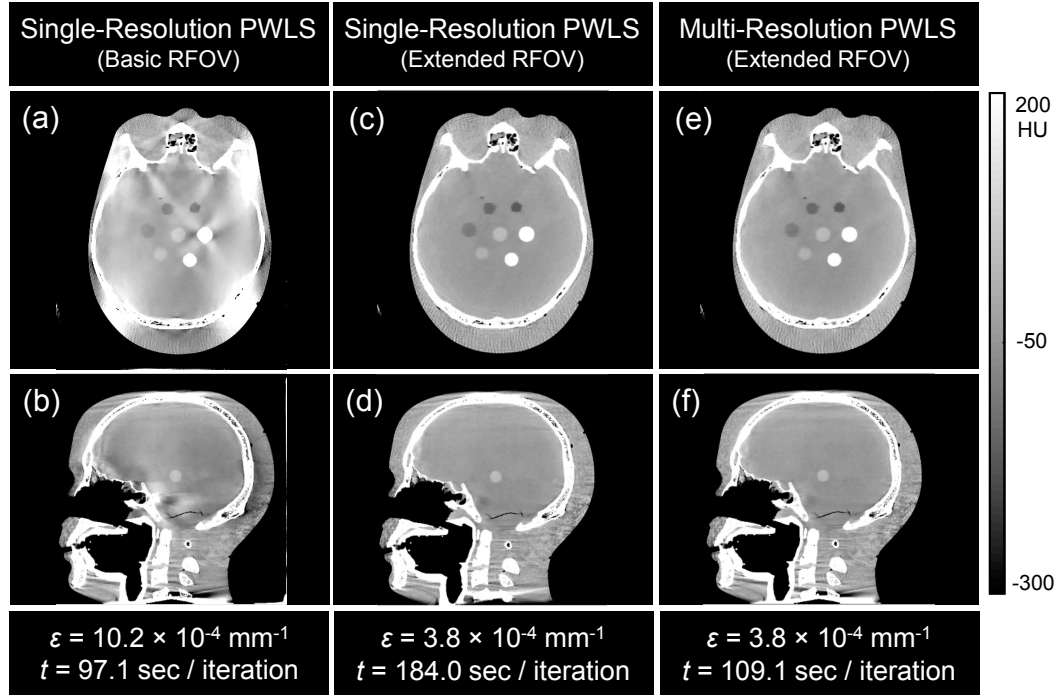


Figure 4.12: Comparison of single-resolution and multi-resolution PWLS reconstruction. (a-b) Single-resolution PWLS reconstruction using the basic RFOV ( $400 \times 480 \times 480$  voxels). (c-d) Single-resolution PWLS reconstruction using an extended RFOV ( $600^3$  voxels assuming an isotropic voxel size of 0.5 mm). (e-f) Multi-resolution PWLS reconstruction using the same extended RFOV. The symbols  $\varepsilon$  and  $t$  denote RMSD and computation time, respectively, quantifying the reduction in artifact using an extended RFOV and the benefit to computation time using the multi-resolution method.

## 4.4 Conclusions and discussion

This chapter reported a multi-resolution MBIR method to mitigate truncation effects with specific application to CBCT of the head, for which the main source of truncation is the patient support / head holder. While conventional reconstruction methods employ a fixed voxel size throughout the image, the multi-resolution method uses a fine voxel size within the untruncated region (i.e., inside the SFOV encompassing the area of interest – in this case, the head) and a coarse voxel size in the truncated region outside the SFOV (i.e., outside the area of interest). The approach was implemented in a PWLS reconstruction framework and evaluated in experiments involving a head phantom imaged on a CBCT test-bench with varying levels of truncation using a commercially available carbon-fiber head holder. The multi-resolution method demonstrated substantial mitigation of truncation effects and major reduction in computational cost compared to single-resolution reconstruction with an extended RFOV.

Investigation of the main algorithm parameters suggest that: (1) reconstruction accuracy in the fine region (the head) does not depend strongly on the regularization parameter in the coarse region  $\beta_C$  as long as the parameter is below a regularization “limit,” which in turn was found not to depend on the location of the head holder and can therefore be held fixed; (2) use of a larger voxel size in the coarse region (larger DS) reduces computational complexity but slightly reduces reconstruction accuracy, suggesting an optimal DS such that the voxel size in the coarse region was  $\sim 4$  times that in the fine region; and (3) reconstruction accuracy improved with a larger RFOV up to a certain extent ( $600^3$  voxels assuming isotropic voxel size in this work) beyond which accuracy was modestly improved. In the current work, truncation was due solely to the head holder (which varied in location but not in size or mass), and more severe truncation (i.e., greater mass of attenuation outside the SFOV) may require larger RFOV. In summary, the method presents a promising means to mitigate truncation effects in CBCT of the head and supports translation of a newly developed CBCT head scanner in point-of-care imaging applications. It is also compatible

with the method reported in Chapter 2 (modified PWLS weights to account for scatter and beam-hardening corrections – as demonstrated in the results above) and potentially with the method of Chapter 3 (selection of spatially varying regularization strength to maximize task performance), although the latter was not investigated in the current work and requires validation of the task-based prediction framework and spatially varying  $\beta_F$ .

A variety of alternative methods to managing truncation artifacts have been reported. For example, some methods treat missing projection data as a continuous extension of the projection at the edge of the detector and extrapolate the missing data before image reconstruction.<sup>200–203</sup> These methods have demonstrated reduction of truncation effects to various extents but the assumption on the continuous extension of the projection may not hold well when the truncation is primarily due to patient support. Other methods attempt to directly reconstruct a ROI inside the patient anatomy that has not been truncated during the scan.<sup>209–215</sup> For example, a widely recognized approach in ROI reconstruction is to backproject the derivative of the projection data and apply Hilbert filtering along certain lines covering the ROI.<sup>209,212</sup> These ROI reconstruction methods have demonstrated substantial reduction of truncation artifacts in the ROI, but as analytical methods, they usually do not enjoy the noise-resolution benefits exhibited by MBIR. The method proposed in this chapter allows more general treatment of the source of truncation than extrapolation-based methods and therefore can be used to manage truncation effects that do not arise from the patient (e.g., due instead to the patient support). Moreover, the proposed method is formulated within a MBIR framework, which allows the use of advanced system models and regularization techniques.

An alternative method to mitigate truncation effects in CBCT of the head is to include a model of the head holder within the image reconstruction process. For example, previous work<sup>216</sup> reported a known-component reconstruction (KCR) approach that could be extended to include the known shape of the head holder. This could yield even better agreement with the measured projection data and mitigate truncation artifacts in the image. Moreover, the multi-resolution

approach proposed above could be combined with the KCR approach to improve computational efficiency.

The multi-resolution method presents a more efficient means to recover attenuation information from truncated objects than simple extension of the RFOV. This is a particularly important consideration in MBIR, which can be sensitive to truncation effects not only in terms of artifacts and accuracy of reconstruction but also in the speed and stability of convergence. MBIR also carries a fairly high computational burden, and straightforward extension of the RFOV could lead to impractical reconstruction time. The current work focused on a particular form of data truncation encountered in CBCT of the head but offers a potential general solution for other scenarios in CT or CBCT. In C-arm CBCT for interventional imaging, for example, the patient periphery, interventional tools, and operating table are often truncated due to the limited SFOV. Moreover, in diagnostic imaging, truncation can occur for obese patients or (purposeful or inadvertent) setup of the patient off center. Such scenarios are the subject of future work, where the proposed method may offer a means to mitigate truncation effects without major increase in computational cost.

## Chapter 5

# 5. Incorporation of Prior Images in Statistical Reconstruction

## 5.1 Introduction

### 5.1.1 Prior-image-based reconstruction

As discussed in Chapter 1, there are many scenarios in medical imaging - including screening, diagnosis, and image-guided intervention - in which the patient undergoes sequential imaging studies with repeated scans over a period of time. Conventionally, each scan is treated in isolation with a full dose protocol, and the accumulated radiation dose in sequential imaging studies can be a concern to both the patient and surgical staff. Knowledge of patient-specific anatomy gained from an image previously acquired in sequential studies (referred to below as the "prior image") can potentially be leveraged in a model-based image reconstruction process, presenting increased opportunities for improved image quality and/or dose reduction.

The importance of prior images in image reconstruction has been recognized in recent years in a number of prior-image-based reconstruction (PIBR) approaches. In Prior Image Constrained Compressed Sensing (PICCS),<sup>110</sup> Chen *et al.* formed an objective function that seeks sparse differences between the reconstruction of current anatomy and a prior image, using compressed sensing concepts whereby a sparse signal can be recovered by  $L^1$  minimization under certain assumptions.<sup>217</sup> The original formulation of PICCS has also been modified with the inclusion of statistical weights<sup>218</sup> and can be applied in situations where the forward model can be transformed

into a linear relationship between the image volume and processed line integrals. In contrast, Stayman *et al.*<sup>111</sup> presented a PIBR approach termed Prior Image Registration, Penalized-Likelihood Estimation (PIRPLE), employing a MBIR framework that integrates: 1) a statistical objective function with nonlinear forward model and noise model for the unprocessed measurements; and 2) a generalized regularization term based on a prior image. This framework permits flexibility in the selection of the forward model and the noise model and does not necessitate a linearizable forward model. In addition to direct use of a patient-specific prior image in the reconstruction objective function, prior images have also been utilized indirectly.<sup>219,220</sup>

Although the use of prior images in PIBR dramatically reduces the data fidelity requirements and demonstrates good image quality under conditions of substantial downsampling and photon starvation,<sup>111,221</sup> a critical aspect of effectively using prior images is the ability to compensate for deformation between the prior image and subsequent image acquisitions. Such deformation is typically caused by patient motion between the baseline scan (i.e., the scan that forms the prior image) and the follow-up scan (i.e., the scan that acquires measurements of current anatomy). Potential sources of mismatch between scans include re-positioning of a patient and acquisition during differing physiological states (e.g., acquisitions at different phases of respiratory or cardiac motion). If these mismatches are not compensated or are incorrectly compensated, incorrect information from the misaligned prior image will be injected into the image reconstruction. Moreover, PIBR without registration cannot differentiate between *true anatomical changes* (e.g. tumor growth, bone drill-out, tissue ablation, etc.) and *changes due to motion*. Ambiguity between these two types of changes in the PIBR results would make true anatomical change difficult to recognize and may introduce false anatomical changes.

Compensation of deformation was typically not considered in the PICCS implementations for cardiac imaging<sup>110</sup> and dose reduction,<sup>218</sup> because in these applications a prior image was formed from a full set of current measurements. Nett *et al.*<sup>222</sup> adopted a staged approach using a preregistration of the prior image followed by PICCS reconstruction in the context of C-arm

interventional imaging where patient motion is commonly seen between the baseline and follow-up scans. Specifically, in this work the prior image was registered to a FBP of sparse measurements of the current anatomy using a rigid 6 degree-of-freedom (DOF) transformation. However, the 6 DOF registration may not accurately capture the non-rigid nature of organ motion, and the accuracy of the staged registration will be limited by the image quality of initial sparse FBP reconstruction. Similarly, the initial PIRPLE implementation<sup>111</sup> entailed a rigid registration of the prior, and although the registration estimate was joint (not staged) with the PL reconstruction estimate, the 6 DOF pose estimate does not resolve deformations between the baseline and follow-up scans. Other methods have been developed that deformably register a prior image acquired from a planning CT with sparse projections from a subsequent CBCT.<sup>223–225</sup> However, in these methods, the follow-up data were used to estimate patient motion but not to reconstruct new images. Deformable registration of a prior image has also been used recently for artifact correction.<sup>226</sup>

A *joint* estimation of both the deformation and image reconstruction may be used to overcome the limitations of staged registration. The idea of joint estimation has been widely studied in MBIR in many modalities.<sup>227–231</sup> For example, in cardiac gated emission computed tomography, Gilland *et al.*<sup>227</sup> designed an objective function that jointly estimated cardiac images at two different frames and the cardiac motion between the two frames, thereby solving both deformation and attenuation together using a conjugate gradient method. In PET, Fessler<sup>229</sup> used an objective function that jointly estimated a single image (at a specific time point) and a series of deformation fields (used to match motions at other time points), achieving a joint solution by alternately updating the deformation and attenuation parameters. Despite the varied use of joint estimation in MBIR, the joint approach has received less attention in PIBR.

This chapter introduces a model-based approach that incorporates a Poisson noise model and a high-quality patient-specific prior image to reconstruct images from sparse and/or noisy measurements. Deformation between the baseline and follow-up scan introduced by patient motion is estimated jointly through a cubic B-spline-based free-form deformation (FFD) model. By

extension of the PIRPLE framework, this approach is referred to as deformable Prior Image Registration, Penalized-Likelihood Estimation (dPIRPLE). First, a detailed description of the dPIRPLE algorithm and an alternating maximization strategy for solving dPIRPLE are presented. Additionally, the convergence properties and the scheduling of registration / reconstruction updates are analyzed. Finally, qualitative and quantitative comparisons of reconstruction results from dPIRPLE and other algorithms are performed under various conditions of data sparsity and exposure levels.

### **5.1.2 Acknowledgement and unique contributions**

The methods and results reported below were reported in conference proceedings and journal articles as follows:

- (1) H. Dang, A. S. Wang, Z. Zhao, M. S. Sussman, J. H. Siewerdsen, and J. W. Stayman, "Joint estimation of deformation and penalized-likelihood CT reconstruction using previously acquired images," *The 12th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, Lake Tahoe, CA, 424-427 (2013).
- (2) H. Dang, A. S. Wang, M. S. Sussman, J. H. Siewerdsen, and J. W. Stayman, "dPIRPLE: a joint estimation framework for deformable registration and penalized-likelihood CT image reconstruction using prior images," *Physics in Medicine and Biology*, 59 (17), 4799-4826 (2014).

with permission from the publisher for reproduction of content in this dissertation. The author's primary contributions in this work were as follows: development of a joint estimation framework for deformable registration and penalized-likelihood reconstruction; physical experimentation; quantitative evaluation of the reported algorithm including parameter selection, convergence properties, alternating maximization scheduling, reconstruction accuracy for different measurement sparsity and exposure levels, and capture range of the deformable registration. The

author gratefully acknowledges the contributions of coauthors, including: Dr. Adam S. Wang for providing valuable input over the course of this project; Dr. Mark S. Sussman for assistance with the cadaver specimen; and Ms. Zhe Zhao for assisting with the physical experiments. Valuable discussion with Dr. Rick Colbeth, Dr. Sungwon Yoon, Mr. Edward Shapiro at Varian Medical Systems (Palo Alto, CA) is also gratefully appreciated. The work was supported by research collaboration with Varian Medical Systems (Palo Alto, CA).

## 5.2 Methods

### 5.2.1 Penalized Likelihood Estimation

Penalized likelihood estimation (PLE) makes more efficient use of measurement data in CT by incorporating a measurement noise model and encourages desired properties in the reconstructed image by using a regularization term. The PLE objective function has been defined in Sec. 1.4.8 assuming a Poisson noise model and is given here again for reader's convenience:

$$\hat{\mu} = \arg \max_{\mu} \log L(y; \mu) - R(\mu), \quad R(\mu) = \beta_R \|\Psi_R \mu\|_{p_R}^{p_R} \quad (5.1)$$

Note that the regularization term has been expressed in vector form. The specific form of the penalty here includes the operator  $\Psi_R$ , a  $p$ -norm metric (with  $p = p_R$  and an exponent  $p_R$ ), and a scalar regularization strength  $\beta_R$ . This general form of penalty allows for varied control of image properties and has a number of different interpretations including penalties on roughness,<sup>106</sup> total variation,<sup>108</sup> and other decompositions of the image. (E.g., one may select  $\Psi_R$  to be an arbitrary sparsifying transform, particularly for  $p_R \leq 1$ , which encourages sparse solutions in the transformed domain.) In this chapter, a  $\Psi_R$  operator is selected that computes the pairwise difference between voxels in a first-order neighborhood around each voxel as in traditional roughness penalties.

Among popular approaches for solving PLE, the SQS approach<sup>232</sup> is used in this chapter that allows for highly parallelizable image updates. To satisfy the five conditions of finding a parabolic surrogate in SQS approach,<sup>232</sup> a modified  $p$ -norm<sup>111</sup> was introduced which replaced a  $\delta$ -neighborhood about the origin in the traditional  $p$ -norm with a quadratic function. This modification ensures the differentiability of the  $p$ -norm operator at the origin. Both the function values and the derivatives match at the transition point  $\pm\delta$  after the modification. Throughout this chapter, either  $p = 1$  or  $p = 2$  is used, and the modified  $p$ -norm becomes equivalent to the Huber function<sup>131</sup> or the standard  $L^2$  norm, respectively. In both cases, the modified  $p$ -norm can be easily shown to satisfy the conditions for application of SQS.

## 5.2.2 Deformable Prior Image Registration, Penalized-Likelihood

### Estimation

Prior images typically contain a great deal of patient-specific anatomical information. Thus, such images have great potential for regularization of the reconstruction problem and consequent dose reduction. One specific way to do this is to modify Eq. (5.1) with an additional penalty term (referred to as the *prior image penalty*), which encourages similarity between the estimated image and the prior image by penalizing their differences. However, one must recognize that changes between a prior image and the current patient anatomy can be introduced in two possible ways: 1) *true anatomical changes*, which are the result of disease progression or surgical interventions (e.g. tumor growth, bone drill-out); and 2) *changes due to motion*, which are caused by patient motion between scans (e.g. patient re-positioning, respiratory or cardiac motion). Thus, directly enforcing similarity to a prior image without the compensation of *changes due to motion* will limit the utility of the prior image and potentially provide incorrect information due to the misregistration. In other words, “efficient” or “complete” extraction of information from a prior image is only possible when proper registration is used to eliminate mismatches due to motion. Moreover, recognizing that the

accuracy of an initial one-time registration procedure (e.g., staged registration followed by reconstruction) is limited by registration of low-fidelity data, the registration parameters are incorporated into a *joint* image reconstruction and registration objective function - solving for both registration and reconstruction parameters. This joint approach is expected to achieve improved results since updated image estimates can help to refine registration estimates and vice versa. This approach is referred to as dPIRPLE, and its objective function can be written as

$$\begin{aligned} \{\hat{\mu}, \hat{\lambda}\} &= \arg \max_{\mu, \lambda} \Phi(\mu, \lambda) \\ &= \arg \max_{\mu, \lambda} \log L(y; \mu) - \beta_R \|\Psi_R \mu\|_{p_R}^{p_R} - \beta_P \|\Psi_P (\mu - \mathbf{W}(\lambda) \mu_P)\|_{p_P}^{p_P} \end{aligned} \quad (5.2)$$

where the last term denotes the prior image penalty with  $\mu_P$  denoting the prior image. The prior image penalty term in Eq. (5.2) has a distinct set of parameters ( $\beta_P$  and  $p_P$ ), which allows control of this penalization independent from that of the roughness penalty. Varying  $\beta_P$  controls the strength of prior image information in the reconstruction. The parameter  $p_P$  also affects the balance of prior information relative to other terms and can be freely chosen to penalize the difference image using different norms. For example,  $p_P=2$  tends to apply larger penalties for greater differences, thereby enforcing smooth differences and blending features in the prior image and current measurements. In contrast,  $p_P=1$  tends to apply relatively smaller penalties for greater differences and therefore allowing or encouraging large and sparse differences. Similar ideas can be found in compressed sensing theory that recovers sparse solutions by using an  $L^1$  norm constraint.<sup>217</sup> This norm selection is important to minimize bias in regions where change occurs between the prior and current scans. Specifically, in this study,  $p_P=1$  is used which encourages similarity between the current reconstruction and prior image data but allows for potentially large but sparse differences. With proper selection of  $\beta_P$ , this penalty is thereby expected to allow the sparse anatomical change (e.g. lung nodule) to appear in the reconstructed image and not to significantly bias the reconstructed image towards the prior data in the region of anatomical change. In this case, the operator  $\Psi_P$  can

be interpreted as a sparsifying operator. Options for this operator include an image gradient operator (e.g., finite differences on local voxels) or the identity matrix (if image changes are already sparse).

In Eq. (5.2),  $\mathbf{W}(\lambda)$  represents a deformation operator that is a function of the deformation parameters  $\lambda$ . While there exist many methods to deformably register a 3D volume, the cubic B-spline-based FFD<sup>233</sup> is chosen as the deformation model. Specifically, one can write

$$\mathbf{W}_x(\lambda) = x + \sum_{x_i \in N_x} \lambda_i \beta\left(\frac{x - x_i}{\sigma}\right) \quad (5.3)$$

where  $\beta(\cdot)$  is the tensor product of cubic B-spline functions,  $x_i$  are the control points,  $\sigma$  is the control point spacing,  $\lambda_i$  are the B-spline coefficient vectors (i.e., control point displacements), and  $N_x$  is the set of control points within the B-spline support of  $x$ . Cubic B-spline FFD has several advantageous properties as a deformation model.<sup>234</sup> First, it has low-dimensional parameterizations and local support, which reduces computational complexity and allows highly localized deformation to be modeled. Second, it provides  $C^2$  continuity at the knots, allowing gradient-based optimization approaches to be used. Third, the FFD grids can be constructed hierarchically to allow the registration to be performed using morphological pyramids, reducing the susceptibility to local minima during optimization.

The estimator in Eq. (5.2) is a general estimator, but one can identify a number of specific forms. For example, if  $\mathbf{W}(\lambda)$  in Eq. (5.2) is replaced with an identity matrix, the objective function will not consider any deformation when incorporating the prior image. This approach is referred to as Prior Image, Penalized-Likelihood Estimation (PIPLE). If  $\mathbf{W}(\lambda)$  is replaced with a rigid transformation (3 translation and 3 rotation parameters), this approach is referred to as rigid PIRPLE. The general form in which  $\mathbf{W}$  represents a FFD is termed dPIRPLE.

### 5.2.3 Strategy for solving dPIRPLE

The implicitly defined dPIRPLE approach of Eq. (5.2) requires a strategy for finding both the attenuation and the deformation parameters. One straightforward method might be to combine both types of parameters and solve by a general "off-the-shelf" gradient-based optimization approach. However, methods that are not tomography-specific or registration-specific, may be exceptionally time-consuming due to the large scale parameter space and further complicated by local minima due to the non-convexity in FFD registration. As such, an alternating maximization approach is used that (a) maximizes the dPIRPLE objective with respect to attenuation parameters with fixed registration (referred to as "image update") using a tomography-specific optimizer and (b) maximizes over registration parameters with fixed attenuation (referred to as "registration update") using a FFD-specific algorithm. This alternating approach between image and registration updates can be expected to handle the large scale parameter space and local minima better than general optimization approaches.

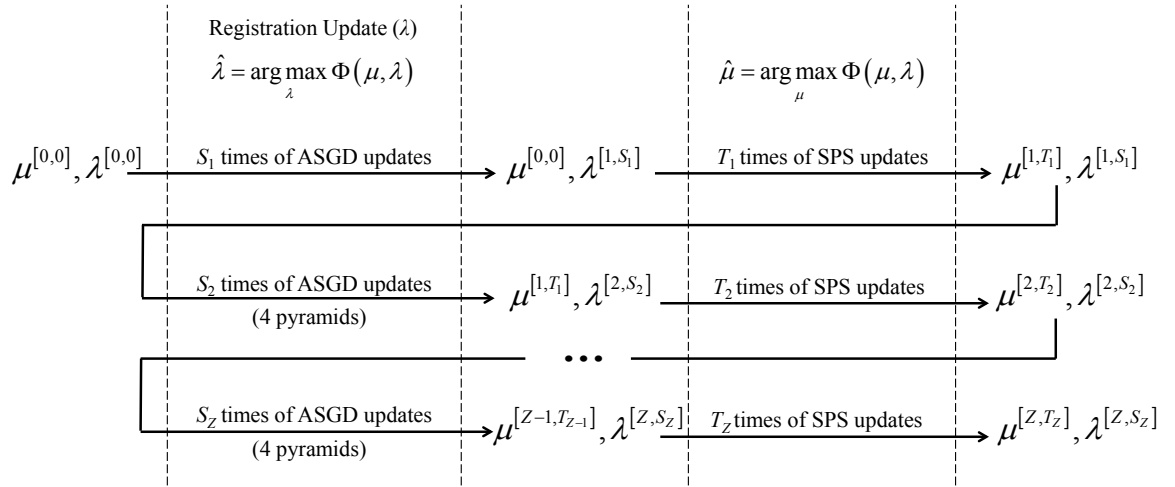


Figure 5.1: Flow chart for solving the dPIRPLE objective function using an alternating maximization approach.

The flow chart in Fig. 5.1 illustrates this alternating approach. After initializing the parameter pair with  $(\mu^{[0,0]}, \lambda^{[0,0]})$ , a number of registration updates ( $S_1$ ) are applied to  $\lambda$  in a

registration block, followed by a number of image reconstruction updates ( $T_1$ ) applied to  $\mu$  in an image block. The superscripts in the square bracket  $[i, j]$  denote the "outer loop" alternation number ( $i$ ) and the "inner loop" number ( $j$ ) of image or registration updates applied. After a first alternation,  $S_2$  registration updates are applied to  $\lambda$  in a new registration block to start another alternation. Starting the workflow with a registration block is preferred since patient motion can then be at least partially compensated before any use of the prior image.

In the image block, since the deformation parameters are fixed, the dPIRPLE objective function becomes dependent only on the attenuation parameters and therefore is equivalent to a standard PLE with a prior image penalty without registration. Since the prior image penalty shares the same structure as the image roughness penalty, it can be easily shown that the prior image penalty also satisfies the conditions of finding a parabolic surrogate in SQS approach.<sup>232</sup> Therefore, the objective function in the image block can be optimized by SQS as summarized in Table 5.1. The optimum curvature<sup>232</sup> is used for the surrogate of the likelihood term in this study.

In the registration block, since the deformation parameters only appear in the prior image penalty term, for registration updates, the dPIRPLE objective function can be reduced to only this term. The objective function can be further transformed into a minimization of image differences after applying the deformation operation, which is essentially a standard image registration problem with a modified  $p$ -norm operator as similarity metric as shown below

$$\begin{aligned}
\hat{\lambda} &= \arg \max_{\lambda} \Phi(\mu, \lambda) \\
&= \arg \max_{\lambda} -\beta_p \left\| \Psi_p(\mu - \mathbf{W}(\lambda) \mu_p) \right\|_{p_p}^{p_p} \\
&= \arg \min_{\lambda} \left\| \Psi_p(\mu - \mathbf{W}(\lambda) \mu_p) \right\|_{p_p}^{p_p}
\end{aligned} \tag{5.4}$$

For example, when  $p_p=2$ , the objective function is equivalent to the common Sum of Squared Differences (SSD) similarity metric for registration. Therefore, the objective function in the registration block can be solved using any number of existing methods for deformable registration

with minor modifications. Moreover, morphological pyramids can be used as part of the update strategy to prevent local minima in the registration update.

Although keeping the objective functions strictly the same between the two alternating updates is preferred in general, it has been found in this study that differing  $p_P$  values in each scenario can provide better convergence behavior in the registration update while still encouraging sparse differences in the image update. Specifically, a higher value of  $p_P$  is used in the registration update than in image update. While a deterministic approach (limited-memory BFGS) was used previously as the optimization algorithm in registration update,<sup>235</sup> a stochastic approach named Adaptive Stochastic Gradient Descent (ASGD)<sup>236</sup> is used here. This algorithm employs a strategy of randomly sampling a subset of image voxels and achieves a substantial reduction in computation time per iteration, while keeping favorable convergence properties. The gradient of the objective function used in ASGD approach can be written as

$$\left[ \nabla_{\lambda} \Phi(\mu, \lambda) \right]_i = \frac{\partial}{\partial \lambda_i} \Phi(\mu, \lambda) = -\beta_P \left[ \Psi_P \dot{\mathbf{W}}_i(\lambda) \mu_P \right]^T \dot{f}_P \left( \Psi_P (\mu - \mathbf{W}(\lambda) \mu_P) \right) \quad (5.5)$$

where  $\dot{f}_P$  denotes the derivative of the modified  $p$ -norm function in prior image penalty on each element of the operand.  $\dot{\mathbf{W}}_i(\lambda)$  denotes the derivative of the deformation operator with respect to the  $i^{\text{th}}$  deformation parameter.

Table 5.1 presents the pseudocode for the alternating maximization approach. In the registration block, the 1<sup>st</sup> (outer) loop corresponds to the morphological pyramids and the 2<sup>nd</sup> (inner) loop corresponds to the ASGD updates. The details of the random sampling and step size in ASGD updates can be found in the paper by Klein *et al.*<sup>236</sup> In the image block,  $\dot{h}_i$  denotes the derivatives of the marginal log-likelihoods,  $\dot{f}_R$  is the derivatives of the modified  $p$ -norm function in image roughness penalty,  $c_i$  is the optimum curvature of the marginal log-likelihoods,  $\omega_f(t)$  is the curvature of the penalty function defined as  $\omega_f(t) = \dot{f}(t)/t$ ,  $K$  is the number of neighboring voxels

used to penalize the  $j^{\text{th}}$  voxel, and  $[\cdot]_+$  is the non-negativity constraint. The image update equation Eq. (5.7) in Table 5.1 is a modified form of Eq. (10) in the paper by Stayman *et al.*<sup>111</sup> that includes a deformation operator with fixed  $\lambda$  on the prior image. The convergence properties and the alternating maximization schedules of the proposed algorithm are discussed in Sec. 5.3.2.

Table 5.1: Pseudocode for the dPIRPLE algorithm.

---

```

Input  $\mu^{[0,0]}, \lambda^{[0,0]}$ 
for  $z = 1$  to  $\text{max\_alternations } (Z)$ 
  % Registration Update Block
  for each image pyramid
    for  $s = 1$  to  $\text{number\_of\_ASGD\_updates } (S_z)$ 
      Randomly sample a subset of image voxels
      Approximate gradient  $\tilde{g}_s = \nabla_{\lambda} \Phi(\mu^{[z-1, T_{z-1}]}, \lambda^{[z, s-1]})$  using sampled voxels and Eq. (5.5)
      Compute an adaptive step size  $\gamma_s$ 
       $\lambda^{[z, s]} = \lambda^{[z, s-1]} - \gamma_s \tilde{g}_s$  (5.6)
    end
  end
  % Image Update Block
  for  $t = 1$  to  $\text{number\_of\_SQS\_updates } (T_z)$ 
    for  $j = 1$  to  $\text{number\_of\_voxels}$ 
      
$$\mu_j^{[z, t]} = \left[ \mu_j^{[z, t-1]} + \frac{\left( \sum_{i=1}^{N_y} \mathbf{A}_{ij} \dot{h}_i \left( \left[ \mathbf{A} \mu^{[z, t-1]} \right]_i \right) - \beta_R \sum_{k=1}^K \left[ \mathbf{\Psi}_R \right]_{kj} \dot{f}_R \left( \left[ \mathbf{\Psi}_R \mu^{[z, t-1]} \right]_k \right) \right.}{\left( \sum_{i=1}^{N_y} \left[ \mathbf{A} \right]_{ij}^2 c_i \left( \left[ \mathbf{A} \mu^{[z, t-1]} \right]_i \right) + \beta_R \sum_{k=1}^K \left[ \mathbf{\Psi}_R \right]_{kj}^2 \omega_{f_R} \left( \left[ \mathbf{\Psi}_R \mu^{[z, t-1]} \right]_k \right) \right.} \right. \quad (5.7)$$


$$\left. \left. - \beta_P \sum_{k=1}^K \left[ \mathbf{\Psi}_P \right]_{kj} \dot{f}_P \left( \left[ \mathbf{\Psi}_P \left( \mu^{[z, t-1]} - \mathbf{W} \left( \lambda^{[z, S_z]} \right) \mu_P \right) \right]_k \right) \right) \right. \\ \left. \left. + \beta_P \sum_{k=1}^K \left[ \mathbf{\Psi}_P \right]_{kj}^2 \omega_{f_P} \left( \left[ \mathbf{\Psi}_P \left( \mu^{[z, t-1]} - \mathbf{W} \left( \lambda^{[z, S_z]} \right) \mu_P \right) \right]_k \right) \right) \right]_+$$

    end
  end
end
return  $\mu^{[Z, T_Z]}, \lambda^{[Z, S_Z]}$ 

```

---

## 5.2.4 Computational complexity and implementation

The computational complexity of dPIRPLE is divided between the image update and the registration update. In the image update, the complexity may be characterized by the number of projection operations needed (forward projections and backprojections), which are the dominant factors for the computation time. Every SQS update requires 3 projection operations if optimum curvature is used. As such, the execution of the dPIRPLE algorithm with respect to image reconstruction in Table 5.1 requires  $3 \times \sum_{z=1}^Z T_z$  back/projections, where  $Z$  denotes the number of alternations. In the registration update, computation time stems from four main sources: 1) computing the gradient of the reduced objective function; 2) evaluating the reduced objective function when computing the step size; 3) computing the Jacobian matrix (of the deformation operator  $\mathbf{W}$  over deformation parameters  $\lambda$ ) at the beginning of each level of the pyramid<sup>236</sup>; and 4) warping and interpolating the moving image at the end of each registration block. The use of the ASGD approach substantially reduces the computation time from the first two sources by randomly sampling a subset of the image voxels to compute the gradient and computing the step size adaptively based on the gradient information instead of function evaluation. As such, the execution of the dPIRPLE algorithm with respect to registration in Table 5.1 is dominated by  $Z$  Jacobian calculations at each of the 4 levels of the pyramid and  $Z$  warping and interpolation operations.

The dPIRPLE algorithm was implemented in Matlab (The Mathworks, Natick, MA), while computationally intensive functions were calculated using optimized C++ libraries. Specifically, the projection operations in the image update used CUDA-based libraries, and the registration update used the image registration toolbox *Elastix*.<sup>237</sup>

The performance of dPIRPLE was compared with a number of other approaches including FBP, PLE, PIPLE, and rigid PIRPLE. All the iterative approaches utilized matched separable

footprint projection operations, which is a modified form<sup>93</sup> of the separable footprint technique,<sup>98</sup> and FBP used a voxel-driven interpolating backprojection.

### 5.2.5 Cadaver experiments

Experiments were carried out on a cadaver torso on a CBCT test-bench. This study emulated a clinical scenario in which an initial diagnostic image has been acquired and a period of time has elapsed and a subsequent follow-up image is necessary. The initial image volume serves as a patient-specific prior image that is used to improve image quality and reduce the required radiation dose in subsequent follow-up scans. Specifically, a lung nodule surveillance scenario was emulated in which a suspicious nodule is imaged in a follow-up study to determine growth rates. In the cadaver experiments, a baseline scan was first acquired using a sufficiently large number of projections, to form a high-quality patient-specific prior image. Then, a follow-up scan was acquired with a substantial reduction in the radiation dose through a reduction in either the number of projections or the exposure per projection. The patient-specific prior image was then used in conjunction with the follow-up scan to reconstruct current anatomy and to detect a newly formed nodule in the presence of patient motion between scans.

The CBCT system consisted an x-ray source (DU694 in an EA10 housing; Dunlee, Aurora, IL), a flat-panel detector (PaxScan 4343CB with 1536×1536 pixels at 0.278 mm pixel pitch after 2×2 binning; Varian Medical Systems, Palo Alto, CA), and a motion control system (Parker Hannifin, OH), as shown in Fig. 5.2(a). While the bench offers a wide range of source-detector motions, in this study, a system geometry was chosen that contains a 150 cm SDD and 120 cm SAD. All images were reconstructed with  $260 \times 300 \times 330$  voxels and  $1 \times 1 \times 1 \text{ mm}^3$  voxel size. The system geometry was calibrated using a separate scan following the calibration method described by Cho *et al.*<sup>238</sup>

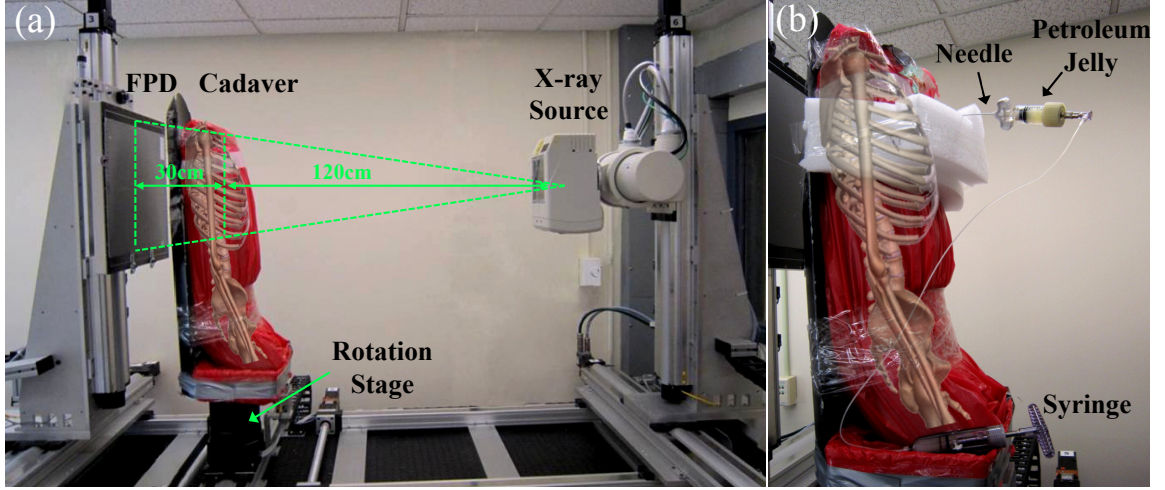


Figure 5.2: (a) Experimental setup on the CBCT test-bench for cadaver experiments. (b) Simulation of lung tumor growth via petroleum jelly injection into the cadaver lung. A semiopaque rendering of a generic skeleton is overlaid on the photograph to illustrate the anatomical position of the cadaver.

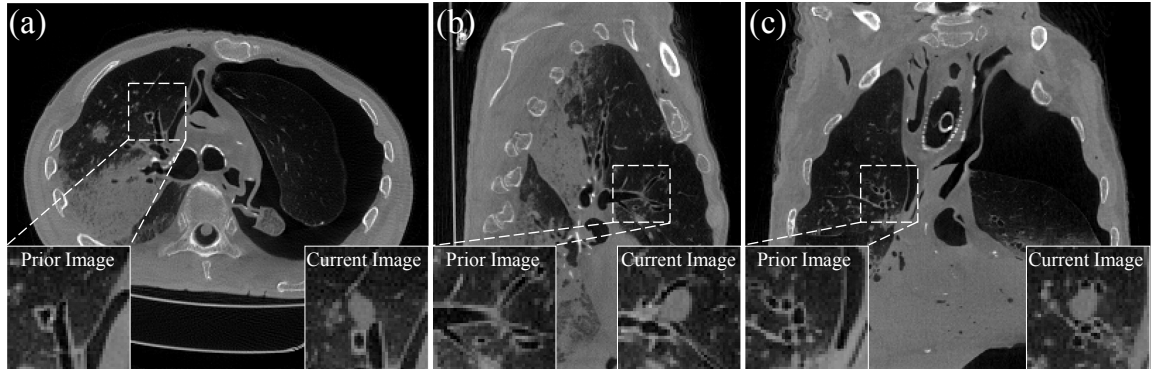


Figure 5.3: Three views of the patient-specific prior image, formed by PLE ( $p_R = 2$ ,  $\beta_R = 10^6$ ) reconstruction of the fully sampled dataset before the injection. Two zoomed-in regions in each view correspond to prior image (left) and current anatomy (right).

A baseline scan was first acquired with 360 projections over  $360^\circ$  (referred to as fully sampled dataset), 100 kVp and 1.25 mAs/projection (referred to as standard exposure). A PLE reconstruction ( $p_R=2$ ,  $\beta_R=10^6$ ) of this dataset was used as a patient-specific prior image, as shown in Fig. 5.3. Approximately  $1 \text{ cm}^3$  petroleum jelly ( $\sim 0.013 \text{ mm}^{-1}$  attenuation) was then injected into the right lung of the cadaver by a thoracic surgeon, as shown in Fig. 5.2(b). Many types of motion were imparted to the cadaver during this procedure, including the deformation of the abdominal

soft tissues, flexing of the spine, and contraction of the chest wall. Two follow-up scans were acquired after the procedure: one with the same fully sampled protocol as used in the baseline scan and one with 0.1 mAs/projection (referred to as the low exposure dataset). A PLE reconstruction ( $p_R=2$ ,  $\beta_R=10^6$ ) of the fully sampled follow-up scan was used as the “ground truth” (and is referred to as the current anatomy). Various sparse datasets were investigated through retrospectively selecting a reduced number of projections from fully sampled data.

To determine the robustness of the deformable registration method with regard to patient motion, a *simulation* study was conducted. In this study, a large number of instances of random rigid motion were applied to the high-quality prior image. These misregistered prior images were used as inputs to the dPIRPLE reconstruction and registration performance after a single registration was assessed. Performance analysis after a single registration leads to a somewhat conservative estimate of the capture range of patient motion since subsequent alternations have the potential to recover from a poor first registration. However, focusing on a single registration allowed for an assessment of a relatively large ensemble of misregistrations, and recovery after a poor first registration was generally not observed.

The first registration is comprised of an initial rigid registration step followed by the deformable registration step described in Sec. 5.2.2. The rigid registration uses Mutual Information<sup>239</sup> as the similarity metric and ASGD as the optimization. (The initial rigid registration was not used in the results above.) Random rigid misregistration was generated as perturbations from the nominal pose of the prior image. The nominal pose corresponded to the optimal rigid alignment of the prior image and the current anatomy, derived from an accurate rigid registration of the prior image and the fully sampled current anatomy reconstruction. Perturbations in each of the six parameters of rigid motion (i.e. translations in  $X$ ,  $Y$ , and  $Z$  and rotations about  $X$ ,  $Y$ , and  $Z$ ) were described by Gaussian probability distributions with *three* standard derivations in each distribution as 100 mm in  $X$  (Anterior-Posterior), 100 mm in  $Y$  (Left-Right), 200 mm in  $Z$  (Superior-Interior), 10 degrees about  $X$ , 10 degrees about  $Y$ , and 10 degrees about  $Z$ . This perturbation model,

adapted from previous work by Otake *et al.*,<sup>240</sup> was intended to emulate realistic variability in clinical setup of the patient in a diagnostic CT scanner. Each misregistration was quantified by computing the mean displacement averaged over all the voxels in the entire FOV of the CBCT.

### 5.2.6 Evaluation methods

To assess the accuracy and image quality of the reconstructed images, reconstructions were compared to the ground truth current anatomy using three metrics, including global root mean square error (RMSE), local RMSE, and structural similarity (SSIM) index.<sup>241</sup>

The RMSE between two images  $u$  and  $v$  is:

$$RMSE = \sqrt{\frac{1}{N_\mu} \sum_{i=1}^{N_\mu} (u_i - v_i)^2} \quad (5.8)$$

where  $N_\mu$  is the number of voxels in either image. The RMSE values have the same units as the image values (i.e. linear attenuation coefficients), that is,  $\text{mm}^{-1}$ . The global RMSE of a reconstructed image is computed over the entire FOV of the CBCT, which reflects the overall accuracy of the reconstructed image. The local RMSE was computed in a region of interest (ROI,  $100 \times 100 \times 100$  voxels) that included the nodule and the adjacent soft tissue in the lung, reflecting reconstruction accuracy in close proximity to the nodule.

SSIM is a good complementary metric to RMSE, since it relates to the perceptual quality of an image, whereas the latter relates to quantitative accuracy. The SSIM is defined as a linear combination of luminance ( $l$ ), contrast ( $c$ ), and structure ( $s$ ):

$$SSIM(u, v) = l(u, v) \cdot c(u, v) \cdot s(u, v) \quad (5.9)$$

where

$$l(u, v) = (2\bar{u}\bar{v} + C_1) / (\bar{u}^2 + \bar{v}^2 + C_1) \quad (5.10a)$$

$$c(u, v) = (2\sigma_u\sigma_v + C_2) / (\sigma_u^2 + \sigma_v^2 + C_2) \quad (5.10b)$$

$$s(u, v) = (\sigma_{uv} + C_3) / (\sigma_u \sigma_v + C_3) \quad (5.10c)$$

and  $\bar{u}$  and  $\bar{v}$  denote the mean voxel values,  $\sigma_u$  and  $\sigma_v$  are standard deviations, and  $\sigma_{uv}$  is the sample covariance. The constants  $C_1$ ,  $C_2$ , and  $C_3$  were chosen as in the paper by Wang *et al.*<sup>241</sup> to prevent instability in the computation of these three similarity measures. SSIM was computed in a ROI ( $16 \times 16 \times 16$  voxels) about the nodule and the adjacent tissues.

To assess registration accuracy, the error vectors of the deformation field were generated from subtracting this field with the “true” field, which was approximated by registering the prior image with current anatomy as estimated by the fully sampled dataset. The deformation field vectors, the error vectors, and the error vector magnitudes were all visualized. Quantitatively, the average magnitude of the error vectors was computed over the object region (no air voxels) and compared to the average magnitude of the deformation field vectors. Additionally, both the difference images between the deformed prior image and the “true” current anatomy and the RMSE of the difference images were computed.

### 5.3 Results

The experimental results are organized into three parts. First, parameter selection, convergence properties, and the alternating maximization schedule for dPIRPLE were investigated. All these results used the same sparsely sampled dataset (20 projections equally spaced over  $190^\circ$ ) with a fixed exposure (1.25 mAs per projection), representing an 18-fold exposure reduction over a fully sampled dataset. Second, the reconstruction results of dPIRPLE using the same dataset were then compared to FBP and other iterative approaches (PLE, rigid PIRPLE, single-registration dPIRPLE). Third, the reconstruction results were evaluated at different measurement sparsity and exposure levels to investigate the robustness of dPIRPLE in preserving image quality and its limits in enabling exposure reduction.

### 5.3.1 Parameter Selection for dPIRPLE

Table 5.2 summarizes the major parameters used in dPIRPLE in this study. Variables are divided into two categories: 1) parameters that define the dPIRPLE objective function (i.e. first 7 parameters); and 2) parameters involved in solving the dPIRPLE objective function (i.e. last 3 parameters).

Table 5.2: Summary of major parameters in dPIRPLE.

Symbol	Name	Nominal Values or Range
$\Psi_R$	Sparsifying operator in image roughness penalty	First order difference operator
$p_R$	Modified $p$ -norm in image roughness penalty	1
$\beta_R$	Penalty strength in image roughness penalty	$10^2 \sim 10^{3.5}$
$\Psi_P$	Sparsifying operator in prior image penalty	Identity matrix
$p_P$	Modified $p$ -norm in prior image penalty	1 (Image update); 2 (Registration update)
$\beta_P$	Penalty strength in prior image penalty	$10^{2.5} \sim 10^4$
$\delta$	Size of quadratic region in modified $p$ -norm	$10^{-4} \text{ mm}^{-1}$
$\sigma$	B-spline control point spacing	10 voxels
$Z$	Maximum number of alternations in optimization	20
$\{S_1, S_2, \dots, S_Z\}$	Number of ASGD updates per pyramid	$\{1000, 1000, \dots, 1000\}$
$\{T_1, T_2, \dots, T_Z\}$	Number of SQS updates per alternation	$\{50, 50, \dots, 50\}$

In these studies,  $p_R=1$  was selected for its edge-preserving effect in the reconstructed image, and  $\Psi_P$  was chosen as the identity matrix since  $\mu\text{-}\mathbf{W}(\lambda)\mu_P$  is already very sparse for the lung nodule growth scenario. In the alternating optimization algorithm, a lower  $p_P$  ( $=1$ ) was used in the image update to encourage sparse differences, while a higher  $p_P$  ( $=2$ ) was used in the registration update to obtain better convergence. A four-level morphological pyramid was used in all the registrations with a control point spacing of 10 voxels at the final pyramid and a downsampling scheme of 8, 4, 2, and 1.

Previous work<sup>111</sup> found that the optimal values of the three regularization parameters  $\delta$ ,  $\beta_P$ , and  $\beta_R$  in the PIRPLE objective function vary with differing datasets including differences in x-ray technique, system geometry, object, number of projections, and volume size/sampling. Therefore, parameter values need to be carefully selected for each dataset to achieve optimal image quality. That study showed that  $\delta$  needs to be below a certain threshold relative to the expected attenuation differences in the reconstruction (and values larger than this threshold have little effect on reconstruction quality). In the current study,  $\delta=10^{-4} \text{ mm}^{-1}$  was found small enough for all the datasets. For each dataset, optimal  $\beta_P$  and  $\beta_R$  were then selected by performing a pair of one-dimensional parameter sweeps as described in previous work,<sup>111</sup> that is, optimizing first over  $\beta_P$  while holding  $\beta_R$  fixed at an initially small value, and then optimizing over  $\beta_R$  at the optimal  $\beta_P$ , using RMSE as the metric. It has been observed in this study that as long as the initial  $\beta_R$  is chosen to be a *small* value and within the wide range of proper starting points (e.g.  $10^0$  to around  $10^{3.5}$  in Fig. 5.4), this pair of 1D sweeps would always reach close to the optimum. (More details on the double 1D sweep can be found in the sensitivity to parameter selection section in previous work.<sup>111</sup>) This pair of 1D sweeps (c.f., a full 2D sweep over  $\beta_P$  and  $\beta_R$ ) is computationally convenient and takes advantage of the typical shape of  $\text{RMSE}(\beta_P, \beta_R)$  as illustrated in Fig. 5.4 (dashed lines). The optimal  $(\beta_P, \beta_R)$  resulting from the double 1D sweep was nearly identical to the true 2D optimum (asterisk). This result demonstrates that the method proposed in previous work<sup>111</sup> is also useful with respect to a deformable registration estimate (previously demonstrated with respect to rigid registration).

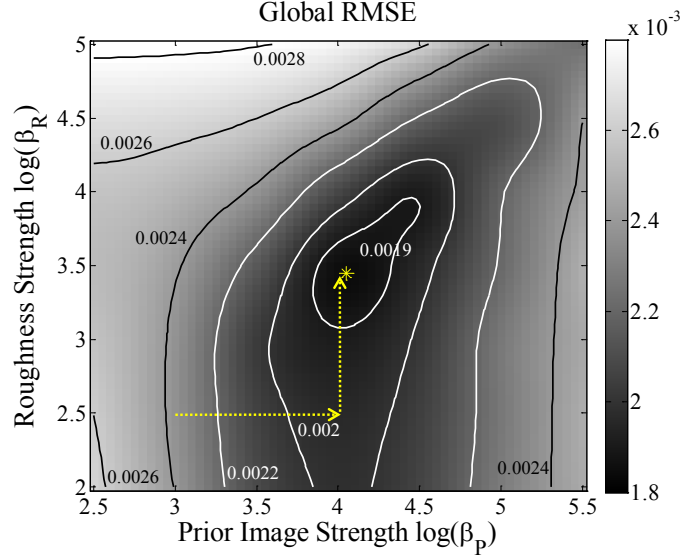


Figure 5.4: Example optimization of  $(\beta_P, \beta_R)$  from a pair of 1D parameter sweeps (dashed lines) compared to the true 2D optimum (asterisk). Results here used the global RMSE as a metric; however, very similar results are seen when using local RMSE.

The last three parameters in Table 5.2 are important in determining the schedule of the proposed alternating maximization approach, and the choice of those parameters is investigated in Sec. 5.3.3. A PLE image was found to give excellent initialization of the attenuation parameters in dPIRPLE. This choice provided a better starting point than a flat (zero) image, helping to speed convergence in the image update. Moreover, the PLE image also provided a feature rich dataset for estimating the deformation as compared to the prior image, which is critical in the first registration block. In contrast, initializing with a flat image can be detrimental to registration and is not particularly good for reconstruction. The deformation parameters of dPIRPLE were initialized with all zeros in the first level of the morphological pyramid in the first registration block. For fair comparison, all prior-image-based approaches (PIPLE, rigid PIRPLE, and dPIRPLE) used the same initialization image and  $\delta$ , except that  $\beta_R$  and  $\beta_P$  were optimized separately for each approach and each dataset. PLE was initialized with a FBP image and used the same  $\delta$ , while  $\beta_R$  was optimized for each dataset.

### 5.3.2 Convergence properties of dPIRPLE

Before evaluating the convergence properties of dPIRPLE with respect to registration, the accuracy of the final registration estimate was measured. Figure 5.5(a-b) shows the final deformation field in the axial and sagittal views, which clearly depict the movement of the body toward the posterior and the right side, as well as the anterior displacement of the heart. Note that each vector in the deformation field is plotted in the reverse direction from the physical deformation (i.e., pointing from a voxel in the current anatomy to the corresponding voxel in the prior image). Also note that the visualized deformation field vectors have a subsampled density (with a factor of 15) and a scaled vector length (with a factor of 2) from the original deformation field vectors to help better visualize the field. Figure 5.5(c-d) shows the error vectors generated from subtracting this final deformation field with the “true” field. Note that within the object the average magnitude of the error vectors is 0.77 mm, smaller than the 1 mm voxel size and much smaller than the average magnitude of the vectors in the estimated deformation field (6.56 mm). Error vectors with larger magnitudes outside the object are attributed to noise in air. To further visualize the deformation field error, the magnitudes of the error vectors were plotted in Fig. 5.5(e-f). Small to moderate errors (~0-2.5 mm) are evident within the object, such as the superior medial region in (e) and the inferior right region in (f), corresponding to errors in estimating the deformation of vasculature in the mediastinum and a slight mismatch in the displacement of the heart, respectively.

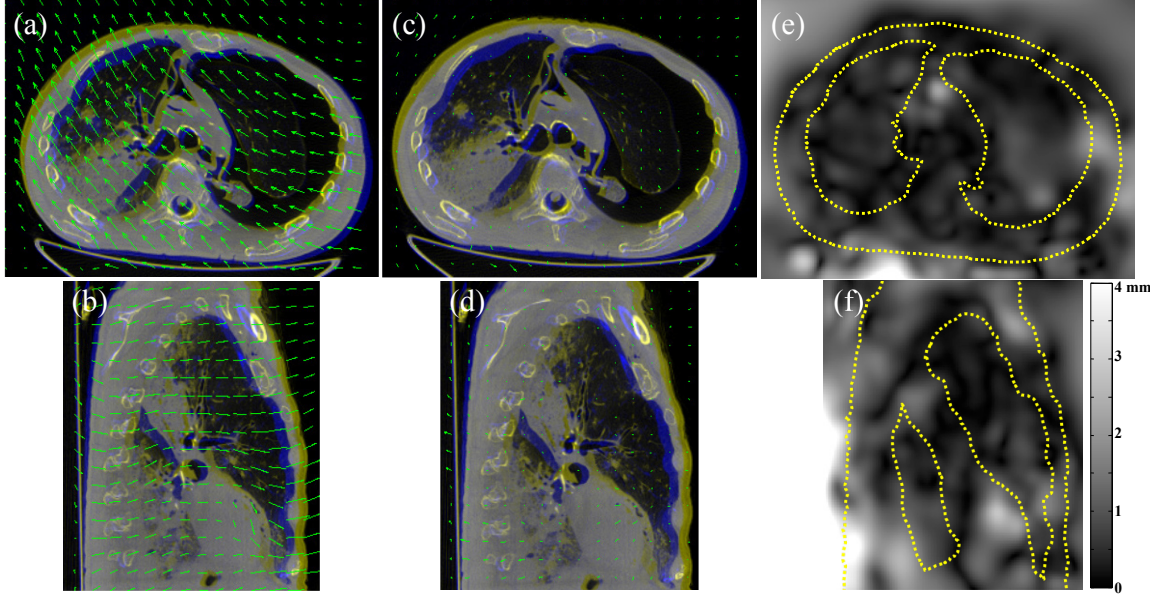


Figure 5.5: (a, b) Deformation field estimated by dPIRPLE using 20 projections. The image below the deformation field is a merged image from the prior image and the current anatomy. The yellow and blue color correspond to positive and negative image value differences between the prior image and the current anatomy, while the original greyscale is used in the region where the image value differences are small. (c, d) Error vectors between deformation field estimated by dPIRPLE and the “true” deformation field approximated by registering the prior image with current anatomy. (e, f) Plot of the magnitude of error vectors (with a trace of the body outline superimposed).

The convergence properties of dPIRPLE in registration were evaluated by examining the progression of the registration residual error at different alternations (i.e., outer loop iterations) as shown in Fig. 5.6. The registration residual error was computed using the difference image between the deformed prior image and current anatomy. It can be seen that the majority of mismatch was compensated after the 1<sup>st</sup> alternation of registration, which substantially prevented incorrect structures from being injected into subsequent image updates. The moving image used in this 1<sup>st</sup> registration was the initialization image in dPIRPLE, that is, a PLE image reconstructed using the sparsely sampled dataset (shown later in Fig. 5.9, third row). However, remaining differences continued to be reduced between the 1<sup>st</sup> and the 20<sup>th</sup> alternation, especially in the areas of the ribs and primary bronchi, demonstrating the importance of the joint estimation. Note that the bright spot at  $z=1$  and  $z=20$  was not registration error but was created due to the introduced nodule which is inherent in the difference image between deformed prior and current anatomy. The local RMSE

between the deformed prior image and current anatomy was  $76.5 \times 10^{-4} \text{ mm}^{-1}$  at  $z=0$ ,  $29.3 \times 10^{-4} \text{ mm}^{-1}$  at  $z=1$ ,  $24.6 \times 10^{-4} \text{ mm}^{-1}$  at  $z=20$ , and  $22.6 \times 10^{-4} \text{ mm}^{-1}$  using the “true” registration. These values reflect the same qualitative improvements as can be seen throughout iterations providing additional indication of the advantage of the joint estimation approach.

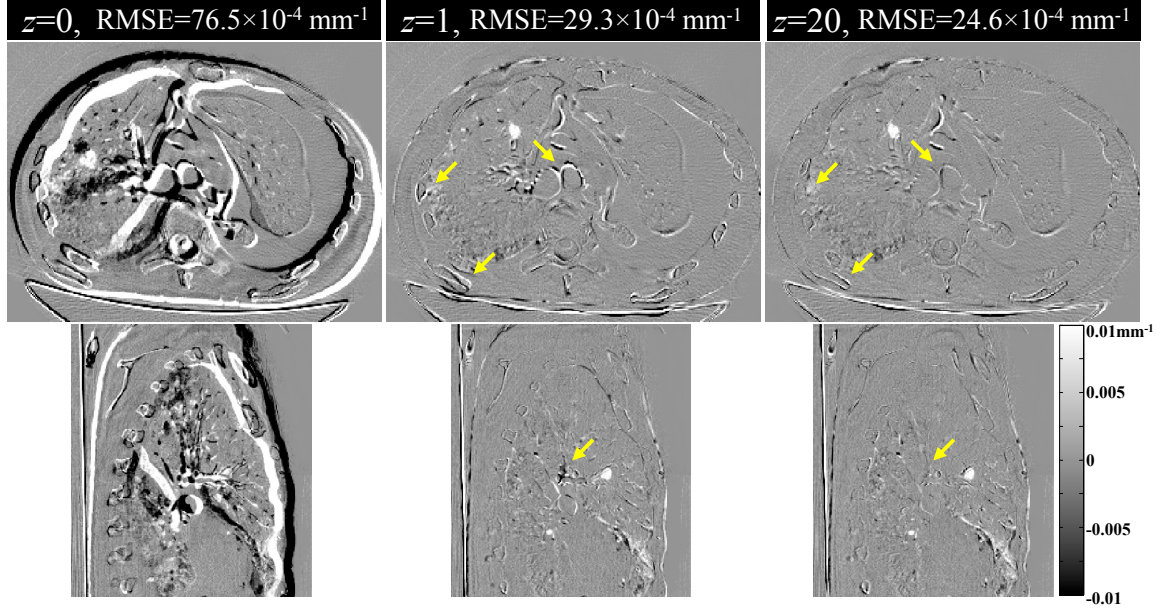


Figure 5.6: Axial and sagittal views of a sequence of registration residual errors in dPIRPLE using 20 projections over  $190^\circ$ . The value  $z$  reflects the estimation at the  $z^{\text{th}}$  alternation of registration updates.

The convergence of dPIRPLE was also quantified by plotting its objective function difference versus iteration in Fig. 5.7(a). The objective function value at the solution, denoted as  $\Phi^{[\infty]}$ , was approximated by the value at the end of all 20 alternations ( $Z=20$ ) including 20 registration blocks and 20 image blocks. Note that the cumulative number of image updates  $n$  labeled on the horizontal axis was a measurement of all the SQS updates spanning over alternations ( $n = \sum_{z=1}^Z T_z$ ). In each image block, a number of  $T_z$  ( $=50$ ) SQS updates were performed, followed by a registration block with a number of  $S_z$  ( $=1000$ ) registration updates in each level of the pyramid, followed by the next image block, etc. Note that the objective function values in Fig. 5.7(a) were computed using Eq. (5.2) with  $p_r=1$  (no change between image and registration

blocks). Figure 5.7(a) shows that the objective function increased monotonically within every image block due to the monotonicity of the SQS approach. The increase in objective function values was greatest in the 1<sup>st</sup> registration block and exhibited smaller improvement in subsequent registration blocks, consistent with the progression of the residual registration error in Fig. 5.6.

The convergence of dPIRPLE was also evaluated in comparison to other iterative approaches using local RMSE. As shown in Fig. 5.7(b), PLE quickly reduced the RMSE but plateaued at a relatively high RMSE. PIPLE did not substantially reduce the RMSE due to the mismatched prior image. In contrast, both rigid PIRPLE and dPIRPLE saw steady reduction throughout the iterations, while dPIRPLE showed a higher reduction rate than rigid PIRPLE. Note that the RMSE was compared at the same iteration rather than at the same computation time. dPIRPLE and rigid PIRPLE with  $T_z=50$  required 77.5% and 4.7% more computation time than PIPLE due to the registration step, while PLE required 15.5% less time than PIPLE due to the lack of a prior image penalty. The amount of time needed for the deformable registration in dPIRPLE is discussed in the following section.

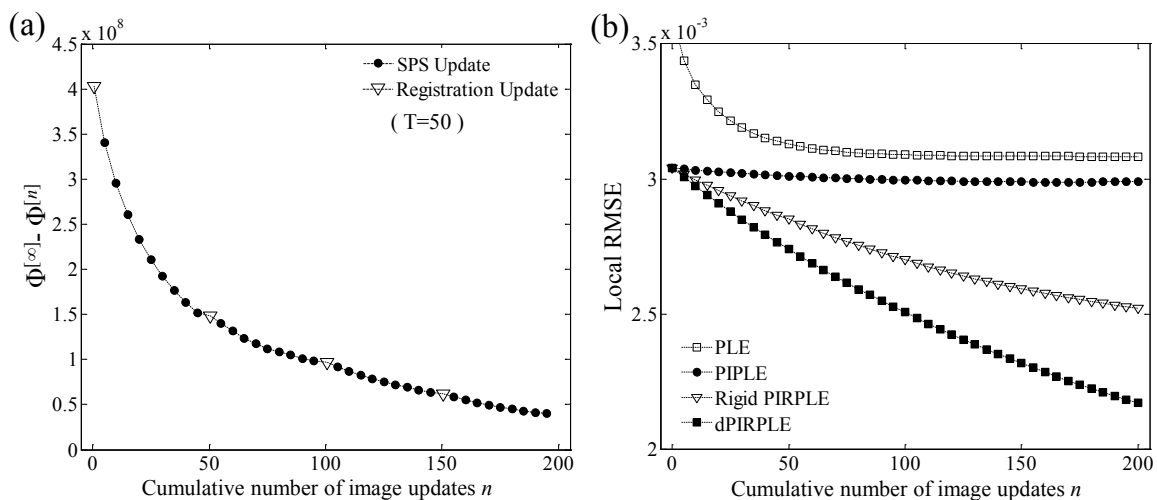


Figure 5.7: (a) The dPIRPLE objective function difference analyzed as a function of the cumulative number of image updates,  $T_z=50$ . (b) Local RMSE versus iteration number for all approaches.

### 5.3.3 Alternating maximization schedule

The alternating maximization schedule is determined by three parameters: the maximum number of alternations in the optimization ( $Z$ ), the number of ASGD updates per pyramid level in the  $z^{\text{th}}$  alternation ( $S_z$ ), and the number of SQS updates in the  $z^{\text{th}}$  alternation ( $T_z$ ).  $Z=20$  was found to be sufficient to produce a nearly converged dPIRPLE image with little or no change in the image after  $Z=20$ . While the proposed alternating algorithm allows  $S_z$  and  $T_z$  to vary during alternations, only the case of constant  $S$  and  $T$  was considered in this study. (Finding optimal values of  $S_z$  and  $T_z$  at each alternation to achieve fastest convergence is the subject of future work.) Within each level of the pyramid, since the registration updates are dominated by the Jacobian calculation, one can perform many registration updates after the Jacobian calculation without substantially increasing the computation time. In this work,  $S=1000$  was selected since larger  $S$  leads to marginal changes in the deformation field. In contrast, the number  $T$  has more influence in determining the computation time of the dPIRPLE technique. Increasing  $T$  means more SQS updates in one image block and relatively less frequent registration. If registration computation time were not an issue, it is arguably preferable to do registration updates more frequently so that the deformation estimation is more accurate at each image update. However, because multi-resolution deformable registration does have significant computational cost, there exists a trade-off between computation time and registration update frequency (and therefore reconstruction accuracy), primarily determined by  $T$ .

To investigate this trade-off, Fig. 5.8 depicts the local RMSE after  $n$  cumulative image updates for various  $T$  values. Note that  $T=1000$  means one registration followed by 1000 SQS updates, corresponding to the simple staged estimation scenario with one registration followed by reconstruction. The  $T=1000$  case shows significantly higher local RMSE, again demonstrating the importance of the joint estimation. In the zoomed region,  $T=50$ , 25, and 5 all exhibit similarly low local RMSE at the 1000<sup>th</sup> iteration. However, since  $T=50$  requires the least computation time among those three choices (as shown in Table 5.3),  $T=50$  was selected as the optimal  $T$  in these studies.

With more relaxed time constraints, one might choose different optimal  $T$  values according to different clinical requirements. The computation times in Table 5.3 were measured on a high performance workstation equipped with two Intel Xeon E5-2600 processors and one Nvidia GeForce GTX 680 graphics card.

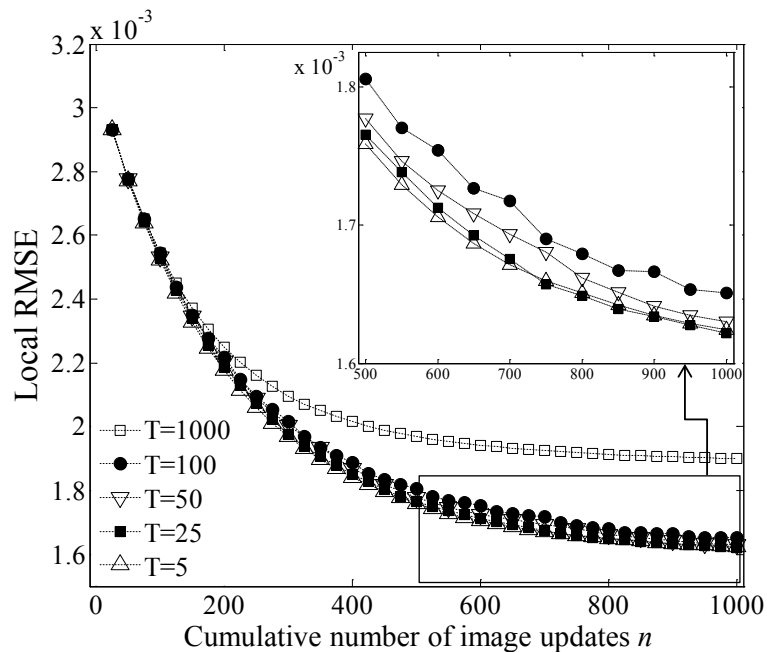


Figure 5.8: Local RMSE computed versus the cumulative number of image updates for dPIRPLE at various  $T$  values.

Table 5.3: Computation time at various settings of  $T$ .

$T$	Registration Time (min)	Reconstruction Time (min)	Total Time (h)
5	1100	142	20.7
25	220	142	6
50	110	142	4.2
100	55	142	3.3
1000	5.5	142	2.5

### 5.3.4 Reconstruction results

The reconstruction results of dPIRPLE were compared with other approaches as well as the "true" current anatomy, as shown in Fig. 5.9. All of these results used the same sparsely sampled dataset (20 projections equally spaced over 190°) with a fixed exposure (1.25 mAs per projection), representing an 18-fold exposure reduction over a fully sampled dataset. All iterative approaches (PLE, rigid PIRPLE, single-registration dPIRPLE, and dPIRPLE) used 1000 iterations (i.e. 1000 SQS updates) to generate nearly converged images. FBP exhibited substantial artifacts from sparse sampling which made reliable nodule detection very difficult. PLE reduced the artifacts, but strong regularization resulted in relatively low spatial resolution which made nodule detection and volume changes difficult to assess. Rigid PIRPLE appeared to overcome ambiguous structures in areas of more rigid motion such as the ribs, spine, and primary bronchi, but mismatch in the more deformable areas resulted in errors in reconstruction of the nodule. Single-registration dPIRPLE corresponds to a scenario of one deformable registration followed by 1000 SQS updates (i.e.  $T=1000$  case in Sec. 5.3.3). Despite fairly accurate nodule reconstruction, incorrect anatomy can be seen in a few locations across the anatomy such as inside the dashed circle, mainly caused by registration residual errors after the 1<sup>st</sup> registration. In contrast, dPIRPLE presented a highly accurate estimate of the true anatomy and the nodule.

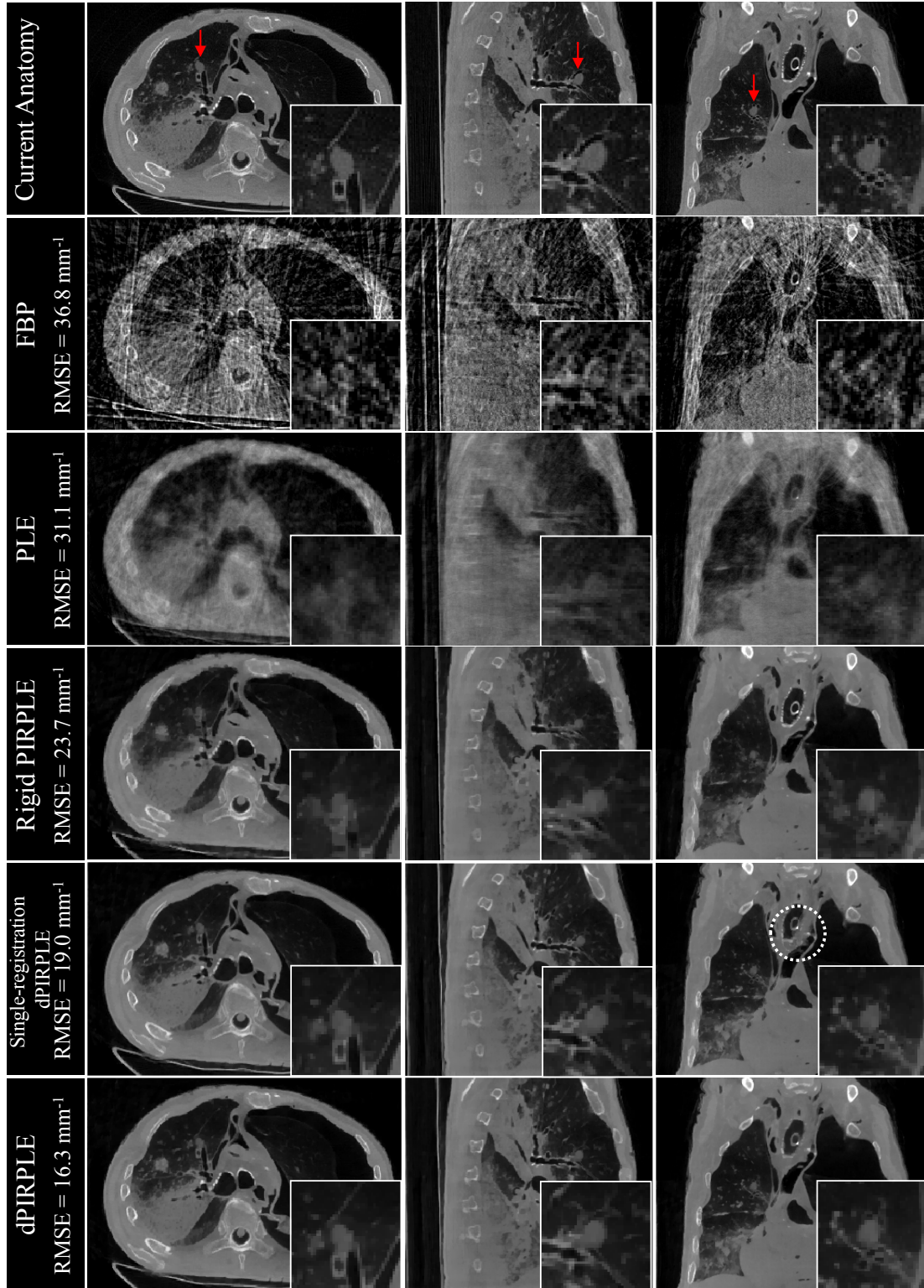


Figure 5.9: Reconstruction results of FBP, PLE ( $\beta_R=10^3$ ), rigid PIRPLE ( $\beta_P=10^4$ ,  $\beta_R=10^{3.5}$ ), single-registration dPIRPLE ( $\beta_P=10^4$ ,  $\beta_R=10^{3.5}$ ), and dPIRPLE ( $\beta_P=10^4$ ,  $\beta_R=10^{3.5}$ ) with 20 projections and 1.25 mAs/projection. The local RMSE of the reconstruction result is displayed in each case. Note the accurate estimate of nodule and lung structures (e.g. airways) in dPIRPLE.

### 5.3.5 Performance with varying sparsity

To investigate the performance of dPIRPLE in comparison with other iterative approaches and varying levels of projection sparsity, subsets of data from a full set of 360 projections over 360° were selected to form sparse datasets of 200, 100, 40, 20, and 10 projections in a short scan orbit that covers 200°. Moreover, the standard exposure (1.25 mAs/projection) used in Sec. 5.3.1-5.3.4 was replaced with a low exposure (0.1 mAs/projection) acquisition to further challenge dPIRPLE at every sparsity level. For each approach and each dataset, an optimal  $\beta_R$  (PLE) or pair of  $(\beta_P, \beta_R)$  (PIPLE, rigid PIRPLE, and dPIRPLE) was chosen using the aforementioned 1D parameter sweep to find the lowest global RMSE and SSIM. If the two metrics resulted in different optima, an average value between the two optima was used.

Figure 5.10 summarizes the reconstructed images for each approach over a range of sparse sampling. While all approaches exhibit accurate estimates at 200 projections (i.e., one projection per degree covering 180° plus fan angle), as the sparsity increases, PLE, PIPEL, and rigid PIRPLE all exhibit apparent reductions in image quality. In PLE, as the sparsity increases, both the spatial resolution and the contrast over the entire image drop quickly. Visualization of the nodule becomes less accurate and hard to identify when the number of projections drops to 20 or below. In PIPEL, with the addition of patient-specific prior information, both the contrast and spatial resolution are better maintained compared to PLE at each level of sparsity. However, due to a lack of registration, anatomical mismatches are readily apparent. As the sparsity increases, the nodule becomes more and more distorted due to the reduced information provided by the measurements, and the nodule largely disappears at 10 projections. The joint registration in rigid PIRPLE helps to reduce a number of mismatches associated and improves overall image quality compared to PIPEL at each level of sparsity. However, errors are still visible in areas of more deformable anatomy. In particular, artifacts and a loss of spatial resolution are evident at soft tissue and airway boundaries in the lung. Moreover, the accuracy in reconstruction of the nodule is degraded as the sparsity increases. The

remaining mismatches due to deformation are further corrected in dPIRPLE, and a high degree of accuracy is maintained over a broad range in sparsity. At the lowest number of projections (10), dPIRPLE was still somewhat able to preserve the nodule shape and achieve much better nodule reconstruction compared to other approaches, although a reduction in the nodule contrast can be seen.

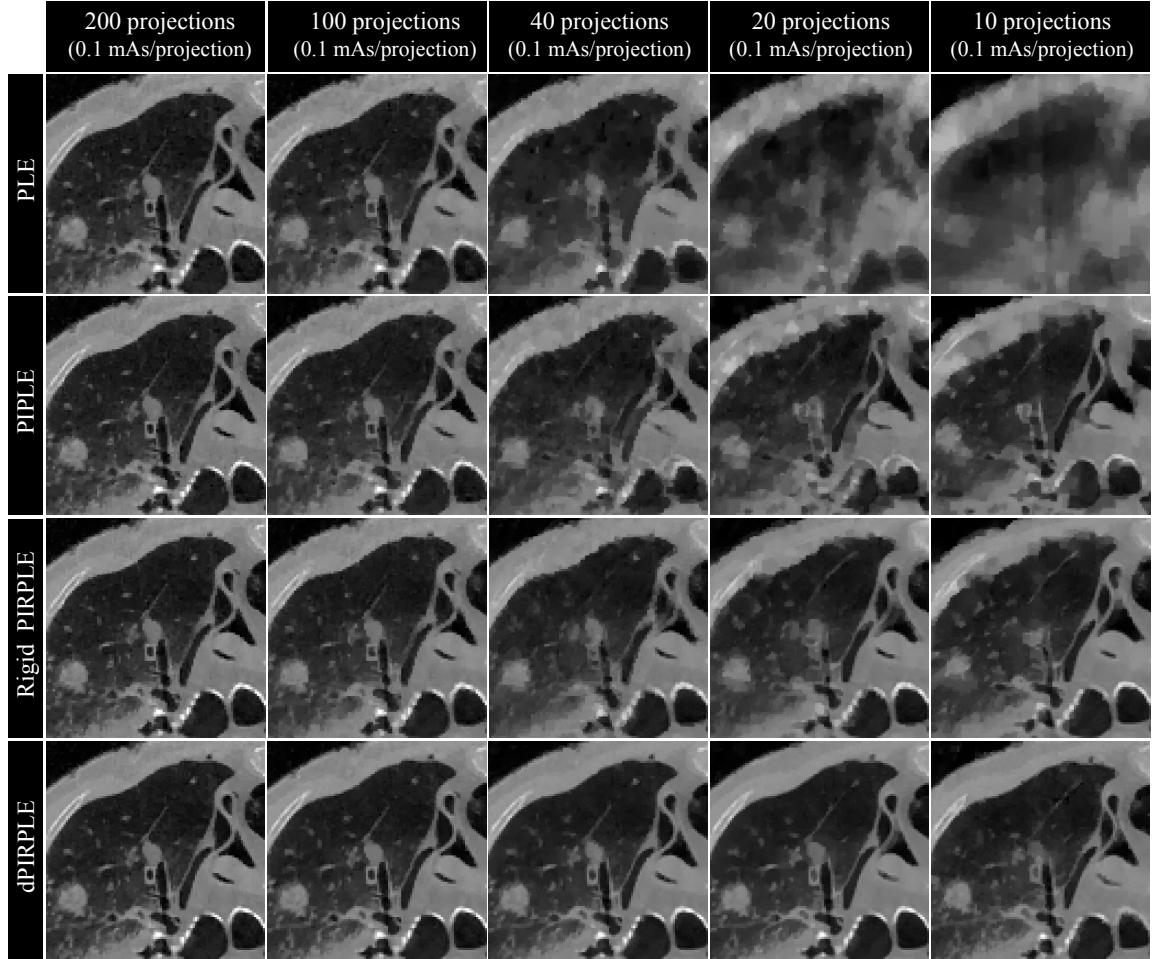


Figure 5.10: PLE, PIPLE, rigid PIRPLE and dPIRPLE reconstruction images at different levels of projection sparsity at low exposure (0.1 mAs/projection).

The accuracy of the reconstructed images was quantified using global RMSE and SSIM. Figure 5.11(a) shows that RMSE exhibited a clear monotonic dependence on the sparsity for all iterative approaches. Moreover, the sensitivity to sampling was substantially lower for dPIRPLE than for other iterative approaches. Note that the use of a prior image without registration (PIPLE)

resulted in even higher error than the case without a prior image (PLE), illustrating the importance of proper registration. The SSIM results in Fig. 5.11(b) show similar trends as in Fig. 5.11(a), and a notable difference in the sensitivity to sampling is observed for dPIRPLE in comparison to the other approaches. A possible reason is that the SSIM metric in Fig. 5.11(b) only considers the nodule area instead of the entire image and the residual registration errors in the other approaches have more influence on the image quality in the vicinity of the nodule. Figure 5.11 also shows better image estimation in dPIRPLE than PLE at 200 and 100 projections, indicating that incorporating a patient-specific prior image can increase imaging performance even at moderate or full sampling.

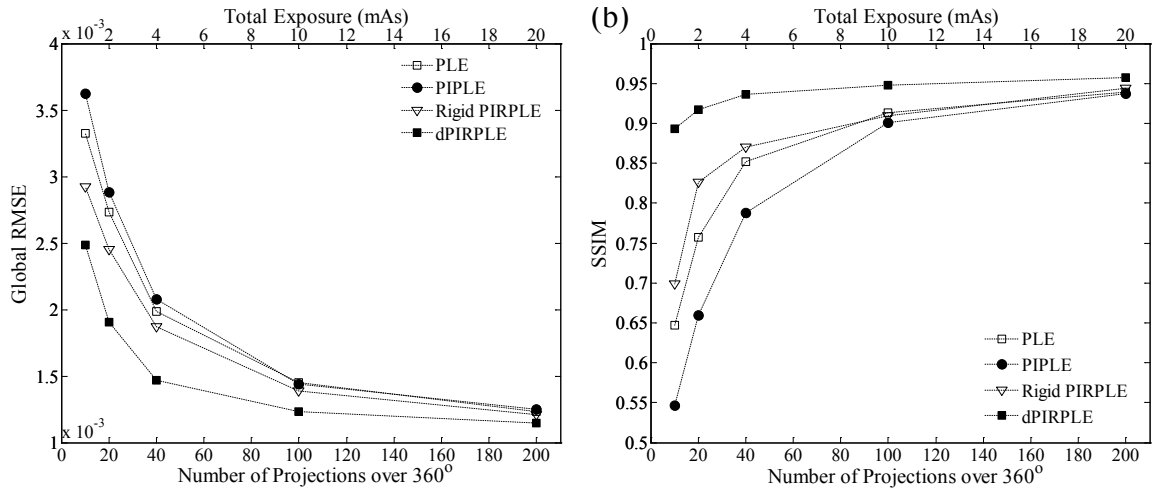


Figure 5.11: Global image accuracy [(a) RMSE] and local image quality [(b) SSIM] for PLE, PIPE, rigid PIRPLE, and dPIRPLE reconstruction computed as a function of sparsity at low exposure (0.1 mAs/projection).

### 5.3.6 Capture range of the deformable registration in dPIRPLE

Previous sections demonstrated accurate registration from the deformable registration method used in dPIRPLE for the specific patient motion introduced in the cadaver experiment. A *simulation* study was also conducted to determine the robustness (i.e. capture range) of the deformable registration method with regard to patient motion. Figure 5.12 shows the RMSE between the

registered prior image and the current anatomy as a function of the mean displacement after both the initial rigid registration and after both rigid and deformable registration. The RMSE remained consistently small when the mean displacement was less than 125 mm and became relatively large when the mean displacement was over 125 mm. Even though the deformable registration is able to decrease the RMSE substantially across the entire range, those cases with  $\text{RMSE} > 30 \text{ mm}^{-1}$  do not provide clinically useful image reconstructions. Thus, deformable registration does not compensate for a poor rigid registration and the delineation between success and failure cases lies at a mean displacement of 125 mm (indicated by the dashed line). While the 125 mm capture range may be appropriate for many clinical tasks, this range may be extended with a more robust initialization of the rigid registration.<sup>240</sup>

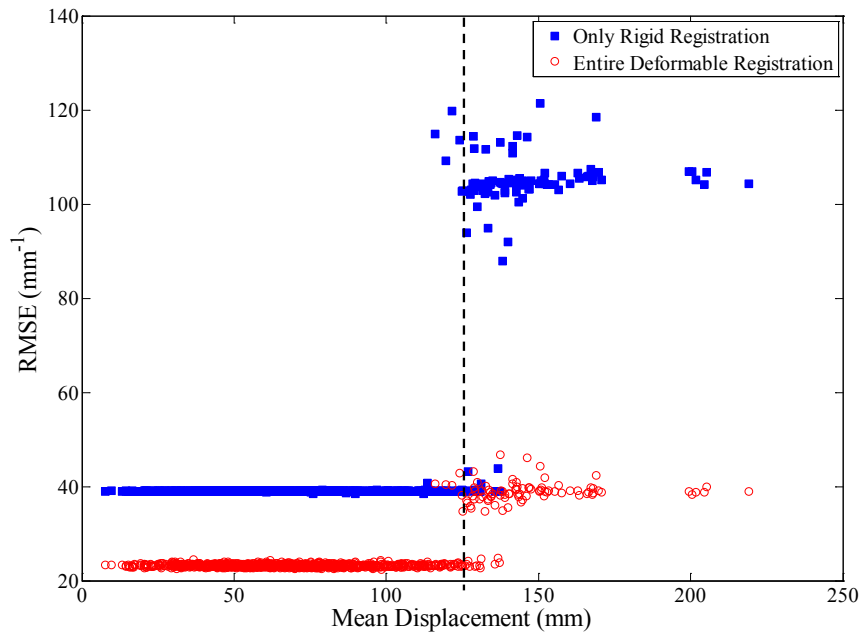


Figure 5.12: RMSE between the registered prior image and the current anatomy as a function of the mean displacement after both the initial rigid registration and after both rigid and deformable registration. Each vertical pair of data points correspond to one instance of random rigid motion that was added to the prior image before registration. The dashed line indicates the delineation between success and failure cases at a mean displacement of 125 mm.

## 5.4 Conclusions and discussion

This chapter introduced a model-based framework, dPIRPLE, that jointly: 1) estimates 3D deformation between a high-quality patient-specific prior image and current anatomy based on subsequent data acquisitions, and 2) estimates a reconstructed image from noisy measurements using the deformed prior image. The proposed framework is solved using an alternating maximization strategy, and both the parameter selection and optimization schedule were investigated. Cadaver experiments were conducted emulating a lung nodule surveillance scenario, and the dPIRPLE algorithm demonstrated superior reconstruction accuracy and image quality compared to FBP, PLE, PLE with an unregistered prior image (PIPLE), and PLE with rigid registration of a prior image (PIRPLE) over a wide range of sampling sparsity and exposure levels.

While dPIRPLE provides a general framework that allows for different kinds of deformable registration in the registration block, a specific registration method utilizing a cubic B-spline-based FFD model and ASGD algorithm was selected in this work. This particular registration method was able to capture a large portion of the deformation in the very first registration block and continued to be iteratively refined with additional alternations of the maximization algorithm. These alternations were found to be essential in minimizing residual errors and maximizing imaging performance. However, despite the success of the dPIRPLE approach using B-spline-based FFD, some residual errors were observed even after a large number of alternations. These residual mismatches could possibly be resolved using more sophisticated registration methods, and is the subject of future work. For example, the deformation field estimation might be further refined by applying the Demons algorithm and its variants.<sup>242–244</sup> The capture range of the deformable registration in dPIRPLE was characterized by adapting a large number of random rigid motions to the prior image. A more realistic way may be to perform cadaver or phantom studies with increasing levels of non-rigid motion and is the subject of future work.

Investigation of the schedule for dPIRPLE alternating maximization demonstrated the trade-off between the reconstruction image accuracy and the computation time. In this chapter, the trade-off was primarily determined by the parameter  $T$ , the number of image updates in one image block. However, this chapter also presumed a constant number of image updates ( $T$ ) and registration updates ( $S$ ) per block during alternations. The more generalized scenario where  $T_z$  and  $S_z$  may vary at the  $z^{\text{th}}$  alternation may potentially provide greater opportunity to balance accuracy versus computation time. For example, since more dramatic changes in the image estimate typically happen in earlier image updates, one might start with relatively low  $T_i$  for the up-to-date deformation estimate and gradually increase  $T_i$  to reduce the computation time. Such studies are the subject of ongoing and future work.

The use of the modified  $p$ -norm in the prior image penalty term allowed additional flexibility to encourage or enforce differences from the prior image in the reconstruction. In this chapter, investigations used a strategy where norms were mismatched between the image update (a modified  $L^1$  norm) and the registration update ( $L^2$  norm). This approach was found to be advantageous in improving convergence and avoiding local minima. However, despite the desirable convergence properties seen in practice with this method, this norm mismatch leads to solutions that do not strictly maximize the original objective (either the  $L^1$  or  $L^2$  form). Although beyond the scope of this chapter, this issue may be handled by applying methods similar to graduated non-convexity optimization.<sup>245</sup> For example, one could start the first registration block with  $p_P=2$  and gradually reduce the value of  $p_P$  in subsequent registration blocks to  $p_P=1$ . In this fashion, the desirable convergence properties observed above could be combined with a stricter convergence in a single objective.

Proper selection of regularization parameters is important for all model-based reconstruction approaches including dPIRPLE. Incorrect selection of regularization parameters, especially prior image penalty strength  $\beta_P$ , will not only reduce reconstructed image quality but have the potential to bias the reconstruction towards the prior data (i.e. to show no change). The

studies in this chapter employed the same method as in previous work<sup>111</sup> to select optimal regularization parameters and demonstrated that knowledgeable 1D parameter sweeps in  $\beta_P$  and  $\beta_R$  individually can be used to identify optimal dPIRPLE parameters accurately and much more efficiently than an explicit, computationally intense 2D parameter sweep. When a truth image is not provided, one could still perform the basic search method with visual assessment instead of using a RMSE metric. These approaches still require several reconstructions to perform the optimization sweep. One method that can be used to help reduce this computational burden will be introduced in the next chapter, which leverages analytical approximations to the implicitly defined estimator along with predictive performance metrics to determine the optimal penalty strength. This method enables efficient selection of proper regularization parameters especially prior image penalty strength  $\beta_P$ , ensuring minimal bias of the reconstructed image towards the prior data in regions where change is present between the current and prior anatomy. Moreover, although the penalty strength  $\beta_P$  and  $\beta_R$  are treated as scalars in this work, this method is sufficiently flexible to allow the use of a spatially varying map of  $\beta_P$  or  $\beta_R$  to design and customize a spatially varying penalty strength. This is similar to other intentionally spatially varying regularization approaches used in PLE to enforce spatially uniform resolution<sup>134</sup> or optimization of task-based detectability.<sup>136</sup> In the context of prior-image-based reconstruction, spatially varying penalty design could enforce more uniform inclusion of prior image information or intentionally nonuniform designs based on where anatomical change is found. More about the design of prior image strength including spatially varying penalty design will be covered in the next chapter.

One very important question associated with all prior-image-based reconstruction methods concerns the veracity of features found in the reconstruction. In these studies, false structures were observed in reconstructions where registration was absent or insufficiently accurate (e.g., doubling of anatomy in extreme circumstances). Moreover, despite these obvious flaws in the image, the apparent image quality (e.g., sharpness) of the image can remain high. Thus, even though such image defects were not observed with dPIRPLE, it would be valuable to have an assessment that

identifies potential false structure. In the paper by Stayman *et al.*,<sup>246</sup> a methodology was proposed whereby the reconstruction is decomposed into portions individually attributable to the data and the prior image. Such an approach could potentially be applied in dPIRPLE to ensure that false structures are not present in reconstructions.

While this chapter focused on the lung nodule surveillance application, the general framework is applicable to many other scenarios where patient-specific prior images, imaging sequences, or longitudinal studies are available, including image-guided radiation therapy, post-operative treatment assessment, and monitoring of patients in the intensive care unit. Many of these applications have more stringent timing constraints and will require refinements and more computationally efficient implementations of the dPIRPLE approach. For example, a fully GPU-based implementation of dPIRPLE is expected to accelerate parallelizable operations (the current implementation only uses the GPU for projection operations and requires frequent and inefficient data transfers between CPU and GPU). Employing GPU for registration in dPIRPLE alone could potentially reduce each registration update from current minutes to seconds.<sup>247</sup> Furthermore, combining ordered subsets and momentum methods has recently been shown to substantially accelerate the x-ray CT image reconstruction.<sup>167,168</sup> Recent experiments<sup>191</sup> have suggested that standard non-sparse reconstructions for interventional systems can be computed in a few minutes. With such improvements, the dPIRPLE approach is expected to be valuable in a wide range of imaging scenarios and in different anatomical sites.

Moreover, extensions to the approach whereby prior image data is available from alternate imaging modalities can further generalize the application of dPIRPLE. In image-guided surgery, for example, dPIRPLE could be adapted to using a preoperative diagnostic CT prior image to improve image quality and reduce dose in intraoperative CBCT. A similar scenario can be envisioned for image-guided radiotherapy. dPIRPLE may also be applied to dual-energy CT in which a high-fidelity high-energy image is used to improve low-fidelity low-energy image with certain mechanism to prevent from perturbing the accuracy of the attenuation coefficients. Even

more varied approaches with cross-modality registration are potentially possible using MRI and nuclear imaging studies. Such extensions would bring a wider spectrum of interesting challenges to the dPIRPLE framework, such as other forms of reduced fidelity data (e.g., alternate sparse samplings) and the need for novel similarity metrics for inclusion of cross-modality prior images. To address the potential additional difficulties in inconsistent intensity values, possible solutions may include matching the intensity values of the prior image to those of the current anatomy,<sup>248</sup> employing a similarity metric that does not require consistent intensity values such as a mutual-information-based metric,<sup>249</sup> or exploiting feature-based registration methods such as registering the surface point sets of both images.<sup>250</sup> Another potential difficulty lies in that motion between scans is likely to be much larger than the one introduced in the cadaver experiment for some applications, for example when a prior image is acquired months before. Section 5.3.6 has demonstrated that dPIRPLE has a fairly wide capture range thereby being able to compensate fairly large motion. For even larger motion, more advanced methods are needed that would tolerate large initial misregistration. For pulmonary imaging, this might include approaches that exploit airway bifurcation structures and lung surface meshes.<sup>251</sup>

## Chapter 6

# 6. Control of Change Admission in Prior-Image-Based Reconstruction

## 6.1 Introduction

### 6.1.1 Balancing prior information with new data

As shown in the previous chapter, MBIR can exploit anatomical information from prior images to improve image quality and/or reduce radiation dose. The general framework referred to as PIRPLE<sup>111</sup> and the dPIRPLE variant incorporating a deformable registration model demonstrates the strong advantages presented by prior-image-based reconstruction (PIBR) methods, particularly under conditions of sparse, low-dose sampling, and sequential imaging studies.

Despite the benefits of using patient-specific prior images in image reconstruction, a key question needs to be answered in PIRPLE (and dPIRPLE) as well as other PIBR methods: To what extent should prior image information be used to achieve accurate image reconstruction? That is, how much information should come from the prior image and how much should come from the measurements. Such balance is typically controlled via regularization parameters, such as  $\beta_P$  in PIRPLE (and dPIRPLE),  $\alpha/\lambda$  in PICCS with statistical weights,<sup>218</sup>  $\beta_M$  in the Reconstruction of Difference method,<sup>252</sup> and  $\beta$  in image reconstruction that penalizes differences between image patches.<sup>219</sup> Inappropriate selection of the prior image strength can lead to poor reconstruction. Specifically, the use of too little prior image information fails to produce a significant imaging benefit with low fidelity data, while the use of too much prior image information can force the

reconstructed image to simply replicate the prior image, potentially obscuring anatomical changes and producing false structures. Further complicating the balance is that the optimal prior image strength can vary across different patients, anatomical changes (e.g., attenuation, shape, and size), acquisition geometry, x-ray techniques, and image reconstruction parameters (e.g., voxel size, volume size).

Traditional methods for choosing the proper prior image strength include exhaustive searches and heuristics / look-up tables. Exhaustive search involves performing a large number of image reconstructions with different regularization parameter values and choosing a value corresponding to an image that optimizes a certain image quality metric; however, this method can be extremely time-consuming, since each reconstruction requires iterative solution. Heuristics and look-up tables do not require image reconstruction and therefore require much less time, but they can be subject to error and suboptimal solutions due to the aforementioned variations of the optimal strength across imaging studies.

In this chapter, a novel method is proposed that prospectively estimates the optimal prior image strength for PIBR without heuristics or exhaustive search. This method leverages an analytical approximation to PIBR objective functions containing non-quadratic penalties.<sup>246</sup> A predictive performance metric is introduced that utilizes the approximate analytical solution and a specification of an anticipated change (i.e., an anatomical change for which accurate reconstruction is to be ensured). This performance metric is, in turn, used to estimate the optimal prior image strength. Additionally, because optimal prior image strength can depend on the location of anatomical change, a spatially varying map of prior image strength is proposed to optimally admit changes everywhere in the image. Thus, the proposed design can ensure accurate reconstructions without *a priori* knowledge of the change location.

The proposed methodology is investigated both in an ellipse phantom and in a realistic thorax phantom emulating a lung nodule surveillance scenario. Performance is compared with traditional exhaustive searches, and the optimality of the spatially varying design is explored.

Additionally, the dependence of optimal prior image strength on different properties of the change (i.e., attenuation, shape, size) is also investigated.

### **6.1.2 Acknowledgement and unique contributions**

The methods and results reported below were reported in conference proceedings and journal articles as follows:

- (1) H. Dang, Jeffrey H. Siewerdsen, and J. Webster Stayman, "Regularization design and control of change admission in prior-image-based reconstruction," *SPIE Medical Imaging*, San Diego, CA, Vol. 9033, 90330O (2014).
- (2) H. Dang, J. H. Siewerdsen, and J. W. Stayman, "Prospective regularization design in prior-image-based reconstruction," *Physics in Medicine and Biology*, 60 (24), 9515-9536 (2015).

with permission from the publisher for reproduction of content in this dissertation. The author's primary contributions in this work were as follows: development of a method that prospectively estimates optimal prior image strength for accurate reconstruction of anatomical changes in prior-image-based reconstruction; introduction of spatially varying prior image penalty for optimal admission of the anatomical change everywhere in the image; phantom experiments; quantitative evaluation of the reported method including estimation accuracy and computation time; and investigation of the dependence of optimal prior image strength on contrast, location, size, and shape of the anatomical change. Valuable discussion with Dr. Rick Colbeth, Dr. Sungwon Yoon, and Mr. Edward Shapiro at Varian Medical Systems (Palo Alto, CA) is gratefully appreciated. The work was supported by research collaboration with Varian Medical Systems (Palo Alto, CA) and by the National Institutes of Health Grant No. 2R01-CA-112163.

## 6.2 Methods

### 6.2.1 Regularization design in prior-image-based reconstruction (PIBR)

While a number of PIBR approaches exist, the investigations below focused specifically on the PIRPLE methodology. It is expected that the basic framework can be extended to other approaches, including PICCS with statistical weights,<sup>218</sup> image reconstruction using non-local prior functions,<sup>219</sup> and reconstruction of difference using prior images.<sup>252</sup> The PIRPLE method employs a PLE framework which incorporates patient-specific prior image information through a regularization term. The objective function of PIRPLE contains three terms: 1) a data fidelity term that uses the measured data based on measurement statistics; 2) an image roughness regularization term that enforces local smoothness and/or edge-preservation in the image estimate; and 3) a prior image regularization term that enforces similarity of the image estimate to a prior image while allowing sparse differences (i.e., anatomical changes) between the two. The prior image is simultaneously registered to the current patient anatomy in either a rigid fashion<sup>111</sup> or a non-rigid fashion (as shown in the previous chapter) to account for patient motion between the previous scan and the current scan. In this chapter, the PIRPLE objective function without registration is considered:

$$\hat{\mu}_{PIRPLE} = \arg \max_{\mu} L(y; \mu) - \beta_R \|\Psi_R \mu\|_{p_R}^{p_R} - \beta_P \|\Psi_P (\mu - \mu_P)\|_{p_P}^{p_P} \quad (6.1)$$

As in the previous chapter,  $p_R = 1$  is chosen because  $L^1$  norm penalty function or its variants (e.g., Huber function<sup>131</sup>) has been shown to encourage edge preservation and achieve improved noise-resolution tradeoff in the reconstructed image compared to quadratic penalty function,<sup>13,93</sup> and  $p_P = 1$  is chosen because  $L^1$  norm penalty function or its variants has been shown to encourage similarity of the reconstructed image to the prior image but also allow for sparse differences.<sup>110,111</sup> The sparsifying operator  $\Psi_R$  is chosen to compute first-order neighborhood pairwise voxel differences,

and  $\Psi_P$  is the identity matrix since the anatomical changes between the prior image and the current anatomy are already sparse. In situations where the anatomical changes are less sparse (e.g., in the case of a much larger anatomical change), a sparsifying transform other than the identity matrix would be encouraged.

Accurate reconstruction of anatomical changes in PIBR requires a proper balance between the information from the prior image and the measurements. In PIRPLE, this balance is governed primarily by the prior image strength parameter  $\beta_P$ . A larger  $\beta_P$  leads to the use of more information from the prior image in the current image reconstruction, while smaller  $\beta_P$  restricts the amount of information from the prior image. Accurate reconstruction of anatomical changes in PIRPLE is also affected by the image roughness strength  $\beta_R$  through the control of image smoothness. Previous work by Stayman *et al.*<sup>111</sup> suggests that the optimization of these two parameters is often separable, suggesting that the two parameters can be selected independently. Specifically, previous work<sup>111</sup> found that the optimal value of  $\beta_P$  is nearly independent of  $\beta_R$  when  $\beta_R$  is low. Therefore, a 1D optimization over  $\beta_P$  can be first performed with a low  $\beta_R$  to estimate the optimal  $\beta_P$ , which is followed by another 1D optimization over  $\beta_R$  with the optimal  $\beta_P$ . The optimal  $\beta_P$  and  $\beta_R$  estimated from this pair of separate 1D optimization using RMSE from a truth image as the metric has been found to closely match the optimal  $\beta_P$  and  $\beta_R$  estimated from an exhaustive 2D optimization.<sup>111</sup> This chapter focuses on investigating the selection of the optimal prior image strength  $\beta_P$  in the context of a fixed image roughness strength  $\beta_R$ , which is the first step of these two 1D optimizations.

## 6.2.2 Approximate analytical solution of PIBR

The objective function of PIRPLE in Eq. (6.1) does not generally admit a closed-form solution because of the nonlinearities of both the log-likelihood function and the non-quadratic regularization. In the previous chapter, this objective function is solved iteratively using optimization approaches. However, it is possible to derive an *approximate* closed-form solution of

the objective. Such approximation does not necessarily avoid a full iterative solution for the desired reconstruction of subsequent imaging studies, but may suffice for prospective regularization design. In this chapter, an approximate closed-form solution is derived by substituting each non-quadratic term in Eq. (6.1) with a quadratic term.

First, the data fidelity term can be approximated by a weighted least-squares term using a second-order Taylor approximation of the log-likelihood function.<sup>139</sup> The simplified objective function can be written as:

$$\hat{\mu} \approx \arg \min_{\mu} \|\mathbf{A}\mu - l\|_{\mathbf{W}}^2 + \beta_R \|\Psi_R \mu\|_{p_R}^{p_R} + \beta_P \|\Psi_P (\mu - \mu_P)\|_{p_P}^{p_P} \quad (6.2)$$

where the weighted least-squares in Eq. (6.2) are the same as the one defined in Sec. 1.4.9. Note that in some cases the Gaussian assumption of the penalized weighted least-squares data fidelity term is preferred over the nonlinear Poisson likelihood. For example, the diagonal weighting  $\mathbf{W}$  can be modified to accommodate various data corrections that modify the noise structure as illustrated in Chapter 2.

Second, non-quadratic regularization (image roughness regularization and prior image regularization) terms can also be approximated by quadratic terms. Using the modified  $L^1$  norm and choosing a proper operating point about which one wishes to form approximate penalty estimates, the modified norm can be approximated by a quadratic function. One may approximate the scalar function applied to each element of the vector argument of the modified norm as follows:

$$f(x_i) \approx g(x_i) = \kappa_i(\tau_i) x_i^2 \quad (6.3a)$$

$$f(x_i) = \begin{cases} \frac{1}{2\delta} x_i^2 & |x_i| < \delta \\ |x_i| - \frac{\delta}{2} & |x_i| \geq \delta \end{cases}, \quad g(x_i) = \kappa_i(\tau_i) x_i^2 = \begin{cases} \frac{1}{2\delta} x_i^2 & |\tau_i| < \delta \\ \frac{|\tau_i| - \delta/2}{\tau_i^2} x_i^2 & |\tau_i| \geq \delta \end{cases} \quad (6.3b)$$

where  $f$  denotes the modified penalty for the  $p = 1$  scenario which is equivalent to the Huber function. The function  $f$  has a scalar input  $x_i$ , and may be approximated with the quadratic function

$g$ . The function  $g$  includes a coefficient  $\kappa_i$  that is computed as a function of the operating point  $\tau_i$ . As shown in Eq. (6.3b), when  $\tau_i$  is chosen to be within the quadratic neighborhood  $([-\delta, \delta])$  of the function  $f$ , the values of the function  $f$  and  $g$  are exactly matched for any  $x_i$  within the quadratic neighborhood. This scenario is illustrated in Fig. 6.1(a). When  $\tau_i$  is chosen to be outside the quadratic neighborhood, the values of the function  $f$  and  $g$  are exactly matched at  $x_i = \tau_i$  and remain close to each other around  $x_i = \tau_i$ , as shown in Fig. 6.1(b). Such quadratic approximation of the non-quadratic regularization yields two important observations. First, for each input element of  $f$ , a separate operating point will be chosen, and a separate parabola will be constructed indicating a location-dependent approximation. Second, since the approximation is most accurate around the operating point, it is desirable to select an operating point that is equal to or close to the value at which the penalty is evaluated.

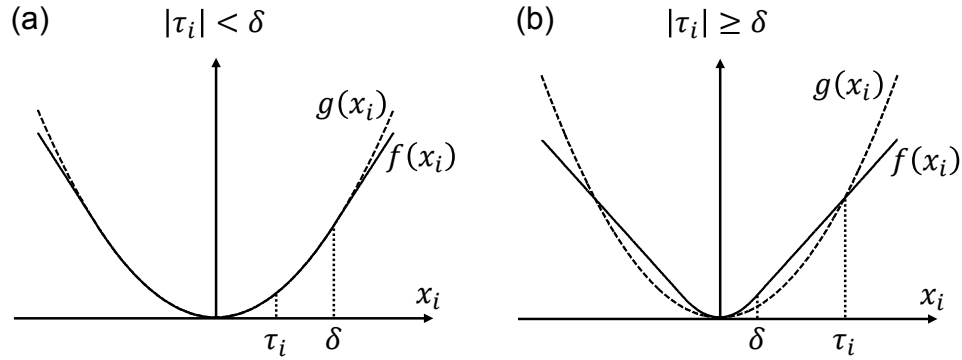


Figure 6.1: Approximation of the modified penalty function  $f$  (equivalent to the Huber function) with a quadratic function  $g$  about an operating point  $\tau_i$ . The operating point is selected either within (a) or outside of (b) the quadratic neighborhood  $([-\delta, \delta])$  of the function  $f$ .

Using the quadratic approximation in Eq. (6.3), the entire modified  $L^1$  norm can be approximated and represented in the following matrix form:

$$\|x\|_1^1 = \sum_i f(x_i) \approx \sum_i \kappa_i(\tau_i) x_i^2 = (x)^T \mathbf{D}\{\kappa(\tau)\}(x) \quad (6.4)$$

where  $\|x\|_1^1$  denotes the modified  $L^1$  norm of a vector  $x$ , and  $\kappa$  and  $\tau$  denote a vector of coefficients and operating points, respectively. One can then apply Eq. (6.4) to the two non-quadratic regularization terms in Eq. (6.2) to yield two quadratic terms:

$$\hat{\mu} \approx \arg \min_{\mu} \left\| \mathbf{A}\mu - l \right\|_{\mathbf{W}}^2 + \beta_R (\Psi_R \mu)^T \mathbf{D}_R \Psi_R \mu + \beta_P \left[ \Psi_P (\mu - \mu_P) \right]^T \mathbf{D}_P \Psi_P (\mu - \mu_P) \quad (6.5a)$$

$$\mathbf{D}_R = \mathbf{D} \left\{ \kappa(\Psi_R \tilde{\mu}) \right\}, \quad \mathbf{D}_P = \mathbf{D} \left\{ \kappa(\Psi_P (\tilde{\mu} - \mu_P)) \right\} \quad (6.5b)$$

For each regularization term, an operating point must be defined in the quadratic approximations. Ideally, this operating point should be an image volume close to the solution of the original objective function. Presuming one has an image estimate  $\tilde{\mu}$ , this leads to an operating point  $\Psi_R \tilde{\mu}$  for the image roughness regularization and  $\Psi_P (\tilde{\mu} - \mu_P)$  for the prior image regularization, as shown in Eq. (6.5b). When the image estimate  $\tilde{\mu}$  is chosen to be close to the PIRPLE solution  $\hat{\mu}_{PIRPLE}$ , such selection of the operation point is expected to yield accurate approximation of the regularization terms evaluated at  $\hat{\mu}_{PIRPLE}$ . Details on choosing a specific image estimate  $\tilde{\mu}$  will be discussed in the next section.

The objective function in Eq. (6.5a) now contains three quadratic terms thereby admitting a closed-form solution, which can be written as:

$$\hat{\mu} \approx \left( \mathbf{A}^T \mathbf{W} \mathbf{A} + \beta_R \Psi_R^T \mathbf{D}_R \Psi_R + \beta_P \Psi_P^T \mathbf{D}_P \Psi_P \right)^{-1} \left( \mathbf{A}^T \mathbf{W} l + \beta_P \Psi_P^T \mathbf{D}_P \Psi_P \mu_P \right) \quad (6.6)$$

This closed-form solution will be used later in Sec. 6.2.4 to estimate the optimal prior image strength.

### 6.2.3 Selection of an operating point

Proper selection of the operating points  $\Psi_R \tilde{\mu}$  and  $\Psi_P (\tilde{\mu} - \mu_P)$  in Eq. (6.5b) is important for accurate approximation of the actual solution of PIRPLE. Ideally, one should use the PIRPLE solution as  $\tilde{\mu}$  so that the value of the approximate quadratic function exactly matches that of the Huber function. This is referred to as the Ideal estimate, whose approximate solution of PIRPLE is denoted  $\hat{\mu}_{Ideal}$ . While this estimate requires a full PIRPLE reconstruction, which is computationally expensive and supersedes the need for an approximate solution, the Ideal estimate

is useful in investigating the accuracy of the underlying quadratic approximation of the non-quadratic regularizations.

A practical operating point is one that can be used for prospective regularization design without having performed the reconstruction. In sequential imaging studies, one often has general knowledge of the anticipated change (or perhaps the kind of changes one might wish to see) in the subsequent scan. For example, in pulmonary nodule surveillance, clinicians may have some knowledge of the nodule's attenuation, size, and shape in a follow-up scan based on the progress of the disease and based on the appearance of the nodule in a previous scan. Similar situations can be found in image-guided radiation therapy, where clinicians may anticipate the location of a tumor or tissue at risk in the current scan based on its location in previous scans. Likewise in image-guided procedures where specific tissue volumes are targeted for resection or specific implants are introduced into the patient, and preoperative scans provide the basis for prior image information.

While knowledge of possible change in the image volume generally includes varying levels of certainty about the specific attenuation, location, shape, and size, it is convenient to start with the assumption that the change is known. Similarly, for prospective regularization design, one might presume that a particular change is present in order to ensure that the regularization design is appropriate should the actual change be present. In this case, one could form an image estimate  $\tilde{\mu}$  as the sum of a prior image  $\mu_P$  and some presumed change  $\mu_C$ :

$$\tilde{\mu} = \mu_P + \mu_C \quad (6.7)$$

This method is referred to as the P+C estimate, with the approximate PIRPLE solution based on Eq. (6.6) denoted as  $\hat{\mu}_{P+C}$ . Since the image estimate  $\tilde{\mu}$  in the P+C estimate does not use the PIRPLE solution, a full image reconstruction is not required in this method. Moreover, one could use the approximate solution from the first P+C estimate (denoted as  $\hat{\mu}_{P+C}^1$ ) as the image estimate  $\tilde{\mu}$  to perform another P+C estimate, and repeat  $n$  times to get the  $n^{\text{th}}$  iteration P+C estimate (denoted as  $\hat{\mu}_{P+C}^n$ ). Iterating on the P+C estimate is expected to yield better selection of the operating point,

thereby resulting in more accurate approximation of the PIRPLE solution, though such a procedure will increase the computation time associated with the estimation. All the estimation methods mentioned in this section along with the full PIRPLE reconstruction method are summarized in Table 6.1.

Table 6.1: Summary of different ways of selecting the operating point for the approximate analytical solution of PIRPLE. An operating point is not needed if doing a full PIRPLE reconstruction (first row), while it is needed and defined as the PIRPLE solution in Ideal estimate (second row), defined as the sum of a prior image and some presumed change in P+C estimate (third row), and defined as the results from the  $(n-1)^{\text{th}}$  iteration P+C estimate in the  $n^{\text{th}}$  iteration P+C estimate (last row).

Method Name	$\tilde{\mu}$ in the Operating Point	Output Image
PIRPLE Reconstruction	N/A	$\hat{\mu}_{PIRPLE}$
Ideal Estimate	$\hat{\mu}_{PIRPLE}$	$\hat{\mu}_{Ideal}$
P+C Estimate (1 iteration)	$\mu_P + \mu_C$	$\hat{\mu}_{P+C}^1$
P+C Estimate ( $n$ iterations)	$\hat{\mu}_{P+C}^{n-1}$	$\hat{\mu}_{P+C}^n$

#### 6.2.4 Predictive performance metric

Previous work by Stayman *et al.*<sup>111</sup> considered optimal prior image strength  $\beta_P$  based on minimizing the RMSE of a PIRPLE-reconstructed image with respect to a truth image in a ROI containing the change. In this chapter, a predictive performance metric is proposed which uses the approximate analytical solution introduced in the previous section to estimate a value of  $\beta_P$  that optimally admits a given anatomical change in the image reconstruction. The metric can be written as:

$$\hat{\beta}_P = \arg \min_{\beta_P} \|\mu_P + \mu_C - \hat{\mu}(\beta_P)\|_S \quad (6.8)$$

which computes the RMSE between the approximate solution and the sum of a prior image and some presumed change in a ROI that contains the change. This ROI is referred to as the Change ROI and is represented using a binary mask  $S$  in Eq. (6.8).

A simple scheme is chosen to minimize the metric in Eq. (6.8) by evaluating at different  $\beta_P$  with regular spacing and choosing the  $\beta_P$  that yields the minimal metric value. More sophisticated searches for the optimal  $\beta_P$  value are the subject of future work. Since the system matrix  $\mathbf{A}$  is typically very large and is not computed explicitly, the inverse operation in Eq. (6.6) is solved in practice using a linear conjugate gradient (CG) method. Specifically, a system of linear equations are formed in a matrix form as  $Zx = b$ , in which matrix  $Z$  corresponds to the term  $\mathbf{A}^T \mathbf{W} \mathbf{A} + \beta_R \Psi_R^T \mathbf{D}_R \Psi_R + \beta_P \Psi_P^T \mathbf{D}_P \Psi_P$  in Eq. (6.6), vector  $b$  corresponds to the term  $\mathbf{A}^T \mathbf{W} l + \beta_P \Psi_P^T \mathbf{D}_P \Psi_P \mu_P$  in Eq. (6.6), and the vector  $x$  computed by a linear CG method corresponds to the image estimate  $\hat{\mu}$  in Eq. (6.6). Since computing an approximate solution is much faster than doing a full image reconstruction, the proposed method is expected to be much more efficient than a traditional exhaustive search that requires full image reconstructions.

### 6.2.5 Spatially varying prior image strength map

The optimal prior image strength for a given anatomical change has been found to be shift-variant from previous work.<sup>253</sup> That is, optimal strength designed at one location is not necessarily optimal for other locations even for identical anatomical changes. Thus, this shift-variance introduces a challenge for regularization design when the location of the change is not known *a priori*. To address this problem, a spatially varying  $\beta_P$  map is proposed, similar to other intentionally spatially varying regularization approaches such as for enforcing uniform resolution<sup>134</sup> and optimization of task-based detectability.<sup>136</sup> Specifically, the approach performs individual optimizations of  $\beta_P$  at every possible location (given a specific presumed anatomical change positioned at each location) to form a  $\beta_P$  map that optimally admits change everywhere. Although repeating the proposed regularization design method at every location requires significantly more computation time than for only one location, such designs can be performed at any time after the prior image scan and before the subsequent scan. For example, surveilling a solitary pulmonary nodule involves a time

between scans of several months,<sup>254</sup> leaving ample time between any two adjacent scans for performing such design. Similarly in fractionated radiotherapy, there is typically several days between the planning scan and the first day of treatment, and one day between subsequent scans at each fraction of treatment. In practice, such design can also be accelerated by estimating the optimal  $\beta_P$  at each grid point and interpolating the results across the image. The design of a spatially varying  $\beta_P$  map is first demonstrated in the ellipse phantom in Sec. 6.3.3 and then used to reconstruct a solitary pulmonary nodule in the thorax phantom in Sec. 6.3.5.

## 6.2.6 Digital phantoms and simulation studies

Two digital phantoms are used in this work and are shown in Fig. 6.2. The ellipse phantom in Fig. 6.2(a) consists of three components: 1) a background ellipse (major axis 41 cm, minor axis 32.4 cm) with attenuation ( $0.021 \text{ mm}^{-1}$ ) similar to soft tissue; 2) a dense circular insert with attenuation ( $0.041 \text{ mm}^{-1}$ ) comparable to bone; 3) and a low-density circular insert with attenuation ( $0.001 \text{ mm}^{-1}$ ) close to air. These components together comprise the anatomical information in the previous scan. For subsequent scans, an anatomical change (a small disc) was introduced in one or both of the two locations shown in Fig. 6.2(a). The ellipse phantom is used in Sec. 6.3.3 to demonstrate the shift-variance of the optimal  $\beta_P$  and the design of a spatially varying  $\beta_P$  map, and used in Sec. 6.3.4 to study the dependence of the optimal  $\beta_P$  on three properties of an anatomical change (attenuation, shape, and size). In all of these studies, a system geometry was chosen that contained a 150 cm SDD, 122 cm SAD, and  $0.556 \times 0.556 \text{ mm}^2$  detector pixel sizes. Only 20 projections equally distributed over  $190^\circ$  were acquired in the subsequent scan, representing a factor of 18 exposure reduction as compared to a typical complete scan (360 projections over  $360^\circ$ ). The 20 projections were simulated using  $10^6$  photons per detector pixel with Poisson noise.

The thorax phantom in Fig. 6.2(b) was generated from an axial slice of a CT scan of a cadaver torso. A lung nodule surveillance scenario was emulated in which a uniform disc emulating

a lung nodule (not present in the prior image scan but present in the subsequent scan) was placed in either the center or the periphery of the lung as shown in Fig. 6.2(b). The thorax phantom is used in Sec. 6.3.1 and 6.3.2 to evaluate the accuracy of the approximate analytical solution and predictive performance metric, and used in Sec. 6.3.5 to evaluate the design of a spatially varying  $\beta_P$  map in a lung nodule surveillance scenario. The projections were generated using the same system geometry and x-ray technique as for the ellipse phantom, except for a lower number of photons per detector pixel ( $10^5$ ).

All PIRPLE reconstructions and image estimates used  $340 \times 420 \times 1$  voxels for the ellipse phantom and  $260 \times 300 \times 1$  voxels for the thorax phantom, both with  $1 \times 1 \times 1$  mm<sup>3</sup> voxel sizes. The Change ROI in the predictive performance metric was set to a circular region with a radius of 30 voxels (large enough to cover the change) as illustrated by the dashed circles in Fig. 6.2. It was found that changing the size of the Change ROI (still large enough to cover the change) did not change the results reported below; therefore, the size of the Change ROI was fixed in this study. The size of the quadratic neighborhood  $\delta$  in the Huber function was set to be  $10^{-4}$  mm<sup>-1</sup> as in the previous chapter. The image roughness strength  $\beta_R$  was fixed at  $10^2$  in the ellipse phantom experiments and 10 in the thorax phantom experiments.

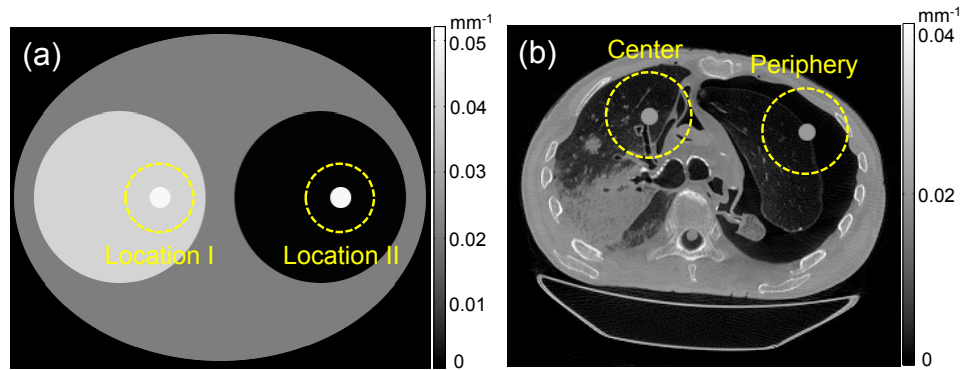


Figure 6.2: (a) Ellipse phantom with attenuation (mm<sup>-1</sup>): 0.021 (background ellipse), 0.041 (bright circular insert), and 0.001 (dark circular insert). An anatomical change is introduced in one or both of Location I and Location II in the subsequent scan. The dashed circles illustrate the Change ROIs used in the predictive performance metric. (b) Thorax phantom generated from an axial slice of a CT scan of a cadaver torso. A uniform circular lung nodule was introduced in the subsequent scan in the center or the periphery of the lung to emulate a lung nodule surveillance scenario.

### **6.2.7 Computational complexity and implementation**

The computational complexity of both the proposed method (using P+C estimate) and the traditional exhaustive search are primarily determined by the total number of projection operations (forward projections and backprojections). Both methods use the same number of projection operations every iteration – CG iteration for the proposed method and SQS iteration for the exhaustive search (assuming pre-computed curvatures in SQS). However, the proposed method tends to require fewer iterations, because the objective function of the P+C estimate is quadratic and thereby easier to solve than the objective function in PIRPLE reconstruction which is not quadratic (not even guaranteed to be concave). The computation time of both methods are compared in the Results Section.

Both the PIRPLE reconstruction and image estimates were implemented in Matlab (The Mathworks, Natick MA), with the projection operations executed on GPU using CUDA-based libraries. The projection operations were implemented based on separable footprints.<sup>98</sup> All experiments were performed on a workstation equipped with one GeForce GTX TITAN (Nvidia, Santa Clara CA) graphics card.

## **6.3 Results**

### **6.3.1 Evaluation of approximate analytical solution**

The proposed approximate analytical solution was evaluated in the thorax phantom by introducing a uniform disc emulating a lung nodule in the center of the lung as shown in Fig. 6.2(b). The uniform disc had a radius of 6 mm and attenuation of  $0.021 \text{ mm}^{-1}$  (i.e., 50 HU assuming  $0.02 \text{ mm}^{-1}$  water attenuation), which is typical values for a solid solitary pulmonary nodule.<sup>254</sup> This experiment assumed that the actual change could be exactly anticipated in the prospective

regularization design, so the presumed change used in the P+C estimate was set to be the same as the actual change. Figure 6.3 illustrates the PIRPLE reconstructions and image estimates using a number of  $\beta_P$  values. In the top row, the nodule reconstructed by PIRPLE exhibited low resolution and a high level of noise when using a very low  $\beta_P$  (due to high angular undersampling in the projection data) while it began to shrink or even disappear when using a very high  $\beta_P$ , demonstrating the importance of using appropriate prior image strength in PIBR. The Ideal estimate exhibited a high level of agreement with the PIRPLE reconstruction for all  $\beta_P$ , suggesting high accuracy in the quadratic approximation of the nonquadratic regularization given an ideal operating point. However, the Ideal estimate still requires a full PIRPLE reconstruction, which supersedes the need for an approximate solution. The P+C estimate did not use PIRPLE reconstruction results but exhibited some level of agreement with the PIRPLE reconstruction after one iteration of the P+C estimate and a high level of agreement after five iterations of the P+C estimate. These results suggest the possibility of approximating PIRPLE results without performing full image reconstructions.

The computation time of traditional exhaustive search (PIRPLE reconstruction) and the proposed method (P+C estimate) were also compared. For this dataset, PIRPLE reconstruction required 2000 SQS iterations to obtain a nearly converged image (RMSE less than 1 HU compared to a PIRPLE image formed using more than 20000 SQS iterations), while the P+C estimate only required 500 CG iterations to obtain a nearly converged image (similarly, RMSE less than 1 HU compared to a P+C estimate using more than 5000 CG iterations). The total computation time was  $\sim 180$  seconds for a PIRPLE reconstruction and  $\sim 10$  seconds for a P+C estimate. This suggests that the proposed method can reduce the computation time required for finding the optimal prior image strength by over an order of magnitude, compared to traditional exhaustive search.

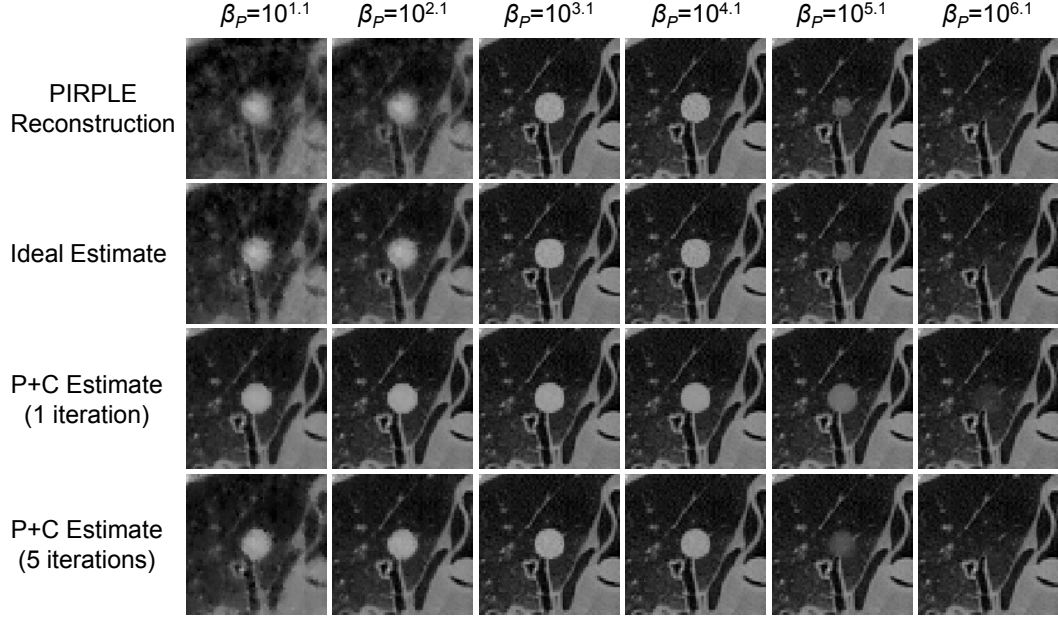


Figure 6.3: PIRPLE reconstruction, Ideal estimate, P+C estimate (1 iteration), and P+C estimate (5 iterations) of a simulated circular solitary pulmonary nodule at various  $\beta_P$ . Note that too small  $\beta_P$  or too large  $\beta_P$  leads to little benefits or false features (false negative) in PIRPLE reconstruction. Only 20 projections equally distributed over  $190^\circ$  were used. (Grayscale window:  $[0 \ 0.04] \text{ mm}^{-1}$ .)

### 6.3.2 Evaluation of predictive performance metric

The qualitative comparisons from the previous section are substantiated with quantitative measures in this section. Specifically, Figure 6.4 illustrates the evaluation of the proposed predictive performance metric at different  $\beta_P$  (with a uniform log spacing of  $10^{0.1}$ ) using either PIRPLE reconstruction or one of the image estimate methods and the previously described experimental setup. The proposed metric using PIRPLE reconstruction exhibited a single well-defined minimum in the range of  $\beta_P$  ( $10^0$  to  $10^7$ ). The minimizer ( $\beta_P = 10^{3.3}$ ) from this method was used as the ground truth for the optimal  $\beta_P$ . The metric using the Ideal estimate closely approximated the results using PIRPLE reconstructions, especially in the region where  $\beta_P$  was greater than  $10^3$ , and estimated the same optimum as the ground truth. This is consistent with the qualitative results in the previous section. When a (single iteration) P+C estimate was used, the metric plot yielded a similar shape and predicted a minimizer ( $10^{3.4}$ ) very close to the ground truth. Using five iterations of the P+C

estimate moved the curve much closer to the curve of PIRPLE reconstruction and predicted the same optimum as the ground truth. These results suggest that the proposed method can yield the same or very similar optimal  $\beta_P$  while having a significant computational advantage over traditional exhaustive search. The well-defined optimum in all plots – including the P+C estimates – suggests that direct minimization of the metric in Eq. (6.8) using more sophisticated optimization approaches may be possible and offer additional computational speedups.

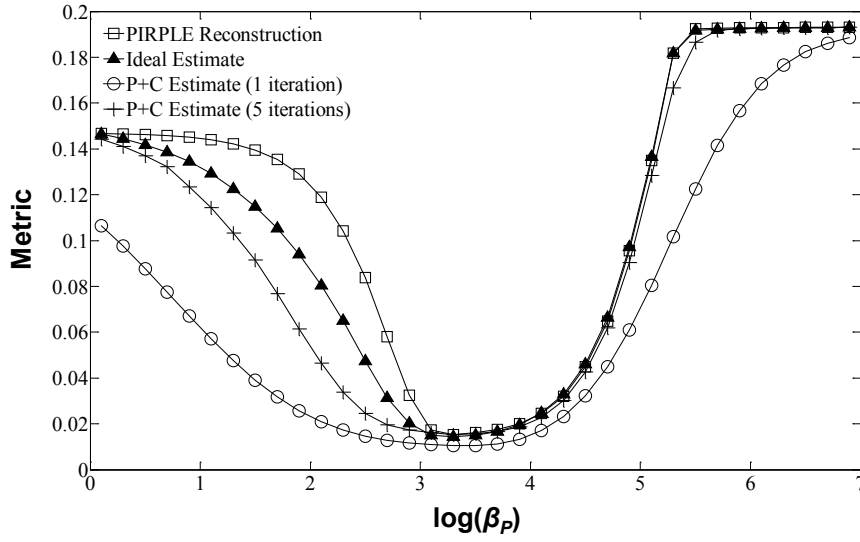


Figure 6.4: Evaluation of the proposed predictive performance metric at different  $\beta_P$  (with a uniform log spacing of  $10^{0.1}$ ) using either PIRPLE reconstruction or one of the image estimation methods. Note that all the methods yielded almost the same optimal  $\beta_P$ , while the proposed method (using P+C estimate) does not require full image reconstruction.

### 6.3.3 Location dependence of regularization design in PIBR and evaluation of spatially varying prior image strength map

Thus far, the location of the anatomical change is assumed to be known exactly. It remains a key question whether an optimal  $\beta_P$  designed for the presumed location remains optimal at other locations. To answer this question, the same change was introduced to either of the two locations in the ellipse phantom as shown in Fig. 6.2(a). The change was a disc with  $0.05 \text{ mm}^{-1}$  attenuation

and 10 mm radius. Figure 6.5 shows the PIRPLE reconstructions and the predictive performance metric values at various  $\beta_P$  for the two locations. A clear difference in the optimal  $\beta_P$  for the two locations can be seen both in the PIRPLE reconstructions (optimal images outlined in black box) and in the metric curves. The optimal  $\beta_P$  was  $10^{3.2}$  for Location I and  $10^{4.5}$  for Location II, which differed over an order of magnitude. As a result, suboptimal reconstruction of the change could be seen at each of the two locations when using a  $\beta_P$  optimized for the other location, as seen in Fig. 6.5(a). Interestingly, the optimal  $\beta_P$  was lower in Location I (higher attenuation area) and higher in Location II (lower attenuation area). This may be interpreted by recognizing the role of  $\beta_P$  in balancing the data fit term with the prior image penalty term in the PIRPLE objective function. The rays that go through Location II tend to have higher fluence than those that go through Location I, thereby leading to a larger value of the data fit term at Location II than at Location I. To maintain the same optimal balance between the two terms, a higher  $\beta_P$  is then needed at Location II than at Location I. Note also that the proposed method (using the P+C estimate after one iteration) predicted the same optimal  $\beta_P$  as the ground truth (using PIRPLE reconstruction) at both locations, again demonstrating the accuracy of the proposed method in estimating the optimal  $\beta_P$ .

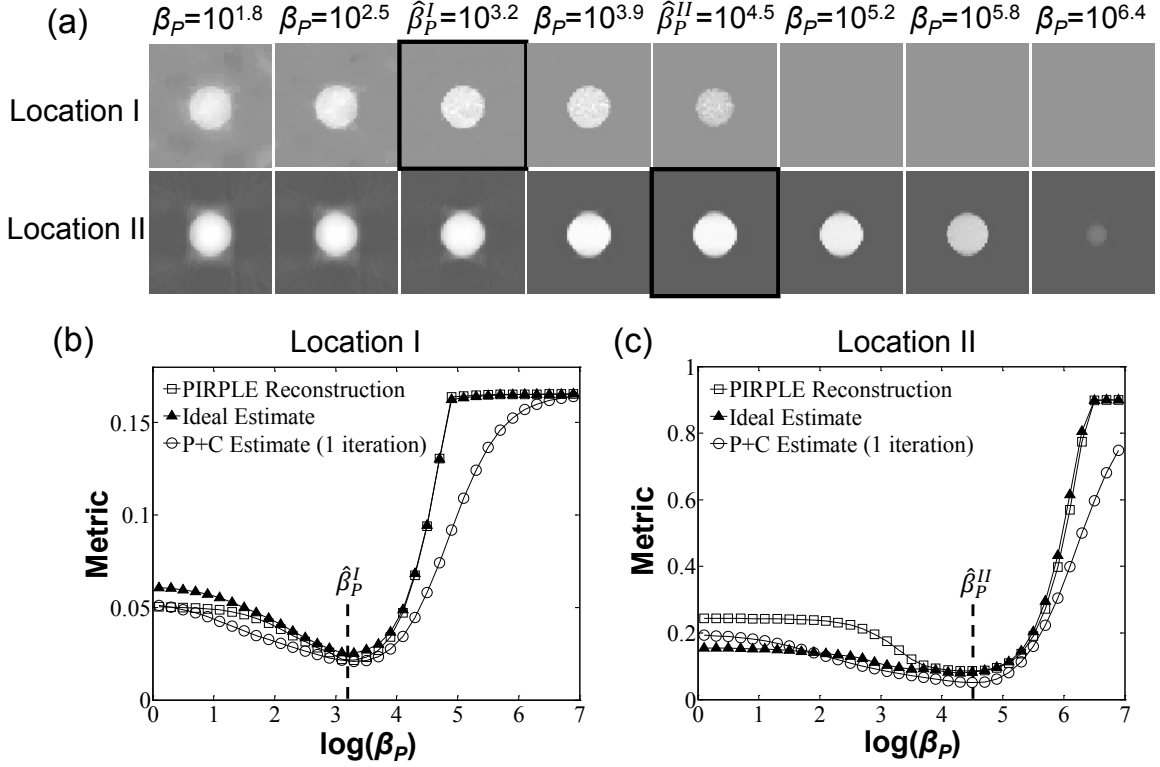


Figure 6.5: Comparison of PIRPLE reconstructions and the prospective performance metric for the same change at two different locations. A clear difference in the optimal  $\beta_P$  (over an order of magnitude) between the two locations can be seen both in the PIRPLE reconstructions and in the metric curves. These results motivate the design of a spatially varying  $\beta_P$  map. (Grayscale window:  $[0.03 \ 0.052] \text{ mm}^{-1}$  for Location I and  $[-0.02 \ 0.052] \text{ mm}^{-1}$  for Location II.)

Because one may not generally know where a change might occur, the location-dependent prior image strength was also investigated. Specifically, a spatially varying  $\beta_P$  map for the circular change mentioned above was generated by estimating the optimal  $\beta_P$  at each grid point of the image (with a spacing of 20 voxels in each dimension and a total number of 229 grid points) and interpolating across the image using radial basis functions.<sup>189</sup> Figure 6.6(a) shows a ground truth  $\beta_P$  map whose optimal  $\beta_P$  at each grid point was estimated using traditional exhaustive search. Spatial variations of the optimal  $\beta_P$  for the same change are seen throughout the image. Figure 6.6(b) shows a  $\beta_P$  map whose optimal  $\beta_P$  at each grid point was estimated by the proposed method. This map shows good agreement with ground truth, (RMSE =  $10^{0.18}$ ) compared to the range of the optimal  $\beta_P$  within each map ( $10^{3.2} \sim 10^{4.6}$ ). The total computation time for a  $\beta_P$  map was  $\sim 15$  hours using

traditional exhaustive search and below 1 hour using the proposed method, demonstrating the computational advantage of prospective design.

Figure 6.6(c) shows PIRPLE reconstructions of the same change at both locations using either a scalar  $\beta_P$  or a spatially varying  $\beta_P$  map. PIRPLE reconstruction using a scalar  $\beta_P$  optimized for Location I ( $\hat{\beta}_P^I = 10^{3.2}$ ) resulted in accurate reconstruction of the change in Location I but streaks and low resolution for the change in Location II (especially visible in difference images). Similarly, PIRPLE reconstruction using a scalar  $\beta_P$  optimized for Location II ( $\hat{\beta}_P^I = 10^{4.5}$ ) lead to accurate reconstruction of the change in Location II but a change with incorrect size in Location I. Whereas PIRPLE reconstruction using the spatially varying  $\beta_P$  map in Fig. 6.6(b) resulted in accurate reconstruction of both changes.

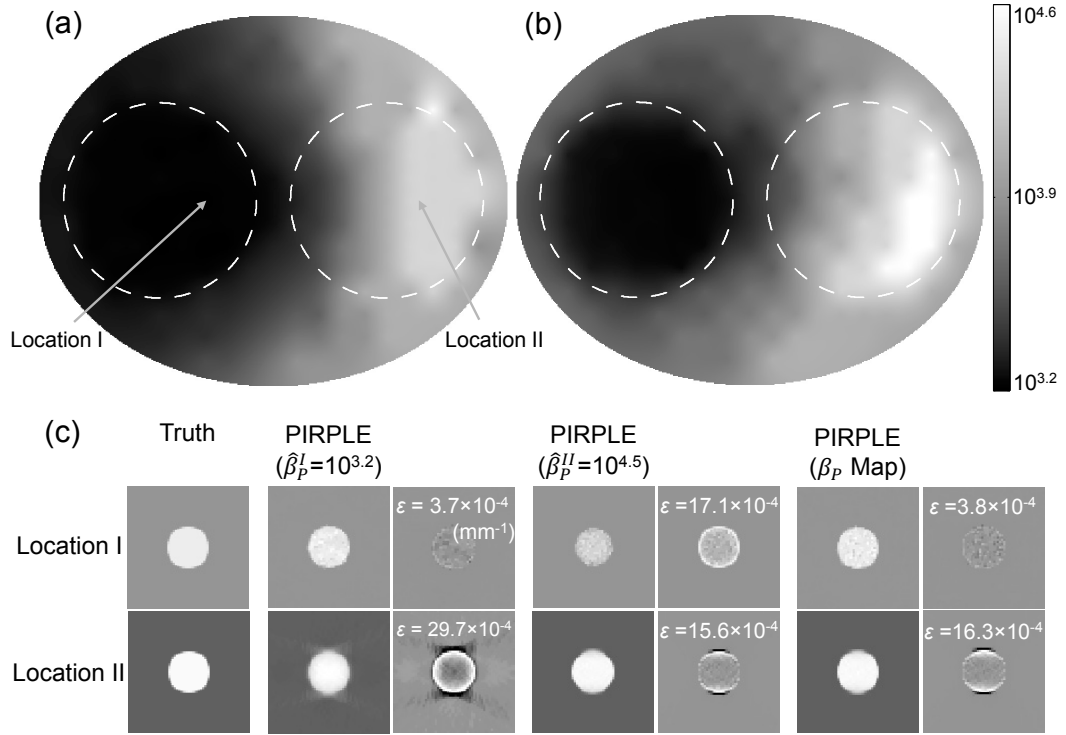


Figure 6.6: (a-b) Spatially varying  $\beta_P$  map generated using traditional exhaustive search (a) or the proposed method (b). The  $\beta_P$  map in (b) exhibits good agreement with the  $\beta_P$  map in (a). (c) PIRPLE reconstruction of the same change at both locations using a scalar  $\beta_P$  optimized for Location I ( $\hat{\beta}_P^I = 10^{3.2}$ ), a scalar  $\beta_P$  optimized for Location II ( $\hat{\beta}_P^I = 10^{4.5}$ ), and the spatially varying  $\beta_P$  map in (b). The parameter  $\varepsilon$  stands for the RMSE of the PIRPLE reconstruction with respect to the truth image. (Grayscale window:  $[0.03 \ 0.052]$  mm<sup>-1</sup> for reconstructed images at Location I,  $[-0.02 \ 0.052]$  mm<sup>-1</sup> for reconstructed images at Location II, and  $[-0.01 \ 0.01]$  mm<sup>-1</sup> for all the difference images.)

### 6.3.4 Attenuation, shape, and size dependence of regularization design in PIBR

The dependence of the optimal  $\beta_P$  on other properties of the anatomical change (besides location) including attenuation, shape, and size was also investigated. The ellipse phantom in Fig. 6.2(a) was used, and a change was introduced to Location II in the subsequent scan. The value of only one of the three properties was varied at a time. The proposed metric was evaluated at different  $\beta_P$  with a uniform log spacing of  $10^{0.01}$ . Since the optimal  $\beta_P$  estimated by the traditional exhaustive search and the proposed method were almost the same, only the optimal  $\beta_P$  estimated by the traditional exhaustive search are shown below.

First, the attenuation of the change was varied from  $0.004 \text{ mm}^{-1}$  to  $0.060 \text{ mm}^{-1}$  (i.e., -800 HU to 2000 HU assuming  $0.02 \text{ mm}^{-1}$  water attenuation) with an increment of  $0.002 \text{ mm}^{-1}$ , covering a broad range of possible changes – e.g. low-attenuating pulmonary ground-glass nodules (about -700 HU)<sup>255</sup> to high-attenuating bones. The shape and size of the change were fixed to a 10 mm radius disc. Figure 6.7(a) shows the estimated optimal  $\beta_P$  as a function of the attenuation of the change. Note that the optimal  $\beta_P$  increased consistently as the attenuation increased, and the rate of the increase was higher for low attenuation changes (e.g., a change in soft tissue) and lower for high attenuation changes (e.g., a change in bone). This indicates a strong dependence of the optimal  $\beta_P$  on the attenuation of the change especially for low attenuation changes. The difference in the dependence of the optimal  $\beta_P$  between low attenuation changes and high attenuation changes may be explained by recognizing the effect of the use of too large  $\beta_P$  – that is, the use of too large  $\beta_P$  will enforce the reconstructed image to simply replicate the prior image, which prevents the change from being reconstructed in the image. Compared to high attenuation changes, low attenuation changes tend to be more vulnerable to such effect because they are more similar to the prior image.

Therefore, their optimal  $\beta_P$  values have a stronger dependence on the attenuation of the change than high attenuation changes.

Shape of the change was varied to simulate different levels of morphologic irregularity (e.g. tumor speculation). Specifically, shape was varied by modeling anatomical changes with a shape whose radius varied as a function of angle using a sinusoid plus a constant. The amplitude of the sinusoid was varied from 0 mm (a circular disc) to 9 mm (highly spiculated) with an increment of 1 mm. Change attenuation was fixed at  $0.02 \text{ mm}^{-1}$  (e.g., soft tissue), and the size was fixed such that the area of the change was  $770 \pm 1 \text{ mm}^2$  for every selected amplitude (This was achieved by tuning the mean radius constant in the shape model). Figure 6.7(b) shows that optimal  $\beta_P$  values exhibited only small variations as the amplitude of the sinusoidal contour increased, indicating a low dependence of optimal  $\beta_P$  on the shape of the change.

Lastly, the dependence on the size of the change was studied by varying the radius of a circular change from 3 mm to 20 mm with an increment of 1 mm. The attenuation was fixed to be  $0.02 \text{ mm}^{-1}$  (e.g., soft tissue). Figure 6.7(c) shows that the optimal  $\beta_P$  exhibited only small variations as the radius of the change increased, indicating a weak dependence of the optimal  $\beta_P$  on the size of the change.

The experiments on the shape and size of the change were also performed with respect to a high attenuation change ( $0.05 \text{ mm}^{-1}$  attenuation), in which similarly weak dependence was observed. The results in this section together suggest that, when performing prospective regularization design, one may need to make sure that the attenuation of the presumed change is consistent with that of the actual change while such consistency may not need to be strictly enforced for the shape or size of the presumed change.

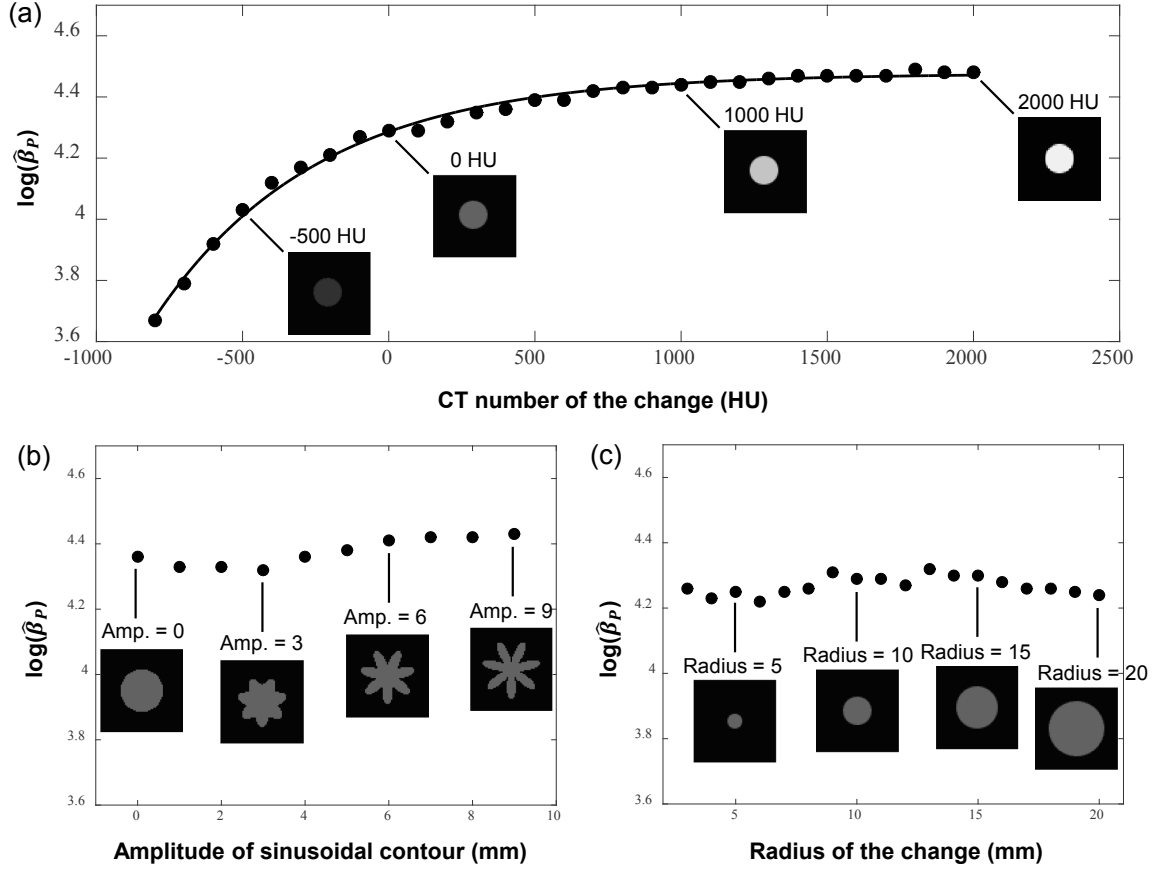


Figure 6.7: Illustration of the dependence of the optimal  $\beta_P$  on the attenuation (a), shape (b), and size (c) of the anatomical change. One anatomical change was introduced to Location II of the ellipse phantom in Fig. 6.2(a), and only one of the three properties of the change mentioned above was varied at a time. The optimal  $\beta_P$  was estimated by evaluating the proposed metric at different  $\beta_P$  with a uniform log spacing of  $10^{0.01}$ . A negative exponential function was fit to the data points in (a) to help illustrate the relationship in (a). (Grayscale window:  $[0 \ 0.052] \text{ mm}^{-1}$ .)

### 6.3.5 Evaluation of regularization design in lung nodule surveillance

Prospective regularization design was applied in a lung nodule surveillance scenario with a solitary pulmonary nodule. The nodule was not present in a previous baseline exam but is present in a subsequent exam. The thorax phantom in Fig. 6.2(b) was used in this study and with a 6 mm radius nodule and an attenuation of  $0.021 \text{ mm}^{-1}$  (i.e., 50 HU). Presuming an unknown location, a spatially varying  $\beta_P$  map was generated using the method described in Sec. 6.2.5. Figure 6.8(a) shows a ground truth  $\beta_P$  map (generated using traditional exhaustive search) which exhibited a shift-variant

optimal  $\beta_P$  within each side of the thoracic cavity as well as between sides. In this specific case, the optimal  $\beta_P$  was higher in the right cavity than in the left cavity, which is due to the asymmetry in the anatomy between the two sides. For example, the heart and trachea were not exactly centered and the lung in the right cavity had collapsed. The optimal  $\beta_P$  spanned almost an order of magnitude across the image ( $10^{2.8} \sim 10^{3.6}$ ). Figure 6.8(b) shows a  $\beta_P$  map generated using the proposed method (including P+C estimate after one iteration), exhibiting good agreement with the ground truth  $\beta_P$  map and at the same time achieving about  $\times 20$  reduction in computation time.

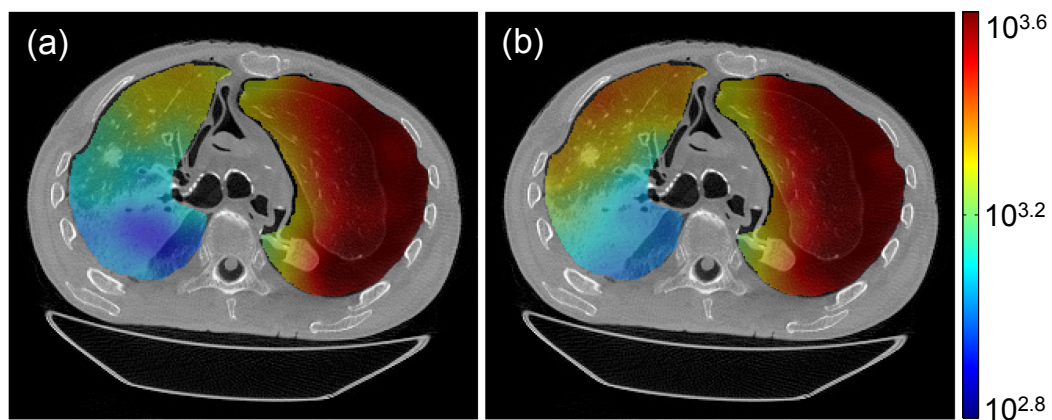


Figure 6.8: Spatially varying  $\beta_P$  map generated using traditional exhaustive search (a) or the proposed method (b) for optimally admitting a solitary pulmonary nodule everywhere in both sides of the thorax cavity. The optimal  $\beta_P$  was estimated on an image grid with a spacing of 20 voxels in each dimension and then interpolated into a  $\beta_P$  map using radial basis functions. The  $\beta_P$  map in (b) exhibited good agreement with the  $\beta_P$  map in (a) while reducing the computation time by a factor of  $\sim 20$ .

The spatially varying  $\beta_P$  map in Fig. 6.8(b) was then used in PIRPLE to reconstruct the actual solitary pulmonary nodule, which was simulated in the periphery of the right lung in the subsequent scan as shown in Fig. 6.2(b). The optimal  $\beta_P$  for a nodule at this location was found to be  $10^{3.5}$  (using uniform log spacing of  $10^{0.1}$ ). Figure 6.9(a-d) shows a ROI of the current anatomy and images reconstructed by FBP, PIRPLE using a suboptimal scalar  $\beta_P$ , and PIRPLE using the spatially varying  $\beta_P$  map. The suboptimal scalar  $\beta_P$  was chosen to be  $10^{2.8}$ , which was optimal for the same nodule at the posterior of the left lung but was not optimal for the true change location. FBP image exhibited a high level of streaks and noise as a result of severe angular undersampling

and lack of support from prior image information ( $\text{RMSE} = 43.7 \times 10^{-4} \text{ mm}^{-1}$ ). PIRPLE image using a suboptimal scalar  $\beta_P$  substantially reduced the streaks and noise, but still exhibited apparent error in the reconstructed nodule especially in the boundary of the nodule (error more pronounced in the difference image) ( $\text{RMSE} = 9.2 \times 10^{-4} \text{ mm}^{-1}$ ). Finally, PIRPLE image using the  $\beta_P$  map exhibited excellent reconstruction of the nodule and the lowest error among all three methods ( $\text{RMSE} = 2.8 \times 10^{-4} \text{ mm}^{-1}$ ).

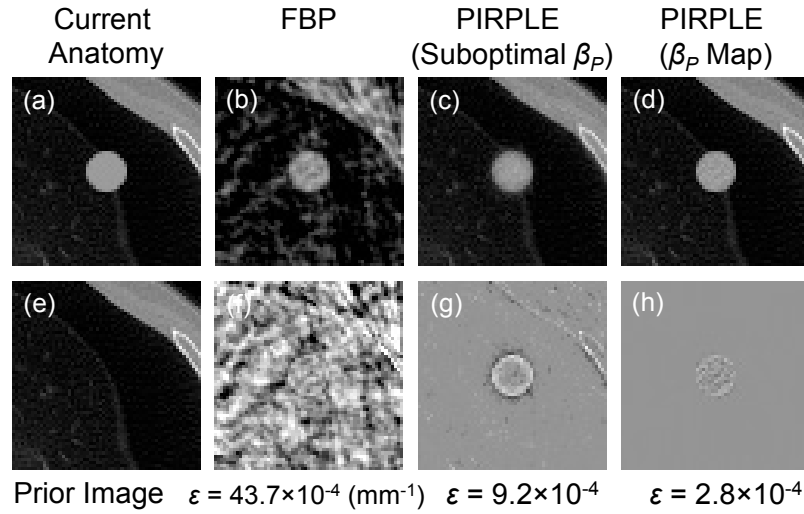


Figure 6.9: Image reconstruction of a solitary pulmonary nodule which was not present in a baseline exam (e) but appeared in the periphery of the right lung in the subsequent exam (a). (b-d) A ROI of the images reconstructed by FBP, PIRPLE using a suboptimal scalar  $\beta_P$ , and PIRPLE using the spatially varying  $\beta_P$  map in Fig. 6.8(b). (f-h) Difference image between each of the images in (b-d) and the current anatomy. The parameter  $\epsilon$  stands for the RMSE of the difference image in (f-h). (Grayscale window:  $[0 \ 0.04] \text{ mm}^{-1}$  for reconstructed images and  $[-0.01 \ 0.01] \text{ mm}^{-1}$  for difference images.)

### 6.3.6 Evaluation of regularization design in a nodule disappearance scenario

The proposed regularization design method was also evaluated in a scenario in which a nodule was present in the prior image but not present in the current anatomy. This scenario is common in lung nodule surveillance, since benign nodules found in a previous exam can often be naturally resolved by the body before a follow-up exam is performed. Similar types of anatomical change can also be

found in radiation therapy, in which tumor shrinks or disappears in follow-up exams after successful treatment. A simulation study was carried out on the thorax phantom, which was the same as the study described in Sec. 6.3.1 except that the lung nodule was present in the prior image but not in the current anatomy. Figure 6.10(a) illustrates the truth image and PIRPLE reconstructions using a number of  $\beta_P$  values. PIRPLE exhibited higher spatial resolution and a lower level of noise as  $\beta_P$  increased from  $10^{1.5}$  to  $10^{3.5}$ , but exhibited features (false positive) as  $\beta_P$  kept increasing, indicating the importance of using proper prior image strength. Note that the false positive nodule information appeared first at the edge of the nodule and then at the interior of the nodule in the PIRPLE image. Figure 6.10(b) shows that the proposed metric using PIRPLE reconstruction still exhibited a single well-defined minimum (ground truth) in the nodule disappearance scenario. Moreover, a (single iteration) P+C estimate was still able to predict a minimizer very close to the ground truth. These results suggest that accurate estimation of the optimal prior image strength with substantial computational speedup can also be achieved in the nodule disappearance scenario.

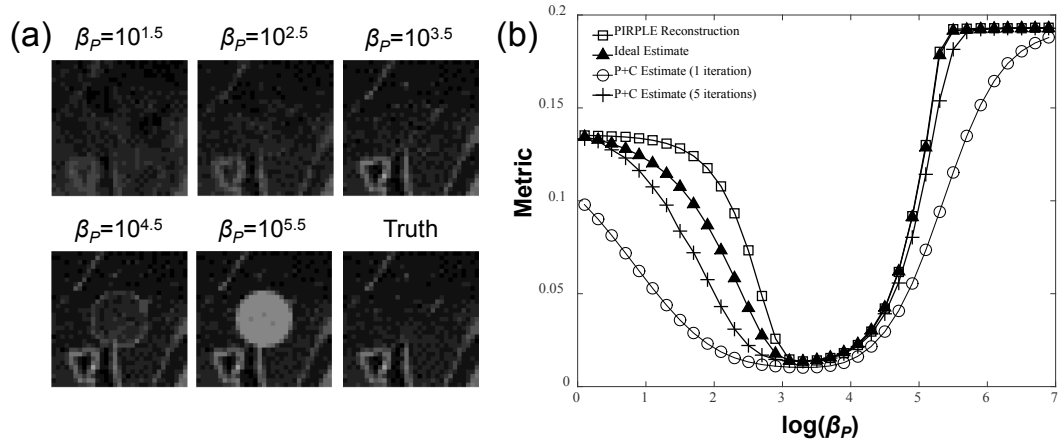


Figure 6.10: (a) PIRPLE reconstruction of the current anatomy (labeled as “Truth”) at various  $\beta_P$ . A lung nodule was present in the prior image but not present in the current anatomy. Note that too small  $\beta_P$  or too large  $\beta_P$  leads to little benefits or false features (false positive) in PIRPLE reconstruction. (Grayscale window:  $[0 \ 0.04] \text{ mm}^{-1}$ ) (b) Evaluation of the proposed metric at different  $\beta_P$  in the nodule disappearance scenario. Note that all the methods yielded almost the same optimal  $\beta_P$ , while the proposed method (P+C estimate) does not require full image reconstruction.

## 6.4 Conclusions and discussion

A novel method has been proposed that prospectively estimates the optimal prior image strength for accurate reconstruction of anatomical changes in PIBR. The approach uses an analytical approximation of PIBR objective functions and a predictive performance metric that leverages knowledge of a presumed change to estimate prior image strength that ensures accurate reconstruction of the change. The prospective regularization strategy yields accurate estimates of the optimal prior image strength and substantially reduces computational time (by a factor of 20) compared to traditional exhaustive search.

A spatially varying prior image strength "map" was also introduced to optimally admit a presumed change everywhere in the image and eliminate the need to know the location of the change *a priori*. Optimal prior image strength was found to vary by at least an order of magnitude throughout the volume in phantom studies, indicating the potential importance of the spatially varying design. The optimal prior image strength was found to vary significantly with the attenuation difference associated with the anatomical change but was relatively insensitive to the shape and size of the change, suggesting accurate specification of change attenuation is important in regularization design in PIBR. Optimal penalty maps were found to improve the accuracy of lung nodule PIBR over uniformly penalized reconstructions. These results suggest great potential for the proposed method to provide prospective patient-, change-, and data-specific customization of the prior image strength to ensure reliable reconstruction of specific anatomical changes. While investigation in this chapter has concentrated on the PIRPLE approach, one might form analogous regularization strength maps for other approaches that require a balance between current imaging data and prior image data. This includes other prior-image-based reconstruction (e.g. PICCS) as well as reconstruction of difference approaches.<sup>252,256</sup>

While the work reported in this chapter provides a general strategy for prospective regularization design in PIBR, there are a number of potential developments that could further

increase the utility of the underlying methodology. First, the predictive performance metric was solved in this chapter by evaluating the metric at different  $\beta_P$  values with regular spacing. While faster than traditional exhaustive search, additional acceleration may be found via more sophisticated minimization methods (e.g. simplex method, etc.). Such directed searches will be more computationally efficient and would be increasingly important for computing spatially varying maps with larger fields-of-view.

A second topic that needs to be investigated is the incorporation of image registration into the regularization design. Patient motion is commonly present between scans (e.g., due to patient re-positioning or respiratory/cardiac motion) and needs to be accommodated to ensure accurate use of the prior image information in PIBR. For the purpose of regularization design, one might adopt a two-step approach – first estimating the patient motion by performing one PIRPLE (for rigid motion) or dPIRPLE (for nonrigid motion) reconstruction using a nominal  $\beta_P$  value or using a dedicated image registration method such as the one by Otake *et al.*,<sup>240</sup> and then performing the proposed regularization design method with a prior image that has been deformed for motion compensation. While one might use the regularization design proposed in this chapter using the presumed change model with a perfectly registered prior image, the sensitivity of regularization design to registration errors also needs to be assessed.

The predictive performance metric proposed in this chapter estimates the prior image strength that minimizes the RMSE of the approximate analytical solution from the prior image plus the anticipated change. While this metric can yield image reconstructions with overall high accuracy, other metrics that are sensitive to specific imaging tasks including detectability index and various observer models<sup>136,172,173,175,257</sup> should also be considered. Such task-based metrics could be used to find optimal prior image strength for particular abnormalities. For example, detectability index<sup>136</sup> may be computed using the approximate analytical solution to estimate the optimal prior image strength for detecting high-contrast, low-frequency lesions such as a solitary pulmonary nodule in lung nodule surveillance.

While the work in this chapter focused on anatomical change within a lung nodule surveillance scenario, there are many other potential applications of optimized PIBR regularization including accurate visualization of resections, device/implant placement, and monitoring of other treatments and interventions. While each of these potential applications deserves additional investigations into the specific challenges associated with PIBR in each area, the proposed methodology for balancing prior image information with measurement data is general. Optimization of this balance is critical for reliable reconstruction and this work represents an important step in providing a degree of robustness and controllability for PIBR approaches. Such reliability is a necessity for delivering on the huge potential of PIBR methods and finding more widespread clinical adoption.

## Chapter 7

### 7. Conclusions

This dissertation has investigated the potential of advanced MBIR methods for improving image quality and reducing radiation dose in CBCT. A number of novel MBIR methods were developed that leverage advanced models of imaging physics, task-based assessment of imaging performance, and incorporation of patient-specific anatomical information from previously acquired images. The approaches uncovered in this work demonstrate substantial improvements in CBCT image quality (especially low-contrast, soft-tissue image quality) and reduction in radiation dose in applications ranging from detection of acute ICH to surveillance of lung nodules. Together, the findings support the proposed thesis statement and goal of this work:

Model-based iterative reconstruction (MBIR) methods can improve image quality and/or reduce radiation dose in cone-beam computed tomography (CBCT) through modeling of imaging physics, development of advanced regularization techniques, and incorporation of patient-specific prior image information.

A major theme of the presented work was the investigation of new approaches to overcome conventional limitations in image quality and dose efficiency in CBCT. MBIR represents a promising software-based approach to achieve this goal that can be implemented in complement to hardware-based approaches such as new detector technology and system design. One powerful feature of MBIR is the inclusion of a comprehensive image reconstruction framework that allows the incorporation of a wealth of information regarding imaging physics, patient anatomy, and the imaging task associated with a particular procedure. Moreover, the overall framework developed

in this dissertation offers a high degree of flexibility to allow the incorporation of different types of physical models and/or prior information in various applications, ranging from point-of-care detection of emergent head trauma to longitudinal screening / surveillance of lung nodules.

## 7.1 Contributions

Specific contributions to the field of medical imaging in this dissertation are summarized as follows.

**Chapter 2** reported a novel PWLS image reconstruction method that incorporates a general framework for accurately accommodating modified noise models for artifact-corrected CBCT data. A specific scenario was considered in which the two dominant artifact correction methods essential to high-quality CBCT of the head – scatter and beam hardening corrections – are applied. The noise amplification imparted by each of the two correction steps were modeled in analytical expressions and compensated by modifying the statistical weights in the resulting PWLS reconstruction method. An anthropomorphic head phantom emulating ICH was built and scanned on a CBCT test-bench equipped with a flat-panel detector. The physical experiments illustrated a high degree of changes in variance (more than two orders of magnitude) after artifact corrections, primarily attributed to scatter correction. The proposed PWLS method demonstrated superior noise-resolution tradeoffs in comparison to FBP and PWLS with conventional statistical weights (viz., at matched 0.50 mm spatial resolution, the contrast-to-noise ratio was 11.9 for the proposed method, compared to 5.6 and 9.9 for FBP and conventional PWLS, respectively). The method was particularly advantageous in regions such as skull base, where scatter and beam-hardening effects tend to be greatest. The results support the hypothesis that with high-fidelity artifact correction and MBIR using an accurate post-artifact-correction noise model, CBCT can provide soft-tissue image quality suitable to reliable detection of ICH. The work supported the development of a dedicated system for point-of-care diagnosis of ICH currently in translation to first clinical studies.

**Chapter 3** proposed a regularization approach in which a spatially varying penalty was determined that maximizes task-based imaging performance at every location in a 3D image. This approach was applied to the PWLS reconstruction method developed in Chapter 2 for high-quality CBCT of the head, focusing on the task of detecting a small, low-contrast ICH. Modeling of local spatial resolution and noise was performed via a predictive framework by which regularization (specifically, the quadratic penalty strength) was varied throughout the image to maximize local task-based detectability index ( $d'$ ). Simulations and test-bench experiments showed that conventional (constant) regularization exhibited a fairly strong degree of spatial variation in  $d'$ , and the task-based method improved detectability by up to ~15%. The improvement was strongest in areas of high attenuation (skull base), where other methods tended to over-smooth the data. The proposed task-based method presents a promising means to improve imaging performance in MBIR, and together with the PWLS method proposed in the previous chapter, could support the development of high-quality CBCT systems for new applications requiring high image quality with respect to specific imaging tasks.

**Chapter 4** continued the investigation on high-quality CBCT of the head and reported a multi-resolution MBIR approach to mitigate truncation effects for which the main source of truncation is the patient support / head holder. While conventional reconstruction methods employ a fixed voxel size throughout the image, the multi-resolution approach uses a fine voxel size within the untruncated region (i.e., inside the scan FOV encompassing the area of interest – in this case, the head) and a coarse voxel size in the truncated region (i.e., outside the scan FOV). The multi-resolution approach allows extension of the reconstruction FOV to mitigate truncation effects without significantly increasing computational complexity of MBIR. The approach was implemented in the PWLS reconstruction framework developed in Chapter 2. Experiments involving an anthropomorphic head phantom with truncation due to a carbon-fiber holder were shown to result in severe artifacts in conventional single-resolution PWLS,

whereas extending the reconstruction FOV strongly reduced truncation artifacts. For the same extended reconstruction FOV, the multi-resolution approach reduced computation time compared to the single-resolution approach (viz., time reduced by 40.7%, 83.0%, and over 95% for an image volume of  $600^3$ ,  $800^3$ ,  $1000^3$  voxels). Algorithm parameters (e.g., regularization strength, the ratio of the fine and coarse voxel size, and reconstruction FOV size) were investigated to guide reliable parameter selection. The method further supports translation of a newly developed CBCT head scanner in point-of-care imaging applications. It is also compatible with the methods reported in Chapters 2 and 3.

**Chapter 5** proposed a MBIR method that leverages patient-specific anatomical information from previously acquired images. The proposed framework jointly estimates the 3D deformation between an unregistered prior image and the current anatomy and reconstructs the current anatomical image using a MBIR approach that includes regularization based on the deformed prior image. This framework is referred to as deformable Prior Image Registration, Penalized-Likelihood Estimation (dPIRPLE). Central to this framework is the inclusion of a 3D B-spline-based free-form-deformation model into the joint framework, which is solved using an alternating maximization strategy allowing for improvements in both the registration and reconstruction throughout the optimization process. Cadaver experiments were conducted on a CBCT test-bench emulating a lung nodule surveillance scenario. The dPIRPLE algorithm demonstrated superior reconstruction accuracy and image quality compared to more traditional reconstruction methods, including FBP, penalized-likelihood estimation, prior image penalized-likelihood estimation without registration, and prior image penalized-likelihood estimation with rigid registration over a wide range of sampling sparsity and exposure levels. This work demonstrates the strong potential of leveraging patient-specific prior image information into MBIR for the reduction of accumulative radiation dose in sequential imaging studies (e.g., an 18-fold exposure reduction in the cadaver experiments).

**Chapter 6** continued the investigation of prior-image-based reconstruction (PIBR) and proposed a method that prospectively estimates the optimal amount of prior image information for accurate admission of specific anatomical changes in PIBR without performing full image reconstructions. This method leverages an analytical approximation to the implicitly defined PIBR estimator proposed in Chapter 5, and introduces a predictive performance metric leveraging this analytical form and knowledge of a particular presumed anatomical change whose accurate reconstruction is sought. Additionally, a map of spatially varying prior image strength is proposed to optimally admit changes everywhere in the image, eliminating the need to know the locations of structural changes *a priori*. Studies were conducted in both an ellipse phantom and a realistic thorax phantom emulating a lung nodule surveillance scenario. The proposed method demonstrated accurate estimation of the optimal prior image strength while achieving a substantial computational speedup (about a factor of 20) compared to traditional exhaustive search. Moreover, in phantoms where the optimal prior image strength varied spatially by an order of magnitude or more, the use of the proposed prior image regularization strength “map” demonstrated accurate reconstruction of anatomical changes without foreknowledge of the location of structural change. Furthermore, the optimal prior image regularization strength was found to vary with attenuation differences associated with anatomical change but exhibited only small variations as a function of the shape and size of the change. The results suggest that, given a target value of change in attenuation, prospective patient-specific, change-specific, and data-specific customization of the prior image regularization strength can be performed to ensure reliable reconstruction of specific anatomical changes.

## 7.2 Future work

The potential of MBIR in improving image quality and reducing radiation dose in CBCT brings numerous opportunities in translating the technology to clinical use and also invites efforts to address a number of important practical challenges. Among the opportunities is the translation of high-quality CBCT of the head to point-of-care environments such as the neuroscience critical care unit (NCCU). The MBIR methods in Chapters 2-4 demonstrated a level of image quality that appears to be sufficient for detecting acute ICH, and the imaging performance in clinical pilot studies is underway at the time of writing. Among the remaining challenges is that of patient motion, which invites future work on motion compensation and its integration into the MBIR framework.<sup>258,259</sup> Another practical challenge is the artifacts caused by the presence of highly attenuating metal objects (e.g., surgical tools or implants) in the scan FOV, which is fairly common for patients in the NCCU and may obscure ICH detection. This challenge invites future work on metal artifact correction<sup>100,101,216,260</sup> either as a separate process (whose effects on noise amplification may be compensated for using the general approach in Chapter 2) or as a part of the MBIR framework (e.g., by modeling the polyenergetic nature of x-rays).

Closely related to the translation of high-quality CBCT to clinical application is the computation time associated with MBIR. Thanks to recent advances in accelerated algorithms for MBIR and computing power (e.g., GPU), a converged MBIR image in CBCT can be obtained in ~2 minutes using hardware and software acceleration.<sup>168,191</sup> The feasibility of the workflow associated with the current acceleration strategies remains to be evaluated. Some clinical applications may be more amenable to such workflow than others. Moreover, the effects of different acceleration methods<sup>168,191,261–264</sup> on MBIR image quality invite more comprehensive investigation.

Another opportunity is the translation of MBIR methods leveraging patient-specific prior image information into sequential imaging studies. Among possible applications is the use of a planning CT image in CBCT scans in radiation therapy to improve the workflow in patient setup and enable online treatment planning.<sup>33,77</sup> Another potential application involves imaging scenarios in which angular undersampling is difficult to avoid, as in perfusion imaging.<sup>265,266</sup> The performance of the methods in Chapter 5-6 in these clinical applications remains to be fully investigated – for example, the robustness of the dPIRPLE algorithm against different deformation patterns. Other practical challenges include intensity inconsistencies between the planning CT and the subsequent CBCT and also invite future investigation. Despite such remaining challenges, MBIR methods such as those developed in this dissertation present an important element to the next generation of CBCT systems.

## Bibliography

1. Meggitt G. *Taming the Rays: A History of Radiation and Protection*. lulu.com; 2010.
2. National Council on Radiation Protection and Measurements. *NCRP Report No. 93, Ionizing Radiation Exposure of the Population of the United States.*; 1987.
3. Brenner DJ, Hall EJ. Computed tomography--an increasing source of radiation exposure. *N Engl J Med*. 2007;357(22):2277-2284. doi:10.1056/NEJMra072149.
4. Bushberg JT, Seibert JA, Leidholdt Jr. EM, Boone JM. *The Essential Physics of Medical Imaging*. 3rd ed. Lippincott Williams & Wilkins; 2001.
5. ICRP. Radiological Protection and Safety in Medicine. ICRP Publication 73. *Annu ICRP*. 1996;26(2).
6. ICRP. Managing Patient Dose in Computed Tomography. ICRP Publication 87. *Annu ICRP*. 2000;30(4).
7. Kalra MK, Maher MM, Toth TL, et al. Strategies for CT Radiation Dose Optimization. *Radiology*. 2004;230(3):619-628. doi:10.1148/radiol.2303021726.
8. McCollough CH, Primak AN, Braun N, Kofler J, Yu L, Christner J. Strategies for Reducing Radiation Dose in CT. *Radiol Clin North Am*. 2009;47(1):27-40. doi:10.1016/j.rcl.2008.10.006.
9. Kalra MK, Maher MM, Toth TL, et al. Techniques and Applications of Automatic Tube Current Modulation for CT. *Radiology*. 2004;233(3):649-657. doi:10.1148/radiol.2333031150.
10. Singh S, Kalra MK, Thrall JH, Mahesh M. Automatic exposure control in CT: Applications and limitations. *J Am Coll Radiol*. 2011;8(6):446-449.

doi:10.1016/j.jacr.2011.03.001.

11. Elbakri I a., Fessler J a. Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Trans Med Imaging*. 2002;21(2):89-99. doi:10.1109/42.993128.
12. Lasio GM, Whiting BR, Williamson JF. Statistical reconstruction for x-ray computed tomography using energy-integrating detectors. *Phys Med Biol*. 2007;52(8):2247-2266. doi:10.1088/0031-9155/52/8/014.
13. Thibault J-B, Sauer KD, Bouman CA, Hsieh J. A three-dimensional statistical approach to improved image quality for multislice helical CT. *Med Phys*. 2007;34(11):4526-4544. doi:10.1118/1.2789499.
14. Rafferty EA, Park JM, Philpotts LE, Poplack SP, Sumkin JH. Digital Mammography and Breast Tomosynthesis Compared with Digital Mammography Alone : Results of a Multicenter, multireader trial. *Radiology*. 2013;266(1):104-113. doi:10.1148/radiol.12120674/-/DC1.
15. Mozzo P, Procacci C, Tacconi A, Martini PT, Andreis IA. A new volumetric CT machine for dental imaging based on the cone-beam technique: preliminary results. *Eur Radiol*. 1998;8(9):1558-1564. doi:10.1007/s003300050586.
16. Pinsky HM, Dyda S, Pinsky RW, Misch KA, Sarment D. Accuracy of three-dimensional measurements using cone-beam CT. *Dentomaxillofacial Radiol*. 2006;35(6):410-416. doi:10.1259/dmfr/20987648.
17. Penninger RT, Tavassolie TS, Carey JP. Cone-beam volumetric tomography for applications in the temporal bone. *Otol Neurotol*. 2011;32(3):453-460. doi:10.1097/MAO.0b013e31820d962c.
18. Xu J, Reh DD, Carey JP, Mahesh M, Siewerdsen JH. Technical assessment of a cone-beam CT scanner for otolaryngology imaging: image quality, dose, and technique

- protocols. *Med Phys*. 2012;39(8):4932-4942. doi:10.1118/1.4736805.
19. Boone JM, Nelson TR, Lindfors KK, Seibert JA. Dedicated Breast CT: Radiation Dose and Image Quality Evaluation. *Radiology*. 2001;221(3):657-667.
  20. Chen B, Ning R. Cone-beam volume CT breast imaging: feasibility study. *Med Phys*. 2002;29(5):755-770. doi:10.1118/1.1461843.
  21. Glick SJ. Breast CT. *Annu Rev Biomed Eng*. 2007;9:501-526. doi:10.1146/annurev.bioeng.9.060906.151924.
  22. O'Connell A, Conover DL, Zhang Y, et al. Cone-beam CT for breast imaging: Radiation dose, breast coverage, and image quality. *Am J Roentgenol*. 2010;195(2):496-509. doi:10.2214/AJR.08.1017.
  23. Wienbeck S, Lotz J, Fischer U. Review of clinical studies and first clinical experiences with a commercially available cone-beam breast CT in Europe. *Clin Imaging*. 2017;42:50-59. doi:10.1016/j.clinimag.2016.11.011.
  24. Kalender WA, Kolditz D, Steiding C, et al. Technical feasibility proof for high-resolution low-dose photon-counting CT of the breast. *Eur Radiol*. 2017;27(3):1081-1086. doi:10.1007/s00330-016-4459-3.
  25. Zbijewski W, De Jean P, Prakash P, et al. A dedicated cone-beam CT system for musculoskeletal extremities imaging: Design, optimization, and initial performance characterization. *Med Phys*. 2011;38(8):4700-4713. doi:10.1118/1.3611039.
  26. Carrino JA, Al Muhit A, Zbijewski W, et al. Dedicated Cone-Beam CT System for Extremity Imaging. *Radiology*. 2014;270(3):816-824. doi:10.1148/radiol.13130225.
  27. Demehri S, Muhit A, Zbijewski W, et al. Assessment of image quality in soft tissue and bone visualization tasks for a dedicated extremity cone-beam CT system. *Eur Radiol*.

- 2015;25(6):1742-1751. doi:10.1007/s00330-014-3546-6.
28. Faccioli N, Foti G, Barillari M, Atzei A, Mucelli RP. Finger fractures imaging: Accuracy of cone-beam computed tomography and multislice computed tomography. *Skeletal Radiol.* 2010;39(11):1087-1095. doi:10.1007/s00256-010-0911-7.
  29. Koskinen SK, Haapamäki V V., Salo J, et al. CT arthrography of the wrist using a novel, mobile, dedicated extremity cone-beam CT (CBCT). *Skeletal Radiol.* 2013;42(5):649-657. doi:10.1007/s00256-012-1516-0.
  30. Masciocchi C, Barile A, Satragno L. Musculoskeletal MRI: dedicated systems. *Eur Radiol.* 2000;10(2):250-255. doi:10.1007/s003300050041.
  31. Naraghi AM, White LM, Patel C, Tomlinson G, Keystone EC. Comparison of 1.0-T Extremity MR and 1.5-T Conventional High-Field-Strength MR in Patients with Rheumatoid Arthritis. *Radiology.* 2009;251(3):829-837. doi:10.1148/radiol.2521081507.
  32. Roemer FW, Lynch JA, Niu J, et al. A comparison of dedicated 1.0 T extremity MRI vs large-bore 1.5 T MRI for semiquantitative whole organ assessment of osteoarthritis: the MOST study. *Osteoarthr Cartil.* 2010;18(2):168-174. doi:10.1016/j.joca.2009.08.017.
  33. Jaffray DA. Image-guided radiotherapy: from current concept to future perspectives. *Nat Rev Clin Oncol.* 2012;9(12):688-699. doi:10.1038/nrclinonc.2012.194.
  34. Herman MG, Balter JM, Jaffray DA, et al. Clinical Use of Electronic Portal Imaging: Report of AAPM Radiation Therapy Committee Task Group 58. *Med Phys.* 2001;28(5):712-737. doi:10.1118/1.1368128.
  35. Siewerdsen JH, Jaffray DA, Edmundson GK, Sanders WP, Wong JW, Martinez AA. Flat-panel cone-beam CT: a novel imaging technology for image-guided procedures. In: *SPIE Medical Imaging.* Vol 4319. ; 2001:435-444. doi:10.1117/12.428085.

36. Jaffray DA, Siewerdsen JH, Wong JW, Martinez AA. Flat-panel cone-beam computed tomography for image-guided radiation therapy. *Int J Radiat Oncol Biol Phys*. 2002;53(5):1337-1349.
37. Oldham M, Létourneau D, Watt L, et al. Cone-beam-CT guided radiation therapy: A model for on-line application. *Radiother Oncol*. 2005;75(3):1-8.  
doi:10.1016/j.radonc.2005.03.026.
38. Mutic S, Dempsey JF. The ViewRay System: Magnetic Resonance-Guided and Controlled Radiotherapy. *Semin Radiat Oncol*. 2014;24(3):196-199.  
doi:10.1016/j.semradonc.2014.02.008.
39. Wooten HO, Rodriguez V, Green O, et al. Benchmark IMRT evaluation of a Co-60 MRI-guided radiation therapy system. *Radiother Oncol*. 2015;114(3):402-405.  
doi:10.1016/j.radonc.2015.01.015.
40. Orecchia R, Surgo A, Muto M, et al. VERO radiotherapy for low burden cancer: 789 patients with 957 lesions. *Ecancermedicalscience*. 2016;10(677):1-11.  
doi:10.3332/ecancer.2016.677.
41. Korreman S, Persson G, Nygaard D, Brink C, Juhler-Nøttrup T. Respiration-correlated image guidance is the most important radiotherapy motion management strategy for most lung cancer patients. *Int J Radiat Oncol Biol Phys*. 2012;83(4):1338-1343.  
doi:10.1016/j.ijrobp.2011.09.010.
42. Cleary K, Peters TM. Image-Guided Interventions: Technology Review and Clinical Applications. *Annu Rev Biomed Eng*. 2010;12(1):119-142. doi:10.1146/annurev-bioeng-070909-105249.
43. Maciunas RJ. Computer-assisted Neurosurgery. *Clin Neurosurg*. 2006;53:267-271.
44. Comeau RM, Sadikot AF, Fenster A, Peters TM. Intraoperative ultrasound for guidance

- and tissue shift correction in image-guided neurosurgery. *Med Phys*. 2000;27(4):787-800.
45. Siewerdsen JH, Daly MJ, Chan H, et al. High-performance intraoperative cone-beam CT on a mobile C-arm: an integrated system for guidance of head and neck surgery. In: *SPIE Medical Imaging*. Vol 7261. ; 2009:72610J. doi:10.1117/12.813747.
  46. Merloz P, Tonetti J, Pittet L, Coulomb M, Lavalée S, Sautot P. Pedicle screw placement using image guided techniques. *Clin Orthop Relat Res*. 1998;354:39-48.
  47. Reininga IHF, Wagenmakers R, van den Akker-Scheek I, et al. Effectiveness of computer-navigated minimally invasive total hip surgery compared to conventional total hip arthroplasty: design of a randomized controlled trial. *BMC Musculoskelet Disord*. 2007;8(4):1-6. doi:10.1186/1471-2474-8-4.
  48. De Silva T, Lo S-FL, Aygun N, et al. Utility of the LevelCheck Algorithm for Decision Support in Vertebral Localization. *Spine (Phila Pa 1976)*. 2016;41(20):E1249-E1256. doi:10.1097/BRS.0000000000001589.
  49. Guttman MA, Ozturk C, Raval AN, et al. Interventional cardiovascular procedures guided by real-time MR imaging: An interactive interface using multiple slices, adaptive projection modes and live 3D renderings. *J Magn Reson Imaging*. 2007;26(6):1429-1435. doi:10.1002/jmri.21199.
  50. Rhode KS, Sermesant M, Brogan D, et al. A system for real-time XMR guided cardiovascular intervention. *IEEE Trans Med Imaging*. 2005;24(11):1428-1440. doi:10.1109/TMI.2005.856731.
  51. Linte CA, Moore J, Wiles AD, Wedlake C, Peters TM. Virtual reality-enhanced ultrasound guidance: a novel technique for intracardiac interventions. *Comput Aided Surg*. 2008;13(2):82-94. doi:10.3109/10929080801951160.
  52. Hasegawa M, Sone S, Takashima S, et al. Growth rate of small lung cancers detected on

- mass CT screening. *Br J Radiol*. 2000;73(876):1252-1259.  
doi:10.1259/bjr.73.876.11205667.
53. Werner-Wasik M, Xiao Y, Pequignot E, Curran WJ, Hauck W. Assessment of lung cancer response after nonoperative therapy: Tumor diameter, bidimensional product, and volume. A serial CT scan-based study. *Int J Radiat Oncol Biol Phys*. 2001;51(1):56-61.  
doi:10.1016/S0360-3016(01)01615-7.
  54. Siewerdsen JH, Moseley DJ, Burch S, et al. Volume CT with a flat-panel detector on a mobile, isocentric C-arm: pre-clinical investigation in guidance of minimally invasive surgery. *Med Phys*. 2005;32(1):241-254. doi:10.1118/1.1836331.
  55. Fahrig R, Dixon R, Payne T, Morin RL, Ganguly A, Strobel N. Dose and image quality for a cone-beam C-arm CT system. *Med Phys*. 2006;33(12):4541-4550.  
doi:10.1118/1.2370508.
  56. Navab N, Heining S-M, Traub J. Camera augmented mobile C-arm (CAMC): Calibration, accuracy study, and clinical applications. *IEEE Trans Med Imaging*. 2010;29(7):1412-1423. doi:10.1109/TMI.2009.2021947.
  57. Dang H, Otake Y, Schafer S, Stayman JW, Kleinszig G, Siewerdsen JH. Robust methods for automatic image-to-world registration in cone-beam CT interventional guidance. *Med Phys*. 2012;39(10):6484-6498. doi:10.1118/1.4754589.
  58. Antonuk LE, Boudry J, Huang W, et al. Demonstration of megavoltage and diagnostic x-ray imaging with hydrogenated amorphous silicon arrays. *Med Phys*. 1992;19(6):1455-1466.
  59. Rowlands JA, Zhao W, Blevins IM, Waechter DF, Huang Z. Flat-panel digital radiology with amorphous selenium and active-matrix readout. *RadioGraphics*. 1997;17(3):753-760.  
doi:10.1148/radiographics.17.3.9153709.

60. Granfors PR, Aufrichtig R. Performance of a 41×41-cm<sup>2</sup> amorphous silicon flat panel x-ray detector for radiographic imaging applications. *Med Phys*. 2000;27(6):1324-1331. doi:10.1118/1.599010.
61. Colbeth RE, Boyce S, Fong R, et al. 40 x 30 cm Flat Panel Imager for Angiography, R&F, and Cone-Beam CT Applications. In: *SPIE Medical Imaging*. Vol 4320. ; 2001:94-102.
62. Tward DJ, Siewerdsen JH. Cascaded systems analysis of the 3D noise transfer characteristics of flat-panel cone-beam CT. *Med Phys*. 2008;35(12):5510-5529. doi:10.1118/1.3002414.
63. Gang G. J, Tward DJ, Lee J, Siewerdsen JH. Anatomical background and generalized detectability in tomosynthesis and cone-beam CT. *Med Phys*. 2010;37(5):1948-1965. doi:10.1118/1.3352586.
64. Gang GJ, Lee J, Stayman JW, et al. Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance. *Med Phys*. 2011;38(4):1754-1768. doi:10.1118/1.3560428.
65. Gang GJ, Zbijewski W, Webster Stayman J, Siewerdsen JH. Cascaded systems analysis of noise and detectability in dual-energy cone-beam CT. *Med Phys*. 2012;39(8):5145-5156. doi:10.1118/1.4736420.
66. Rabbani M, Shaw R, Van Metter R. Detective quantum efficiency of imaging systems with amplifying and scattering mechanisms. *J Opt Soc Am A*. 1987;4(5):895-901. doi:10.1364/JOSAA.4.000895.
67. Cunningham IA, Westmore MS, Fenster A. A spatial-frequency dependent quantum accounting diagram and detective quantum efficiency model of signal and noise propagation in cascaded imaging system. *Med Phys*. 1994;21(3):417-427.
68. Siewerdsen JH, Antonuk LE, El-Mohri Y, et al. Empirical and theoretical investigation of

- the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology. *Med Phys*. 1997;24(1):71-89. doi:10.1118/1.597919.
69. Siewerdsen JH, Antonuk LE, El-Mohri Y, Yorkston J, Huang W, Cunningham IA. Signal, noise power spectrum, and detective quantum efficiency of indirect-detection flat-panel imagers for diagnostic radiology. *Med Phys*. 1998;25(5):614-628. doi:10.1118/1.598243.
  70. Yao J, Cunningham IA. Parallel cascades: new ways to describe noise transfer in medical imaging systems. *Med Phys*. 2001;28(10):2020-2038. doi:10.1118/1.1405842.
  71. Zhao W, Rowlands JA. Digital radiology using active matrix readout of amorphous selenium: theoretical analysis of detective quantum efficiency. *Med Phys*. 1997;24(12):1819-1833. doi:10.1118/1.598097.
  72. Zhao W, Ji WG, Rowlands JA. Effects of characteristic x rays on the noise power spectra and detective quantum efficiency of photoconductive x-ray detectors. *Med Phys*. 2001;28(10):2039-2049. doi:10.1118/1.1405845.
  73. Prakash P, Zbijewski W, Gang GJ, et al. Task-based modeling and optimization of a cone-beam CT scanner for musculoskeletal imaging. *Med Phys*. 2011;38(10):5612-5629. doi:10.1118/1.3633937.
  74. Glick SJ, Vedantham S, Karellas A. Investigation of optimal kVp settings for CT mammography using a flat-panel imager. In: *SPIE Medical Imaging*. Vol 4682. ; 2002:392-402. doi:10.1117/12.465581.
  75. Shen Y, Zhong Y, Lai C-J, Wang T, Shaw CC. Cone beam breast CT with a high pitch (75  $\mu\text{m}$ ), thick (500  $\mu\text{m}$ ) scintillator CMOS flat panel detector: Visibility of simulated microcalcifications. *Med Phys*. 2013;40(10):101915. doi:10.1118/1.4820440.
  76. Moore CJ, Amer A, Marchant T, et al. Developments in and experience of kilovoltage X-ray cone beam image-guided radiotherapy. *Br J Radiol*. 2006;79(Special Issue 1):S66-

- S78. doi:10.1259/bjr/68255935.
77. Boda-Heggemann J, Lohr F, Wenz F, Flentje M, Guckenberger M. kV cone-beam CT-based IGRT. *Strahlentherapie und Onkol.* 2011;187(5):284-291. doi:10.1007/s00066-011-2236-4.
  78. Xing L, Thorndyke B, Schreiber E, et al. Overview of image-guided radiation therapy. *Med Dosim.* 2006;31(2):91-112. doi:10.1016/j.meddos.2005.12.004.
  79. Siewerdsen JH. Cone-beam CT with a flat-panel detector: From image science to image-guided surgery. *Nucl Instruments Methods Phys Res Sect A Accel Spectrometers, Detect Assoc Equip.* 2011;648(SUPPL. 1):S241-S250. doi:10.1016/j.nima.2010.11.088.
  80. Zentai G. Comparison of CMOS and a-Si flat panel imagers for X-ray imaging. In: *IEEE International Conference on Imaging Systems and Techniques.* ; 2011:194-200. doi:10.1109/IST.2011.5962217.
  81. Badal A, Badano A. Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit. *Med Phys.* 2009;36(11):4878. doi:10.1118/1.3231824.
  82. Sisniega A, Zbijewski W, Badal A, et al. Monte Carlo study of the effects of system geometry and antiscatter grids on cone-beam CT scatter distributions. *Med Phys.* 2013;40(5):51915. doi:10.1118/1.4801895.
  83. Kak AC, Slaney M. *Principles of Computerized Tomographic Imaging.*; 2001.
  84. Hsieh J. *Computed Tomography: Principles, Design, Artifacts, and Recent Advances.* Wiley; 2009.
  85. Feldkamp LA, Davis LC, Kress JW. Practical cone-beam algorithm. *Opt Soc Am.* 1984;1(6):612-619.

86. Grass M, Kohler T, Proksa R. 3D cone-beam CT reconstructions for circular trajectories. *Phys Med Biol.* 2000;45(2):329-347.
87. Hsieh J, Tang X. Tilted cone-beam reconstruction with row-wise fan-to-parallel rebinning. *Phys Med Biol.* 2006;51(20):5259-5276. doi:10.1088/0031-9155/51/20/012.
88. Katsevich A. Theoretically exact filtered backprojection-type inverse algorithm for spiral CT. *SIAM J Appl Math.* 2002;62(6):2012-2026.
89. Pack JD, Noo F. Cone-beam reconstruction using 1D filtering along the projection of M - lines. *Inverse Probl.* 2005;21(3):1105-1120. doi:10.1088/0266-5611/21/3/019.
90. Zamyatin AA, Katsevich A, Chiang BS. Exact image reconstruction for a circle and line trajectory with a gantry tilt. *Phys Med Biol.* 2008;53(23):N423-N435. doi:10.1088/0031-9155/53/23/N02.
91. Rockmore AJ, Macovski A. A maximum likelihood approach to transmission image reconstruction from projections. *IEEE Trans Nucl Sci.* 1977;24(3):1929-1935. doi:10.1109/TNS.1977.4329128.
92. Leahy RM, Qi J. Statistical approaches in quantitative positron emission tomography. *Stat Comput.* 2000;10(2):147-165. doi:10.1023/A:1008946426658.
93. Wang AS, Stayman JW, Otake Y, et al. Soft-tissue imaging with C-arm cone-beam CT using statistical reconstruction. *Phys Med Biol.* 2014;59(4):1005-1026. doi:10.1088/0031-9155/59/4/1005.
94. Whiting BR, Massoumzadeh P, Earl OA, O'Sullivan JA, Snyder DL, Williamson JF. Properties of preprocessed sinogram data in x-ray computed tomography. *Med Phys.* 2006;33(9):3290-3303. doi:10.1118/1.2230762.
95. Xu J, Tsui BMW. Electronic noise modeling in statistical iterative reconstruction. *IEEE*

- Trans Image Process.* 2009;18(6):1228-1238. doi:10.1109/TIP.2009.2017139.
96. Yu DF, Fessler JA, Ficarò EP, Yu DF, Ficarò EP. Maximum-likelihood transmission image reconstruction for overlapping transmission beams. *IEEE Trans Med Imaging.* 2000;19(11):1094-1105. doi:10.1109/42.896785.
97. De Man B, Basu S. Distance-driven projection and backprojection in three dimensions. *Phys Med Biol.* 2004;49(11):2463-2475. doi:10.1109/NSSMIC.2002.1239600.
98. Long Y, Fessler JA, Balter JM. 3D forward and back-projection for X-ray CT using separable footprints. *IEEE Trans Image Process.* 2010;29(11):1839-1850. doi:10.1109/TMI.2010.2050898.
99. Tilley II S, Siewerdsen JH, Stayman JW. Model-based iterative reconstruction for flat-panel cone-beam CT with focal spot blur, detector blur, and correlated noise. *Phys Med Biol.* 2015;61(1):296-319. doi:10.1088/0031-9155/61/1/296.
100. De Man B, Nuyts J, Dupont P, Marchal G, Suetens P. An iterative maximum-likelihood polychromatic algorithm for CT. *IEEE Trans Med Imaging.* 2001;20(10):999-1008. doi:10.1109/42.959297.
101. Elbakri IA, Fessler JA. Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation. *Phys Med Biol.* 2003;48(15):2453-2477.
102. Xu S, Uneri A, Khanna AJ, Siewerdsen JH, Stayman JW. Polyenergetic known component reconstruction (KCR) for flat-panel CBCT with unknown material compositions and unknown x-ray spectra. *Phys Med Biol.* 2017.
103. Evans JD, Whiting BR, Polite DG, O'Sullivan JA, Klahr PF, Williamson JF. Experimental implementation of a polyenergetic statistical reconstruction algorithm for a commercial fan-beam CT scanner. *Phys Medica.* 2013;29(5):500-512.

doi:10.1016/j.ejmp.2012.12.005.

104. Snyder DL, O'Sullivan JA, Murphy RJ, Politte DG, Whiting BR, Williamson JF. Image reconstruction for transmission tomography when projection data are incomplete. *Phys Med Biol*. 2006;51(21):5603-5619. doi:10.1088/0031-9155/51/21/015.
105. Stayman JW, Siewerdsen JH. Task-based trajectories in iterative reconstructed interventional cone-beam CT. In: *International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. ; 2013:257-260.
106. Lange K. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Trans Med Imaging*. 1990;9(4):439-446. doi:10.1109/42.61759.
107. Nikolova M, Idier J, Mohammad-Djafari A. Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF. *IEEE Trans Image Process*. 1994;7(4):571-585.
108. Vogel CR, Oman ME. Iterative Methods for total variation denoising. *Sci Comput*. 1996;17(1):227-238.
109. Sidky EY, Pan X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys Med Biol*. 2008;53(17):4777-4807. doi:10.1088/0031-9155/53/17/021.
110. Chen G-H, Tang J, Leng S. Prior Image Constrained Compressed Sensing (PICCS). In: *SPIE Medical Imaging*. Vol 6856. ; 2008:685618. doi:10.1117/12.770532.
111. Stayman JW, Dang H, Ding Y, Siewerdsen JH. PIRPLE: a penalized-likelihood framework for incorporation of prior images in CT reconstruction. *Phys Med Biol*. 2013;58(21):7563-7582. doi:10.1088/0031-9155/58/21/7563.
112. Box GE, Hunter JS, Hunter WG. *Statistics for Experimenters: Design, Innovation, and*

*Discovery*. New York: Wiley; 2005.

113. Baker JR, Budinger TF, Huesman RH. Generalized approach to inverse problems in tomography: image reconstruction for spatially variant systems using natural pixels. *Crit Rev Biomed Eng*. 1992;20(1-2):47-71.
114. Matej S, Lewitt RM. Practical Considerations for 3-D construction Using Spherical Symmetric Volume Elements. *IEEE Trans Med Imaging*. 1996;15(1):68-78.  
doi:10.1109/42.481442.
115. Peters TM. Algorithms for Fast Back- and Re-Projection in Computed Tomography. *IEEE Trans Nucl Sci*. 1981;28(4):3641-3647. doi:10.1109/TNS.1981.4331812.
116. Zhuang W, Gopal SS, Hebert TJ. Numerical evaluation of methods for computing tomographic projections. *IEEE Trans Nucl Sci*. 1994;41(4):1660-1665.  
doi:10.1109/23.322963.
117. Siddon RL. Fast calculation of the exact radiological path for a three-dimensional CT array. *Med Phys*. 1985;12(2):252-255.
118. De Man B, Basu S. Distance-driven projection and backprojection. In: *IEEE Nuclear Science Symposium Conference Record*. ; 2002:1477-1480.  
doi:10.1109/NSSMIC.2002.1239600.
119. Foi A, Trimeche M, Katkovnik V, Egiazarian K. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Trans Image Process*. 2008;17(10):1737-1754. doi:10.1109/TIP.2008.2001399.
120. Gordon R, Bender R, Herman GT. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. *Theor Biol*. 1970;29(3):471-481.
121. Herman GT, Lent A, Rowland SW. ART: Mathematics and applications. A report on the

- mathematical foundations and on the applicability to real data of the algebraic reconstruction techniques. *J Theor Biol.* 1973;42(1):1-32. doi:10.1016/0022-5193(73)90145-8.
122. Andersen AH, Kak AC. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrason Imaging.* 1984;6(1):81-94.
  123. Sonka M, Fitzpatrick JM. *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis.*; 2000. doi:10.1117/3.831079.
  124. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Phys D Nonlinear Phenom.* 1992;60(1-4):259-268. doi:10.1016/0167-2789(92)90242-F.
  125. Panin VY, Zeng GL, Gullberg GT. Total variation regulated EM algorithm. *IEEE Trans Nucl Sci.* 1999;46(6):2202-2210. doi:10.1109/23.819305.
  126. Persson M, Bone D, Elmqvist H. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. *Phys Med Biol.* 2001;46(3):853-866. doi:10.1088/0031-9155/46/3/318.
  127. Song J, Liu QH, Johnson GA, Badea CT. Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT. *Med Phys.* 2007;34(11):4476-4483. doi:10.1118/1.2795830.
  128. Tang J, Nett BE, Chen G-H. Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms. *Phys Med Biol.* 2009;54(19):5781-5804. doi:10.1088/0031-9155/54/19/008.
  129. Hebert TJ. Statistical Stopping Criteria for Iterative Maximum-Likelihood Reconstruction of Emission Images. *Phys Med Biol.* 1990;35(9):1221-1232.
  130. Snyder DL, Miller MI, Thomas LJ, Politte DG. Noise and edge artifacts in maximum-

- likelihood reconstructions for emission tomography. *IEEE Trans Med Imaging*. 1987;6(3):228-238. doi:10.1109/TMI.1987.4307831.
131. Huber PJ. *Robust Statistics*. New York: Wiley; 1981.
  132. Fessler JA, Rogers WL. Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs. *IEEE Trans Image Process*. 1996;5(9):1346-1358. doi:10.1109/83.535846.
  133. Qi J, Leahy RM. Resolution and noise properties of MAP reconstruction for fully 3-D PET. *IEEE Trans Med Imaging*. 2000;19(5):493-506. doi:10.1109/42.870259.
  134. Stayman JW, Fessler JA. Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction. *IEEE Trans Med Imaging*. 2000;19(6):601-615. doi:10.1109/42.870666.
  135. Yang L, Zhou J, Ferrero A, Badawi RD, Qi J. Regularization design in penalized maximum-likelihood image reconstruction for lesion detection in 3D PET. *Phys Med Biol*. 2014;59(2):403-419. doi:10.1088/0031-9155/59/2/403.
  136. Gang GJ, Stayman JW, Zbijewski W, Siewerdsen JH. Task-based detectability in CT image reconstruction by filtered backprojection and penalized likelihood estimation. *Med Phys*. 2014;41(8):81902. doi:10.1118/1.4883816.
  137. Cho JH, Fessler JA. *Regularization Designs for Uniform Spatial Resolution and Noise Properties in Statistical Image Reconstruction for 3D X-Ray CT: Supplementary Material*.; 2015. doi:10.1109/TMI.2014.2365179.
  138. Gang GJ, Siewerdsen JH, Stayman JW. Task-driven tube current modulation and regularization design in computed tomography with penalized-likelihood reconstruction. In: *SPIE Medical Imaging*. Vol 9783. ; 2016:978324. doi:10.1117/12.2216387.

139. Sauer K, Bouman C. A Local Update Strategy for Iterative Reconstruction from Projections. *IEEE Trans Signal Process.* 1993;41(2):534-548. doi:10.1109/78.193196.
140. Hill MD, Silver FL, Austin PC, Tu J V. Rate of stroke recurrence in patients with primary intracerebral hemorrhage. *Stroke.* 2000;31(1):123-127. doi:10.1161/01.STR.31.1.123.
141. Dublin AB, French BN, Rennick JM. Computed tomography in head trauma. *Radiology.* 1977;122(2):365-369.
142. Parizel PM, Makkat S, Van Miert E, Van Goethem JW, Van den Hauwe L, De Schepper AM. Intracranial hemorrhage: Principles of CT and MRI interpretation. *Eur Radiol.* 2001;11(9):1770-1783. doi:10.1007/s0033000000800.
143. Kidwell CS, Wintermark M. Imaging of intracranial haemorrhage. *Lancet Neurol.* 2008;7(3):256-267. doi:10.1212/01.CON.0000333199.98344.da.
144. Kidwell CS, Chalela JA, Saver JL, et al. Comparison of MRI and CT for Detection of Acute Intracerebral Hemorrhage. *JAMA.* 2014;292(15):1823-1830.
145. Masaryk TJ, Kolonick R, Painter T, Weinreb DB. Benefits of Portable Head / Neck CT Imaging in the Intensive Care Unit. *Radiol Manage.* 2008;30(2):50-54.
146. Smith I, Fleming S, Cernaianu A. Mishaps during transport from the intensive care unit. *Crit Care Med.* 1990;18(3):278-281.
147. Indeck M, Peterson S, Smith J, Brotman S. Risk, cost, and benefit of transporting ICU patients for special studies. *J Trauma.* 1988;28(7):1020-1025.
148. Boone JM, Kwan ALC, Yang K, Burkett GW, Lindfors KK, Nelson TR. Computed tomography for imaging the breast. *J Mammary Gland Biol Neoplasia.* 2006;11(2):103-111. doi:10.1007/s10911-006-9017-1.
149. Sisniega A, Zbijewski W, Xu J, et al. High-fidelity artifact correction for cone-beam CT

- imaging of the brain. *Phys Med Biol*. 2015;60(4):1415-1439. doi:10.1088/0031-9155/60/4/1415.
150. Kachelrieß M, Watzke O, Kalender WA. Generalized multi-dimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT. *Med Phys*. 2001;28(4):475-490. doi:10.1118/1.1358303.
  151. La Rivière PJ, Bian J, Vargas PA. Penalized-likelihood sinogram restoration for computed tomography. *IEEE Trans Med Imaging*. 2006;25(8):1022-1036.
  152. Zhu L, Wang J, Xing L. Noise suppression in scatter correction for cone-beam CT. *Med Phys*. 2009;36(3):741-752. doi:10.1118/1.3063001.
  153. Boone JM, Seibert JA. An analytical model of the scattered radiation distribution in diagnostic radiology. *Med Phys*. 1988;15(5):721-725. doi:10.1118/1.596186.
  154. Ning R, Tang X, Conover D. X-ray scatter correction algorithm for cone beam CT imaging. *Med Phys*. 2004;31(5):1195-1202. doi:10.1118/1.1711475.
  155. Siewerdsen JH, Daly MJ, Bakhtiar B, et al. A simple, direct method for x-ray scatter estimation and correction in digital radiography and cone-beam CT. *Med Phys*. 2006;33(1):187-197. doi:10.1118/1.2148916.
  156. Zhu L, Bennett NR, Fahrig R. Scatter Correction Method for X-Ray CT Using Primary Modulation: Theory and Preliminary Results. *IEEE Trans Med Imaging*. 2006;25(12):1573-1587. doi:10.1109/TMI.2006.884636.
  157. Joseph PM, Spital RD. A method for correcting bone induced artifacts in computed tomography scanners. *J Comput Assist Tomogr*. 1978;2(1):100-108.
  158. Hsieh J, Molthen RC, Dawson CA, Johnson RH. An iterative approach to the beam hardening correction in cone beam CT. *Med Phys*. 2000;27(1):23-29.

doi:10.1118/1.598853.

159. Li T, Li X, Jing W, et al. Nonlinear sinogram smoothing for low-dose X-ray CT. *IEEE Trans Nucl Sci.* 2004;51(5):2505-2513. doi:10.1109/TNS.2004.834824.
160. Benaroya H, Han SM, Nagurka M. *Probability Models in Engineering and Science*. CRC Press; 2005.
161. Kwan ALC, Seibert JA, Boone JM. An improved method for flat-field correction of flat panel x-ray detector. *Med Phys.* 2006;33(2):391-393. doi:10.1118/1.2163388.
162. Siewerdsen JH, Jaffray DA. Cone-beam computed tomography with a flat-panel imager: Magnitude and effects of x-ray scatter. *Med Phys.* 2001;28(2):220-231. doi:10.1118/1.1339879.
163. Erdoğan H, Fessler JA. Ordered subsets algorithms for transmission tomography. *Phys Med Biol.* 1999;44(11):2835-2851.
164. Zbijewski W, Stayman JW. Volumetric soft tissue brain imaging on xCAT, a mobile, flat-panel X-ray CT system. In: *SPIE Medical Imaging*. Vol 7258. ; 2009:72582K. doi:10.1117/12.811025.
165. Greenberg MS, Arredondo N. *Handbook of Neurosurgery*. New York: Greenberg Graphics; 2006.
166. Wang AS, Stayman JW, Otake Y, Kleinszig G, Vogt S, Siewerdsen JH. Nesterov's Method for Accelerated Penalized-Likelihood Statistical Reconstruction for C-arm Cone-Beam CT. In: *The Third International Conference on Image Formation in X-Ray Computed Tomography*. ; 2014:409-413.
167. Kim D, Pal D, Thibault JB, Fessler JA. Accelerating ordered subsets image reconstruction for X-ray CT using spatially nonuniform optimization transfer. *IEEE Trans Med Imaging*.

- 2013;32(11):1965-1978. doi:10.1109/TMI.2013.2266898.
168. Kim D, Ramani S, Fessler JA. Combining Ordered Subsets and Momentum for Accelerated X-ray CT Image Reconstruction. *IEEE Trans Med Imaging*. 2015;34(1):167-178.
  169. Siewerdsen JH, Cunningham IA, Jaffray DA. A framework for noise-power spectrum analysis of multidimensional images. *Med Phys*. 2002;29(11):2655-2671. doi:10.1118/1.1513158.
  170. Stayman JW, Zbijewski W, Tilley SI, Siewerdsen JH. Generalized least-squares CT reconstruction with detector blur and correlated noise models. In: *SPIE Medical Imaging*. Vol 9033. ; 2014:903335. doi:10.1117/12.2043067.
  171. Tilley SI, Siewerdsen JH, Stayman JW. Iterative CT Reconstruction using Models of Source and Detector Blur and Correlated Noise. In: *The Third International Conference on Image Formation in X-Ray Computed Tomography*. ; 2014:363-367.
  172. Siewerdsen JH, Antonuk LE. DQE and System Optimization for Indirect-Detection Flat-Panel Imagers in Diagnostic Radiology. In: *SPIE Medical Imaging*. Vol 3336. ; 1998:546-555.
  173. Tward DJ, Siewerdsen JH, Daly MJ, et al. Soft-tissue detectability in cone-beam CT: evaluation by 2AFC tests in relation to physical performance metrics. *Med Phys*. 2007;34(11):4459-4471. doi:10.1118/1.2790586.
  174. Wunderlich A, Noo F. Image covariance and lesion detectability in direct fan-beam x-ray computed tomography. *Phys Med Biol*. 2008;53(10):2471-2493. doi:10.1088/0031-9155/53/10/002.
  175. Barrett HH, Myers KJ. *Foundations of Image Science*. New York: Wiley; 2004.

176. Adams JH, Doyle D, Ford I, Gennarelli TA, Graham DI, McLellan DR. Diffuse axonal injury in head injury: definition, diagnosis and grading. *Histopathology*. 1989;15(1):49-59. doi:10.1111/j.1365-2559.1989.tb03040.x.
177. Dang H, Wang AS, Sussman MS, Siewerdsen JH, Stayman JW. dPIRPLE: a joint estimation framework for deformable registration and penalized-likelihood CT image reconstruction using prior images. *Phys Med Biol*. 2014;59(17):4799-4826. doi:10.1088/0031-9155/59/17/4799.
178. Barrett HH. Objective assessment of image quality: effects of quantum noise and object variability. *J Opt Soc Am A*. 1990;7(7):1266-1278. doi:10.1364/JOSAA.7.001266.
179. Reiser I, Nishikawa RM. Task-based assessment of breast tomosynthesis: effect of acquisition parameters and quantum noise. *Med Phys*. 2010;37(4):1591-1600. doi:10.1118/1.3357288.
180. 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. *Med Phys*. 2012;39(7):4115-4122. doi:10.1118/1.4725171.
181. Samei E, Richard S, Lurwitz L. Model-based CT performance assessment and optimization for iodinated and noniodinated imaging tasks as a function of kVp and body size. *Med Phys*. 2014;41(8):81910. doi:10.1118/1.4890082.
182. Sanchez AA, Sidky EY, Pan X. Task-based optimization of dedicated breast CT via Hotelling observer metrics. *Med Phys*. 2014;41(10):101917. doi:10.1118/1.4896099.
183. Xu J, Sisniega A, Zbijewski W, et al. Modeling and design of a cone-beam CT head scanner using task-based imaging performance optimization. *Phys Med Biol*. 2016;61(8):3180-3207. doi:10.1088/0031-9155/61/8/3180.
184. Xu J, Sisniega A, Zbijewski W, et al. Evaluation of detector readout gain mode and bowtie

- filters for cone-beam CT imaging of the head. *Phys Med Biol*. 2016;61(16):5973-5992. doi:10.1088/0031-9155/61/16/5973.
185. Gang GJ, Stayman JW, Ehtiati T, Siewerdsen JH. Task-driven image acquisition and reconstruction in cone-beam CT. *Phys Med Biol*. 2015;60(8):3129-3150. doi:10.1088/0031-9155/60/8/3129.
  186. International Commission on Radiation Units and Measurements. *ICRU Report 54: Medical Imaging - The Assessment of Image Quality.*; 1995.
  187. Pineda AR, Siewerdsen JH, Tward DJ. Analysis of image noise in 3D cone-beam CT: spatial and Fourier domain approaches under conditions of varying stationarity. In: *SPIE Medical Imaging*. Vol 6913. ; 2008:69131Q. doi:10.1117/12.772966.
  188. Brunner CC, Abboud SF, Hoeschen C, Kyprianou IS. Signal detection and location-dependent noise in cone-beam computed tomography using the spatial definition of the Hotelling SNR. *Med Phys*. 2012;39(6):3214-3228. doi:10.1118/1.4718572.
  189. Powell MJ. *Algorithms for Approximation*. Oxford: Clarendon Press; 1987.
  190. Clinthorne NH, Pan TS, Chiao PC, Rogers WL, Stamos JA. Preconditioning methods for improved convergence rates in iterative reconstructions. *IEEE Trans Med Imaging*. 1993;12(1):78-83. doi:10.1109/42.222670.
  191. Wang AS, Stayman JW, Otake Y, Vogt S, Kleinszig G, Siewerdsen JH. Accelerated statistical reconstruction for C-arm cone-beam CT using Nesterov's method. *Med Phys*. 2015;42(5):2699-2708. doi:10.1118/1.4914378.
  192. Huda W, Chamberlain CC, Rosenbaum AE, Garrisi W. Radiation doses to infants and adults undergoing head CT examinations. *Med Phys*. 2001;28(3):393-399. doi:10.1118/1.1350435.

193. Qi J. Analysis of lesion detectability in Bayesian emission reconstruction with nonstationary object variability. *IEEE Trans Med Imaging*. 2004;23(3):321-329. doi:10.1109/TMI.2004.824239.
194. Qi J, Huesman RH. Penalized maximum-likelihood image reconstruction for lesion detection. *Phys Med Biol*. 2006;51(16):4017-4029. doi:10.1088/0031-9155/51/16/009.
195. Yendiki A, Fessler JA. Analysis of observer performance in known-location tasks for tomographic image reconstruction. *IEEE Trans Med Imaging*. 2006;25(1):28-41. doi:10.1109/TMI.2005.859714.
196. Yendiki A, Fessler JA. Analysis of observer performance in unknown-location tasks for tomographic image reconstruction. *J Opt Soc Am A*. 2007;24(12):B99-B109. doi:10.1126/scisignal.2001449.Engineering.
197. Yao J, Barrett HH. Predicting human performance by a Channelized Hotelling Observer Model. In: *SPIE Medical Imaging*. Vol 1768. ; 1992:161-168.
198. Abbey CK, Barrett HH, Wilson DW. Observer signal-to-noise ratios for the ML-EM algorithm. In: *SPIE Medical Imaging*. Vol 2712. ; 1996:47-58. doi:10.1117/12.236860.Observer.
199. Stayman JW, Fessler JA. Efficient Calculation of Resolution and Covariance for Penalized-Likelihood Reconstruction in Fully 3-D SPECT. *IEEE Trans Med Imaging*. 2004;23(12):1543-1556.
200. Ohnesorge B, Flohr T, Schwarz K, Heiken JP, Bae KT. Efficient correction for CT image artifacts caused by objects extending outside the scan field of view. *Med Phys*. 2000;27(1):39-46. doi:10.1118/1.598855.
201. Hsieh J, Chao E, Thibault J, et al. A novel reconstruction algorithm to extend the CT scan field-of-view. *Med Phys*. 2004;31(9):2385-2391. doi:10.1118/1.1776673.

202. Kolditz D, Meyer M, Kyriakou Y, Kalender WA. Comparison of extended field-of-view reconstructions in C-arm flat-detector CT using patient size, shape or attenuation information. *Phys Med Biol.* 2010;56(1):39-56. doi:10.1088/0031-9155/56/1/003.
203. Xia Y, Hofmann H, Dennerlein F, et al. Towards clinical application of a laplace operator-based region of interest reconstruction algorithm in c-Arm Ct. *IEEE Trans Med Imaging.* 2014;33(3):593-606. doi:10.1109/TMI.2013.2291622.
204. Xu J, Sisniega A, Zbijewski W, et al. Technical assessment of a prototype cone-beam CT system for imaging of acute intracranial hemorrhage. *Med Phys.* 2016;43(10):5745-5757. doi:10.1118/1.4963220.
205. Hamelin B, Goussard Y, Dussault JP. Penalized-likelihood region-of-interest CT reconstruction by local object supersampling. In: *Proceedings of the 29th Annual International Conference of the IEEE EMBS.* ; 2007:739-742. doi:10.1109/IEMBS.2007.4352396.
206. Cao Q, Zbijewski W, Yorkston J, Siewerdsen JH, Stayman JW. Multi-Resolution Penalized Weighted Least-Squares Reconstruction for Quantitative Cone-Beam CT Imaging of Bone Morphology. In: *International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine.* ; 2015:452-455.
207. Delaney AH, Bresler Y. Multiresolution tomographic reconstruction using wavelets. *IEEE Trans Image Process.* 1995;4(6):799-813. doi:10.1109/83.388081.
208. Cao Q, Zbijweski W, Sisniega A, Yorkston J, Siewerdsen JH, Stayman JW. Multiresolution iterative reconstruction in high-resolution extremity cone-beam CT. *Phys Med Biol.* 2016;61(20):7263-7281. doi:10.1088/0031-9155/61/20/7263.
209. Noo F, Clackdoyle R, Pack JD. A two-step Hilbert transform method for 2D image reconstruction. *Phys Med Biol.* 2004;49(17):3903-3923. doi:10.1088/0031-

9155/49/17/006.

- 210. Defrise M, Noo F, Clackdoyle R, Kudo H. Truncated Hilbert transform and image reconstruction from limited tomographic data. *Inverse Probl.* 2006;22(3):1037-1053. doi:10.1088/0266-5611/22/3/019.
- 211. Kudo H, Courdurier M, Noo F, Defrise M. Tiny a priori knowledge solves the interior problem in computed tomography. *Phys Med Biol.* 2008;53(9):2207-2231. doi:10.1088/0031-9155/53/9/001.
- 212. Zou Y, Pan X. Exact image reconstruction on PI-lines from minimum data in helical cone-beam CT. *Phys Med Biol.* 2004;49(6):941-959. doi:10.1088/0031-9155/49/6/006.
- 213. Pan X, Zou Y, Xia D. Image reconstruction in peripheral and central regions-of-interest and data redundancy. *Med Phys.* 2005;32(3):673-684. doi:10.1118/1.1844171.
- 214. Yu H, Ye Y, Wang G. Practical cone-beam lambda tomography. *Med Phys.* 2006;33(10):3640-3646. doi:10.1118/1.2348767.
- 215. Yu H, Wang G. Compressed sensing based interior tomography. *Phys Med Biol.* 2009;54(9):2791-2805. doi:10.1088/0031-9155/54/9/014.
- 216. Stayman JW, Otake Y, Prince JL, Khanna AJ, Siewerdsen JH. Model-based tomographic reconstruction of objects containing known components. *IEEE Trans Med Imaging.* 2012;31(10):1837-1848. doi:10.1117/12.911202.
- 217. Candès EJ, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inf Theory.* 2006;52(2):489-509. doi:10.1109/TIT.2005.862083.
- 218. Lauzier PT, Chen G-H. Characterization of statistical prior image constrained compressed sensing (PICCS): II. Application to dose reduction. *Med Phys.* 2013;40(2):21902.

doi:10.1118/1.4773866.

219. Xu J, Tsui BMW. C-arm CT Image Reconstruction From Sparse Projections. In: *International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine.* ; 2013:34-37.
220. Zhang H, Han H, Wang J, et al. Deriving adaptive MRF coefficients from previous normal-dose CT scan for low-dose image reconstruction via penalized weighted least-squares minimization. *Med Phys.* 2013;41(4):41916. doi:10.1118/1.4869160.
221. Ding Y, Siewerdsen JH, Stayman JW. Incorporation of noise and prior images in penalized-likelihood reconstruction of sparse data. In: *SPIE Medical Imaging.* Vol 8313. ; 2012:831324. doi:10.1117/12.911667.
222. Nett B, Tang J, Aagaard-Kienitz B, Rowley H, Chen G-H. Low Radiation Dose C-arm Cone-beam CT based on Prior Image Constrained Compressed Sensing (PICCS): Including compensation for image volume mismatch between multiple acquisitions. In: *SPIE Medical Imaging.* Vol 7258. ; 2009:725803. doi:10.1117/12.813800.
223. Zeng R, Fessler JA, Balter JM. Estimating 3-D respiratory motion from orbiting views by tomographic image registration. *IEEE Trans Med Imaging.* 2007;26(2):153-163. doi:10.1109/TMI.2006.889719.
224. Ren L, Chetty IJ, Zhang J, et al. Development and clinical evaluation of a three-dimensional cone-beam computed tomography estimation method using a deformation field map. *Int J Radiat Oncol Biol Phys.* 2012;82(5):1584-1593. doi:10.1016/j.ijrobp.2011.02.002.
225. Wang J, Gu X. High-quality four-dimensional cone-beam CT by deforming prior images. *Phys Med Biol.* 2013;58(2):231-246. doi:10.1088/0031-9155/58/2/231.
226. Heußner T, Brehm M, Ritschl L, Sawall S, Kachelrieß M. Prior-based artifact correction

- (PBAC) in computed tomography. *Med Phys*. 2014;41(2):21906. doi:10.1118/1.4851536.
227. Gilland DR, Mair BA, Bowsher JE, Jaszczak RJ. Simultaneous reconstruction and motion estimation for gated cardiac ECT. *IEEE Trans Nucl Sci*. 2002;49(5):2344-2349. doi:10.1109/TNS.2002.803820.
  228. Chun SY, Fessler JA. Joint image reconstruction and nonrigid motion estimation with a simple penalty that encourages local invertibility. In: *SPIE Medical Imaging*. Vol 7258. ; 2009:72580U. doi:10.1117/12.811067.
  229. Fessler JA. Optimization transfer approach to joint registration/reconstruction for motion-compensated image reconstruction. In: *IEEE International Symposium on Biomedical Imaging*. ; 2010:596-599.
  230. Yang G, Hipwell JH, Hawkes DJ, Arridge SR. Numerical Methods for Coupled Reconstruction and Registration in Digital Breast Tomosynthesis. *Ann BMVA*. 2012;2012(x):1-29.
  231. Taguchi K, Sun Z, Segars WP, Fishman EK, Tsui BMW. Image-domain motion compensated time-resolved 4D cardiac CT. In: *SPIE Medical Imaging*. Vol 6510. ; 2007:651016. doi:10.1117/12.709980.
  232. Erdoğan H, Fessler JA. Monotonic algorithms for transmission tomography. *IEEE Trans Med Imaging*. 1999;18(9):801-814. doi:10.1109/42.802758.
  233. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging*. 1999;18(8):712-721. doi:10.1109/42.796284.
  234. Holden M. A review of geometric transformation for nonrigid body registration. *IEEE Trans Med Imaging*. 2008;27(1):111-128. doi:10.1109/TMI.2007.904691.

235. Dang H, Wang AS, Zhao Z, Sussman MS, Siewerdsen JH, Stayman JW. Joint Estimation of Deformation and Penalized-Likelihood CT Reconstruction Using Previously Acquired Images. In: *International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine.* ; 2013:424-427.
236. Klein S, Pluim JPW, Staring M, Viergever MA. Adaptive stochastic gradient descent optimisation for image registration. *Int J Comput Vis.* 2009;81(3):227-239.  
doi:10.1007/s11263-008-0168-y.
237. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Trans Med Imaging.* 2010;29(1):1-11.  
doi:10.1109/TMI.2009.2035616.
238. Cho Y, Moseley DJ, Siewerdsen JH, Jaffray DA. Accurate technique for complete geometric calibration of cone-beam computed tomography systems. *Med Phys.* 2005;32(4):968-983. doi:10.1118/1.1869652.
239. Mattes D, Haynor DR, Vesselle H, Lewellyn TK, Eubank W. Nonrigid multimodality image registration. In: *SPIE Medical Imaging.* Vol 4322. ; 2001:1609-1620.  
doi:10.1117/12.431046.
240. Otake Y, Schafer S, Stayman JW, et al. Automatic localization of vertebral levels in x-ray fluoroscopy using 3D-2D registration: a tool to reduce wrong-site surgery. *Phys Med Biol.* 2012;57(17):5485-5508. doi:10.1088/0031-9155/57/17/5485.
241. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image Quality Assessment : From Error Visibility to Structural Similarity. *IEEE Trans Image Process.* 2004;13(4):600-612.
242. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. *Neuro Image.* 2009;45(1 Suppl):S61-S72.  
doi:10.1016/j.neuroimage.2008.10.040.

243. Nithiananthan S, Brock KK, Daly MJ, Chan H, Irish JC, Siewerdsen JH. Demons deformable registration for CBCT-guided procedures in the head and neck: convergence and accuracy. *Med Phys*. 2009;36(10):4755-4764. doi:10.1118/1.3223631.
244. Nithiananthan S, Schafer S, Mirota DJ, et al. Extra-dimensional Demons: A method for incorporating missing tissue in deformable image registration. *Med Phys*. 2012;39(9):5718-5731. doi:10.1118/1.4747270.
245. Blake A, Zisserman A. *Visual Reconstruction*. MIT Press; 1987.
246. Stayman JW, Prince JL, Siewerdsen JH. Information Propagation in Prior-Image-Based Reconstruction. In: *The Second International Conference on Image Formation in X-Ray Computed Tomography*. ; 2012:334-338.
247. Gu X, Pan H, Liang Y, et al. Implementation and evaluation of various demons deformable image registration algorithms on a GPU. *Phys Med Biol*. 2010;55(1):207-219. doi:10.1088/0031-9155/55/1/012.
248. Nithiananthan S, Schafer S, Uneri A, et al. Demons deformable registration of CT and cone-beam CT using an iterative intensity matching approach. *Med Phys*. 2011;38(4):1785-1798. doi:10.1118/1.3555037.
249. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging*. 1997;16(2):187-198. doi:10.1109/42.563664.
250. Reaungamornrat S, Liu WP, Wang AS, et al. Deformable image registration for cone-beam CT guided transoral robotic base-of-tongue surgery. *Phys Med Biol*. 2013;58(14):4951-4979. doi:10.1088/0031-9155/58/14/4951.
251. Uneri A, Nithiananthan S, Schafer S, et al. Deformable registration of the inflated and deflated lung in cone-beam CT-guided thoracic surgery: initial investigation of a

- combined model- and image-driven approach. *Med Phys*. 2013;40(1):17501.  
doi:10.1118/1.4767757.
252. Pourmorteza A, Dang H, Siewerdsen JH, Stayman JW. Reconstruction of Difference using Prior Images and a Penalized-Likelihood Framework. In: *International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. ; 2015:252-255.
  253. Dang H, Siewerdsen JH, Stayman JW. Regularization design and control of change admission in prior-image-based reconstruction. In: *SPIE Medical Imaging*. ; 2014:90330O. doi:10.1117/12.2043781.
  254. Tan BB, Flaherty KR, Kazerooni EA, Iannettoni M. The solitary pulmonary nodule. *Chest*. 2003;123(1\_suppl):89S-96S.
  255. Funama Y, Awai K, Liu D, et al. Detection of nodules showing ground-glass opacity in the lungs at low-dose multidetector computed tomography: phantom and clinical study. *J Comput Assist Tomogr*. 2009;33(1):49-53. doi:10.1097/RCT.0b013e31815e6291.
  256. Abbas S, Min J, Cho S. Super-sparsely view-sampled cone-beam CT by incorporating prior data. *J Xray Sci Technol*. 2013;21(1):71-83. doi:10.3233/XST-130367.
  257. Qi J, Huesman RH. Theoretical study of lesion detectability of MAP reconstruction using computer observers. *IEEE Trans Med Imaging*. 2001;20(8):815-822.  
doi:10.1109/42.938249.
  258. Berger M, Müller K, Aichert A, et al. Marker-free motion correction in weight-bearing cone-beam CT of the knee joint. *Med Phys*. 2016;43(3):1235. doi:10.1118/1.4941012.
  259. Sisniega A, Stayman JW, Cao Q, Yorkston J, Siewerdsen JH, Zbijewski W. Image-based motion compensation for high-resolution extremities cone-beam CT. In: *SPIE Medical Imaging*. Vol 9783. ; 2016:97830K. doi:10.1117/12.2217243.

260. Watzke O, Kalender WA. A pragmatic approach to metal artifact reduction in CT: Merging of metal artifact reduced images. *Eur Radiol.* 2004;14(5):849-856. doi:10.1007/s00330-004-2263-y.
261. Yu Z, Thibault JB, Bouman CA, Sauer KD, Hsieh JH. Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization. *IEEE Trans Image Process.* 2011;20(1):161-175. doi:10.1109/TIP.2010.2058811.
262. Ramani S, Fessler JA. A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. *IEEE Trans Med Imaging.* 2012;31(3):677-688. doi:10.1109/TMI.2011.2175233.
263. Nien H, Fessler JA. Relaxed Linearized Algorithms for Faster X-Ray CT Image Reconstruction. *IEEE Trans Med Imaging.* 2016;35(4):1090-1098. doi:10.1109/TMI.2015.2508780.
264. McGaffin MG, Fessler JA. Alternating Dual Updates Algorithm for X-ray CT Reconstruction on the GPU. *IEEE Trans Comput Imaging.* 2015;1(3):186-199. doi:10.1109/TCI.2015.2479555.
265. Mow M, Zbijewski W, Sisniega A, et al. Brain Perfusion Imaging Using a Reconstruction of Difference Approach for Cone-Beam Computed Tomography. In: *SPIE Medical Imaging.* ; 2017.
266. Niu K, Yang P, Wu Y, et al. C-Arm Conebeam CT Perfusion Imaging in the Angiographic Suite: A Comparison with Multidetector CT Perfusion Imaging. *Am J Neuroradiol.* 2016;37(7):1303-1309. doi:10.3174/ajnr.A4691.

## Curriculum Vitae

Hao Dang was born in Changzhou, China on December 19, 1987. He attended high school at Changzhou No. 1 High School. He received his B.S. in Biomedical Engineering from Shanghai Jiao Tong University (SJTU) in Shanghai, China. At SJTU, Hao was a research assistant in the Image-Guided Surgery and Therapy Lab and worked on image registration using normalized gradient information and saliency information. Following his interests in interdisciplinary research, he was accepted to the Biomedical Engineering Ph.D. Program at Johns Hopkins University. His first semester of research at Johns Hopkins was in the Computer Science Department, where he worked on a lab rotation project with Dr. Peter Kazanzides in the Sensing, Manipulation, and Real-Time Systems (SMART) Lab on the development of a robotic system to assist in skull base surgery. Hao Dang's doctoral research was carried out in the I-STAR Lab (<http://istar.jhu.edu>) in Biomedical Engineering under supervision of Prof. Jeffrey H. Siewerdsen (primary advisor) and Prof. J. Webster Stayman (co-advisor). His Ph.D. dissertation focused on the development of new model-based iterative reconstruction methods to improve image quality – specifically, low-contrast, soft-tissue image quality – and reduce radiation dose in cone-beam computed tomography. His Ph.D. research culminated in multiple novel contributions to the field, leading to several published studies and conference presentations and subsequent translation of technology into clinical studies.