

Statistical Analysis and Evaluation of Near Infrared Tomographic Imaging System

A Thesis
Submitted to the Faculty
In partial fulfillment of the requirements for the
Degree of

Doctor of Philosophy

by

Xiaomei Song
Thayer School of Engineering
Dartmouth College, Hanover, New Hampshire
May 2005

Examining Committee

Brian Pogue, Ph.D. (Chair)

Keith Paulsen, Ph.D.

Tor D. Tosteson, Ph.D.

Marvin Doyley, Ph.D.

Eric Hansen, Ph.D.

Andreas H. Hielscher, Ph.D.

Dean of Graduate Studies

Xiaomei Song, Author

©2004 Trustees of Dartmouth College

ABSTRACT

Near-infrared (NIR) diffuse tomography is an emerging method for imaging the interior of tissues to quantify concentrations of hemoglobin and exogenous chromophores non-invasively in vivo. It often exploits an optical diffusion model-based image reconstruction algorithm to estimate spatial property values from measurements of the light flux at the surface of the tissue. In this thesis, three assessment studies of statistical accuracy of NIR tomographic imaging and computer-assisted diagnosis systems on computer simulated and phantom & patient study test data were present.

In mean-squared error (MSE) based statistical analysis, estimates of image bias and image standard deviation were calculated based upon 100 repeated reconstructions of a test image with randomly distributed noise added to the light flux measurements. It was observed that the bias error dominates at high regularization parameter values while variance dominates as the algorithm is allowed to approach the optimal solution.

In contrast-to-noise ratio (CNR) based statistical analysis, CNR was used to determine the detectability of objects within reconstructed images from diffuse near-infrared tomography. It was concluded that there was a maximal value of CNR

near the location of an object within the image and that the size of the true region could be estimated from the CNR. Experimental and simulation studies led to the conclusion that objects can be automatically detected with CNR analysis and that our current system has a spatial resolution limit near 4 mm and a contrast resolution limit near 1.4. A new linear convolution method of CNR calculation was developed for automated region of interest detection.

In NIR tomographic imaging system observer performance studies, receiver operating characteristic (ROC) Analysis on NIR tomography imaging system were performed and interpreted. Parametric and nonparametric approaches have been applied for estimating the ROC curve and two Matlab program were developed as observer test interface and ROC analysis tool. The results indicate that effects of iteration and algorithm performance significantly alter detectability of objects in NIR tomography images.

ACKNOWLEDGEMENTS

First and foremost of all, I would like to thank my advisor, Professor Brian Pogue, for his support, dedication and enthusiasm. I highly appreciate the experience and expertise he brings to the research topics that have been important elements of my work. I would like to thank him for believing in my abilities, and making the experience more interesting and enjoyable than I had ever imagined.

My thanks also go to Professor Keith Paulsen for his support, encouragement and guidance throughout this PhD research, as well as for sharing his great knowledge in the field of medical imaging. I am indebted to him for originally inspiring me to pursue a career in Biomedical Engineering

I would like to thank Professor Tor Tosteson and Marvin Doyley for thoroughly proofreading my papers and the draft version of this thesis for their useful comments which have been a basis for improvement of the text. Special thanks also go to Professor Eric Hansen and Andreas Hielscher for their thorough reading and comments on this thesis and serving as committee member of it.

My thanks go especially to Professor Hamid Dehghani and Dr. Shudong Jiang. This thesis would not have been possible without their tremendous contribution on software and hardware guidance and support throughout the entire research work of this thesis. Their drive, enthusiasm, work and knowledge have been very valuable for

my research.

I would like to acknowledge help and collaboration from members of NIR imaging group at Dartmouth including Ben Brooksby, Subhadra Srinivasan, Heng Xu, Phaneendra Yalavarthy, Summer Gibbs, Scott Davis and previous research by Troy McBride.

I should also mention the great support I got from the following people during my Ph. D study at Dartmouth College: Qianqian Fang, Chao Sheng, Qing Feng, Xiaodong Zhou, Kyung Park, Zhiliang Fan, Xin Wang, Xiang Li and Yijin He.

Christine Kogel at DHMC is gratefully acknowledged for her support in patient study of this thesis.

Furthermore, I would like to thank all the personnel at Thayer Engineering School for creating a nice working atmosphere.

Financial support was provided by the National Institutes of Health (NIH) under Grant PO1CA80139 and Grant RO1CA69544, and the Dartmouth Society of Engineers Fellowship

Finally my love and thanks go to my parents for their unwavering encouragement and support. I have to thank my husband who has helped give me the motivation to see this thesis through to completion. Now we can go play Wu Zi Qi...

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iv
Table of Contents	vi
List of Figures	ix
List of Tables	xvi
1 Introduction	1
1.1 Optical Imaging for Breast Cancer Detection Image Construction	4
1.2 Near Infrared Tomographic Imaging Research at Dartmouth College	7
1.2.1 Frequency Domain Diffuse Optical Tomography Theory	7
1.2.2 NIR Tomographic Imaging System	10
1.2.3 Experimental Research: Simulation, Phantom and Patient Studies	16
2 Mean-Squared Error (MSE) based Statistical Analysis of Nonlinearly Reconstructed Near-Infrared Tomographic Images	20
2.1 Introduction	21
2.2 Mean-Squared Error Definition	25
2.3 MSE Statistical Analysis of Computer Simulated NIR Tomographic Images	28
2.3.1 NIR Tomography Simulation Setup	28
2.3.2 Images of Interaction Coefficient and Error	29
2.3.3 Fixed versus Levenberg-Marquardt Regularization	32
2.3.4 Image Quality	35

2.3.5	Regularization Method	39
2.3.6	Noise Analysis	40
2.4	MSE Statistical Analysis of Phantom Study Images	42
2.4.1	Phantom Studies Setup	44
2.4.2	Phantom Studies Results	46
2.4.2.1	Images of Interaction Coefficient and Errors	46
2.4.2.2	Projection Error	57
2.4.2.3	Region of Interest (ROI) Analysis versus the Entire Image Field	60
2.5	Conclusions	63
3	Contrast-to-noise ratio based Contrast-Detail Analysis of System Performance	66
3.1	Introduction	66
3.2	Theory and Method	69
3.2.1	Contrast-Detail Curve	69
3.2.2	Contrast-to-Noise Definition of NIR Tomographic Images	72
3.3	Contrast-to-Noise Ratio Analysis of NIR Tomographic Simulation Images	76
3.3.1	Evaluating the Three Methods of Calculating the CNR	76
3.3.2	Comparison of ROI Location Detection: Optimized CNR Method vs. 2-D Convolution Method	79
3.4	Contrast-to-Noise Ratio Analysis of NIR Tomographic Phantom Study Images	82
3.5	Discussion	85
3.6	Conclusions	89
4	Observer Performance Studies with Receiver Operating Characteristic (ROC) Analysis	91
4.1	Introduction	91
4.2	ROC and Localization-Response ROC Analysis in Medical Imaging	94
4.2.1	Sensitivity and Specificity of Diagnostic Tests	94

4.2.2 Concept of ROC Curve	97
4.2.3 ROC Curve Fitting	99
4.2.4 Multiple-Reader Multiple-Case ROC and Location ROC	106
4.2.5 Summary Measures of the ROC Curve	107
4.3 ROC Analysis: Software Development at Dartmouth	109
4.4 Assessment of Near-Infrared Tomography Imaging System	113
4.4.1 Computer Simulations	113
4.4.1.1 NIR Tomography Simulation and Image Preparation	113
4.4.1.2 Detection and Localization Task	115
4.4.1.3 Computational Assessment	117
4.4.1.4 Results	118
4.4.1.5 Computer Simulation Conclusions	135
4.4.2 Phantom Studies	139
4.4.2.1 Phantom Studies Design and Image Preparation	139
4.4.2.2 Results and Discussion with a 6mm ROI	141
4.4.3 Patient Studies	144
4.4.3.1 Patients Studies Image Preparation	144
4.4.3.2 Detection Task	145
4.4.3.3 Results	146
4.5 Conclusions	148
5 Concluding Remarks	151
Appendix	154
References	159

LIST OF FIGURES

1.1	A typical medical diagnostic system diagnosis flow (red arrow) and system assessment process flow	2
1.2	Schematic outline of NIR tomographic imaging system at Dartmouth College	11
1.3	The rotary stage of the detection array	12
1.4	The fiber optic/patient interface	14
1.5	NIR tomographic imaging system at Dartmouth College	15
1.6	Illustration of the geometry used for the simulation studies	17
1.7	Illustration of the two procedures used in phantom study	18
2.1	Illustration of the geometry used for the simulation studies of MSE analysis	28
2.2	Reconstructed images of absorption coefficient and reduced scattering coefficient after 1, 2, 5, & 8 iterations	30
2.3	Calculated images of absorption coefficient bias error and standard deviation error	31
2.4	Calculated images of reduced scattering coefficient bias error and standard deviation error	32
2.5	Absolute bias and standard deviation for the absorption coefficient and reduced scattering coefficient using the Levenberg-Marquardt scheme	33
2.6	Absolute bias and standard deviation for the absorption coefficient and reduced scattering coefficient using the fixed levels of regularization parameter	34

2.7	Projection error as a function of iteration number using the Levenberg-Marquardt scheme	35
2.8	Absolute bias and standard deviation for the absorption coefficient after 1, 2, 5, & 8 iteration	38
2.9	The geometries of the phantoms used in the phantom studies for MSE Analysis	45
2.10	Reconstructed images of absorption coefficient and reduced scattering coefficient by the fixed decreasing regularization method for the phantom #1 after 1, 3, 5, & 8 iterations	46
2.11	Calculated images of bias error and standard deviation error for absorption coefficient after 1, 3, 5, & 8 iterations by the fixed decreasing regularization method for the phantom #1	47
2.12	Calculated images of bias error and standard deviation error for reduced scattering coefficient after 1, 3, 5, & 8 iterations by the fixed decreasing regularization method for the phantom #1	48
2.13	Reconstructed images of absorption coefficient and reduced scattering coefficient by the Levenberg-Marquardt regularization method for the phantom #1 after 1, 3, 5, & 8 iterations	49
2.14	Calculated images of bias error and standard deviation error for absorption coefficient after 1, 3, 5, & 8 iterations by the Levenberg-Marquardt regularization method for the phantom #1	50
2.15	Calculated images of bias error and standard deviation error for reduced scattering coefficient after 1, 3, 5, & 8 iterations by the Levenberg-Marquardt regularization method for the phantom #1	50
2.16	Reconstructed images of absorption coefficient and reduced scattering coefficient by the fixed decreasing regularization method for the phantom #2 after 1, 3, 5, & 8 iterations	51
2.18	Calculations of the absolute bias and standard deviation for absorption coefficient of the phantom #1	53

2.19	Calculations of the absolute bias and standard deviation for reduced scattering coefficient of the phantom #1	54
2.20	Calculations of the absolute bias and standard deviation for absorption coefficient of the phantom #2	55
2.21	Calculations of the absolute bias and standard deviation for reduced scattering coefficient of the phantom #2	56
2.22	Calculated the average projection errors of the phantom #1 & #2 by fixed decreasing regularization method and Levenberg-Marquardt regularization method	58
3.1	Formation of Contrast-Detail curve	70
3.2	Geometry of the simulation for optical property distribution, and (b) a typical reconstructed absorption coefficient image	73
3.3	Schematic showing the three possible choices of backgrounds for calculating CNR	74
3.4	Graphs of CNR distribution obtained by the three different methods of choosing the background areas	78
3.5	Reconstructed absorption images for target with large size, low contrast and target with small size, high contrast	79
3.6	Two methods of ROI detection	80
3.7	Distance error between the detected center of the ROI in the reconstructed image relative to the real location of the circular ROI center is plotted as a function of the diameter of the ROI	81
3.8	Photograph of the phantom used in our study	83
3.9	Reconstructed images of the absorption coefficient with the object size of 4 mm in diameter and the absorption contrast equal to 2:1, 4:1, 6:1, and 9:1.	84
4. 1	Figure 4.1 The 2x2 Diagnosis Table.	95
4. 2	Example of probability density distributions of an observer's confidence in a diagnostic test. The observer's diagnosis is made according to the confidence threshold, which varies to generate different (sensitivity, specificity) pairs and thus an ROC curve.	98

4. 3	The ROC curve achieved from Figure 4.2 as the observer's confidence threshold varies. The square point on the curve corresponds to the observer's confidence threshold line shown in Figure 4.2	98
4. 4	Procedures of Maximum-Likelihood Estimation Method Receiver Operating Characteristic Curve Fitting	111
4. 5	Screen Capture of the MATLAB GUI based ROC/LROC fitting platform, including: Main menu options, Input Data Information, ROC Analysis Output, Confidence Interval Boundary, Input Data And Fitted ROC Plots, Statistical Tests Results of Standard Errors, Data Correlation and Confidence Intervals.	112
4. 6	The NIR tomographic imaging system is shown along with a typical reconstructed image. (a) is the top view of the phantom and fiber interface; (c)-(d) are typical reconstructed NIR tomography images with the same size, and location of ROI, iteration number, and reconstruction algorithm but different contrast levels: (b) $C = 1.1$; (c) $C = 1.4$, and (d) $C = 2.0$	115
4. 7	The MATLAB image detection program is shown. Observers can adjust the image gray scale map with the top two slide bars, and point out the estimated center location of the region with the computer mouse	116
4. 8	(a) The distribution sample of human observer response is shown, which was fitted by kernel density estimation. The window width was chosen from the direct 'plug-in' method and the Epanechnikov kernel function was used (b) ROC and LROC curve for (a)	119
4. 9	Sample images from heterogeneity size studies are shown. The objects all have the same contrast $C = 2.0$, the same reconstruction process iteration number $I = 6$, but with different object sizes. In (a) a small sized object was used, with diameter of 6mm. In (b) a medium sized object was used, with diameter of 10mm. In (c) a large object of diameter 16mm was used.	121
4. 10	The heterogeneity size study for both human and computational observers is shown in (a) with ROC and area under curve data, and in (b) with LROC and area under curve	123

	data. Both of these are shown for data sets where the absorption contrast ranged from 1.1 to 2.0, with a fixed 6 iterations.	
4. 11	Sample images from heterogeneity contrast studies are shown. The objects had same size, with diameter of 10mm, same reconstruction process with iteration number $I = 6$ but with different contrasts. In (a) the contrast was $C = 1.4$; in (b) the contrast was $C = 1.7$, and in (c) the contrast was $C = 2.0$.	124
4. 12	The heterogeneity contrast study for both human and computational observers is shown, with (a) showing the ROC and area under curve, and (b) showing the LROC and area under curve. The diameter of ROI was equal 10 mm for all, and 6 iterations were used	125
4. 13	The mean of the projection error function of 294 heterogeneous reconstructed images is shown as the function of iteration number, using a diameter ROI of 10 mm and an absorption contrast range from 1.1 to 2.0. The error bars are the standard deviations of the 294 images for all of 20 iterations	128
4. 14	Sample images from the reconstruction process are shown for different iteration numbers. The images have same object size, $D = 10\text{mm}$, the same object-to-background contrast of $C = 2.0$, but are from different numbers of iteration in the algorithm. In (a) the result from 4 iterations is shown. In (b) the result from 6 iterations is shown, and in (c) the result from 10 iterations is shown. Higher numbers of iterations tend to recover higher spatial frequency information and higher contrast, albeit with the addition of a non-linear level of noise increased	128
4. 15	The reconstruction process iteration number was studied for both human and computational observers. In (a) the ROC and area under curve are shown, and in (b) the LROC and area under curve are shown, for a fixed diameter equal to 10 mm and with varying absorption contrast from 1.1 to 2.0	130
4. 16	Sample images from heterogeneity location studies are shown. The objects had same size, with diameter of 10mm, same reconstruction process with iteration number $I = 6$, and with the same contrast $C = 1.7$, but at different locations. In (a) the distance between the object and phantom centers was $D =$	133

11mm; in (b) the distance between the object and phantom centers was $D = 21\text{mm}$, and in (c) the distance between the edges of the object and phantom was $D = 6\text{mm}$

- | | | |
|-------|--|-----|
| 4. 17 | The heterogeneity location study for both human and computational observers. In (a) the AUC of ROC are shown, and in (b) the AUC of LROC are shown, for diameter of ROI equal to 6 (top pair of graphs) and 10 mm (bottom pair of graphs) and with varying absorption contrast from 1.1 to 2.0 | 134 |
| 4. 18 | Human observers and computed CNR performance comparisons are shown for paired sets of images analyzed. The human observer decision (confidence level that the image is abnormal) and computational observer decision (contrast-to-noise ratio) for same group of images are plotted in CNR-Human decision 2-D diagram. | 136 |
| 4. 19 | The geometries of the phantoms used in the studies for a circular object in solid phantoms whose size slightly varies. The diameters of the heterogeneity in all the 3 phantoms are 6 mm. (a) Phantom #1: $d = 84.6\text{ mm}$ and the absorption and scattering coefficients of homogeneous area are 0.0041 and 1.31 (b) Phantom #2: $d = 84.3\text{ mm}$ and the absorption and scattering coefficients of homogeneous area are 0.0033 and 1.16 (c) Phantom #3: $d = 84.4\text{ mm}$ and the absorption and scattering coefficients of homogeneous area are 0.0040 and 1.03. | 140 |
| 4. 20 | ROC curves for the heterogeneity location study from (a) the worst human observer performance and (b) the best human observer performance for 6mm ROI and with varying absorption contrast from 1.2 to 2.0 | 143 |
| 4. 21 | A screen shot of the Labview software interface of computer-aid program in our patient studies. the mean and standard deviation values, the maximum value and its location in x, y axis for each image, along with the display controls at the top of the field | 145 |
| 4. 22 | The ROC curves and the 95% confidence interval of HbT for both (a) large (size $> 6\text{mm}$) and (b) small (size $\leq 6\text{mm}$) breast cancer tumors vs. the normal tissue. | 147 |

4. 23 The ROC curves and the 95% confidence interval of StO₂ for both (a) large (size > 6mm) and (b) small (size ≤ 6mm) breast cancer tumors vs. the normal tissue. 148
4. 24 The ROC curves and the 95% confidence interval of water for both (a) large (size > 6mm) and (b) small (size ≤ 6mm) breast cancer tumors vs. the normal tissue. 148

LIST OF TABLES

4.1	Comparison of Parametric and Semi/Non Parametric ROC Fitting Approaches	105
4.2	Summary of ROC Software	109
4.3	Human Observers Performance Comparison	120
4.4	Parameter Settings in the Heterogeneity Size and Contrast Studies (Iteration number = 6)	121
4.5	Parameter Settings in the Reconstruction Process Iteration Number Study	127
4.6a	Parameter Settings in the Heterogeneity Location Study (Iteration number = 6)	131
4.6b	Parameter Settings in the Heterogeneity Location Study (Iteration number = 6)	132
4.7	Parameters of Resin Solid phantoms for Phantom Studies	140
4.8	Human Observers Performance Comparison with Different Contrast (C)	142

CHAPTER 1

Introduction

How accurate and effective is a particular medical diagnostic imaging system? How can the efficacy and efficiency of an existing medical imaging system be improved? These are two of the most common questions radiologists must face throughout their career. A medical diagnostic system can be divided into two composite modules: an image construction module and an image processing and interpretation module, as described in Figure 1.1. In the image construction module, the inputs are patients, diseased or healthy, and the outputs are images containing patient information. The combination of hardware and imaging capacity of any particular medical diagnostic imaging system are distinct from any other imaging system. In the image processing and interpretation module, radiologists read and analyze the images and give final diagnoses with or without the help of computer assisted diagnosis systems. To understand the performance of a medical imaging system and improve its efficacy and efficiency thereafter, an assessment system module, as indicated in Figure 1.1, is also necessary. Generally speaking, the image interpretation module and the imaging system assessment module can be combined and form a so-called "post processing" module.

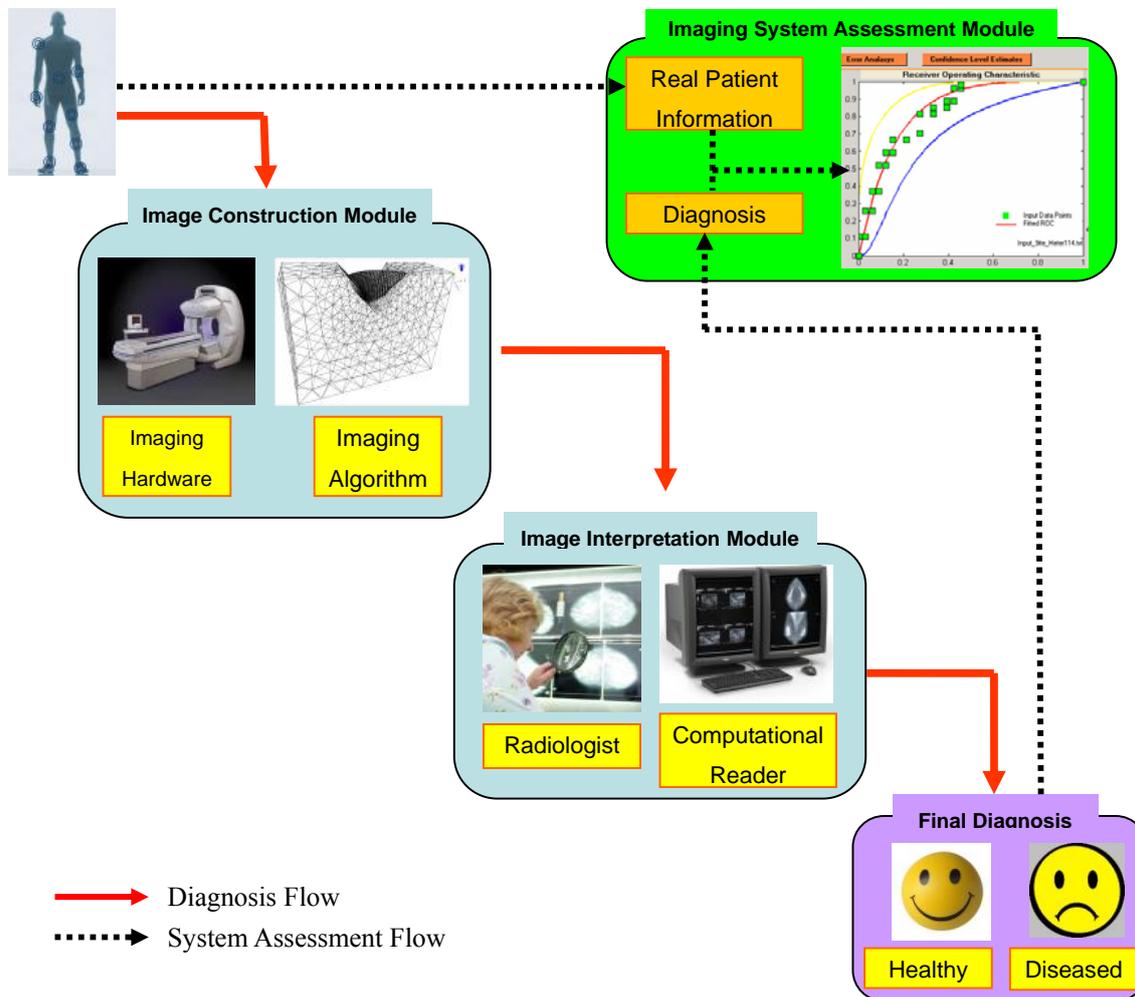


Figure 1.1 A typical medical diagnostic system diagnosis flow (red arrow) and system assessment process flow (black arrow) within image construction module, image processing and interpretation module and imaging system assessment module.

In recent years, as traditional medical imaging modalities mature and many new imaging algorithms emerge, there is more attention devoted to the assessment of medical imaging and computer-assisted diagnosis systems. This trend emerges in two quantifiable features: firstly, technical improvements such as the use of a Contrast-Detail Curve and Receiver Operating

Characteristic (ROC) analysis provide objection assessment of system performance; secondly, the experience and information learned from system assessment studies has led to computer assisted detection systems, which requires more informative and dedicated medical imaging system assessment technologies.

In this thesis, two new post-imaging analysis tools for NIR diffusion tomographic imaging are present and a complete system assessment study are performed and analyzed for the breast imaging work going on at Dartmouth College. The following sections in this chapter briefly reviewed optical imaging in breast cancer early detection with focus on NIR tomographic imaging technologies and the practice currently in clinical trials.

In Chapter 2, a mean-squared (MSE) based image analysis is presented and the results from several simulation, phantom and patient studies are discussed.

The contrast-to-noise ratio (CNR) is introduced for NIR imaging analysis in Chapter 3 to analyze the detectability of objects within reconstructed images from diffuse NIR tomography. A new linear convolution method for CNR calculation is also developed in Chapter 3 for using in an automated region of interest (ROI) detection tool.

In Chapter 4, receiver operating characteristic (ROC) analysis based observer performance studies on the NIR tomography imaging system are performed and interpreted. Parametric and nonparametric approaches have been applied for estimating the ROC curves. The results in Chapter 4 indicate that effects of iteration and algorithm performance significantly alter the detectability of objects in NIR tomography images. Location ROC (LROC) analysis is also introduced to examine the ability to observers to find the position of objects in NIR images. Finally these tools are used to interpret the results of the ongoing clinical trial to assess the benefit of NIR tomography as a tool in breast tumor

characterization.

1.1 Optical Imaging for Breast Cancer Detection

Electromagnetic wave and mechanical wave imaging has been an area of research interest in medical imaging in the past several decades. The idea of using electromagnetic waves in breast cancer diagnosis was first proposed in 1929, when a visible light was used as the projection imaging source and large breast tumors were successfully detected [16]. Since then, many kinds of radiation were reported to be used in breast cancer imaging research. These include various electromagnetic wavelengths, such as X-ray, ultraviolet, visible & infrared light, microwave and RF waves, and mechanical waves such as ultrasound or elastography. Depending on the intrinsic characteristics of these waves and the physics of the interactions between the waves and human tissue, one or several of following mechanisms are used including transillumination, reflection, refraction, diffusion, diffraction, scattering, interference, modulation and re-emission of radiation [6].

While magnetic resonance imaging (MRI) and X-ray computed tomography (CT) give very successful clinical images and have become the standard medical diagnostic methods for visualizing and examining soft human organs and bones, their drawbacks remain. It is well known that CT uses ionizing radiation at high doses and MRI requires a very expensive infrastructure, which lead to recent renewed interest in alternative modalities such as optical imaging. A properly designed optical imaging system is believed to be non-invasive to the human body, relatively inexpensive, physically compact and requires a small space. Another reason for the interest in optical imaging is the fact that it is able to provide broad and direct

physiological information about human organs, such as changes in hemoglobin concentration, oxygen saturation, water and lipid concentrations, while CT and MRI provide primarily anatomical information of the tissue. This difference may lead to advantages of optical imaging over CT or MRI in some medical diagnostic applications, such as breast cancer early detection or brain imaging [29, 34, 73, 83, 87].

The development of photon migration and diffusion theories to model light travel in highly scattering media has led to theoretical and experimental advances in optical spectroscopy and imaging of the human breast in the last few years. Light travels in a diffuse pattern in highly scattering media, and numerical modeling of this phenomenon has allowed the development of efficient and accurate Diffuse Optical Tomography (DOT). The wavelengths used in DOT are at the red end of the visible spectrum and in the near-infrared range, between 650 to 950 nanometers, where light propagates relatively deeply through tissue, up to 10 - 12 centimeters. Therefore diffuse optical tomography is also referred as NIR tomographic imaging in some literature, emphasizing the spectral location. This imaging approach provides a diagnostic modality to non-invasively quantify oxygen saturation, hemoglobin concentration, water concentration, scattering and potentially exogenous chromophores in tissues [62, 73, 99]. It takes advantage of the optical contrast of both absorption and scattering between normal and tumor tissues in vivo. In the approach used in this thesis, for instance, frequency domain light signals are used to quantify both absorption and scattering images of the breast, using transmitted amplitude and phase shift of intensity modulated near-infrared light.

In practice, however, since light traveling between two points on the surface experience multiple scattering events and result in a very weak intensity, multiple sources and detectors

are usually used at the body surface to maximize the response information. Using the surface transmission profile from the imaging system, together with a light propagation model, an iterative algorithm is used to minimize the difference between the measured data and the model data is used to reveal optical properties of the body being imaged. This process is the "image reconstruction". While some sophisticated models are used in DOT [23], the diffusion approximation based model is widely used in biological tissues where scattering is much more probable than absorption [31, 41, 57, 62, 73].

In recent years, a lot of effort has been made to improve separation of absorption from scatter in the image formation process. These efforts include time-domain and frequency-domain techniques. Time-resolved imaging techniques use a picosecond or femtosecond laser to illuminate the Region of Interest (ROI) surface via fiber optics, and various detection optics are employed to collect the response signals, which consist of histograms of photon flight times produced by measuring the photon arrival time with respect to the laser-generated reference signal [31, 32, 58]. Alternatively, in frequency-domain, the intensity of the light sources is sinusoidally modulated at a frequency of up to several hundred megahertz and similar detection schemes are used to measure the demodulation and phase shift of the light diffusely transmitted through the ROI. The independent measure of phase and amplitude provide sufficient information to quantitatively separate absorption and scatter coefficients.

Frequency-domain diffuse optical tomography is used throughout this thesis study. In the next sections, the theory and methods we used for frequency-domain DOT is briefly discussed and the operation of the NIR tomographic imaging system used for phantom and patient studies is presented.

1.2 Near Infrared Tomographic Imaging Research at Dartmouth College

1.2.1 Frequency Domain Diffuse Optical Tomography Theory

The diffusion equation for NIR tomographic imaging can be derived from the radiative transport equation under certain approximations [63, 64, 68] as:

$$-\nabla \cdot D(r)\nabla\Phi(r,\omega) + [\mu_a(r) + i(\omega/c)]\Phi(r,\omega) = S(r,\omega) \quad (1-1)$$

Where $\Phi(r,\omega)$ is the radiance, $D(r)$ is the diffusion coefficient, $\mu_a(r)$ is the absorption coefficient, $S(r,\omega)$ is the source, c is the light speed in the medium, and ω is the modulation frequency. The diffusion coefficient can be written as

$$D(r) = \frac{1}{3[\mu_a(r) + \mu'_s(r)]} \quad (1-2)$$

where $\mu'_s(r)$ is the reduced scattering coefficient (also referred to as transport scattering coefficient).

Forward calculations with the diffusion equation are accomplished with a finite element numerical solution using a circular mesh with a set of regularly spaced node positions [65]. The diffusion equation is solved for the optical flux at the boundary positions corresponding to the detector locations based upon source locations matching the fiber positions used in the system, as described in next section [49]. The amplitude and phase of this signal is calculated at 100 MHz to provide the simulated set of measurements, Φ_M .

The image reconstruction approach we use has been described in detail elsewhere [52, 66, 70, 76]. Briefly, a Newton-Raphson approach to iteratively solve for the spatial distribution of optical properties μ , which minimizes the squared error between sets of measured and calculated data. Thus, the objective function to be minimized has the form:

$$\chi_k^2 = \frac{\|\Phi_M - \Phi_C(\mu^k)\|^2}{M\sigma^2} \quad (1-3)$$

where $\Phi_C(\mu^k)$ is the calculated set of data based upon the estimated distribution of interaction coefficients, μ^k , at the k^{th} iteration. M is the number of measurements and σ is the average standard deviation in these measurements. Minimizing Equation (1-3) by optical property parameter variation is non-linear, but can be realized through iterative solution for μ , based upon an initial estimate of the distribution. The matrix solution has the form [70]:

$$\mu^{k+1} = \mu^k + [\mathbf{J}(\mu^k)^T \mathbf{J}(\mu^k) + \lambda \mathbf{I}]^{-1} [\mathbf{J}(\mu^k)]^T (\Phi_M - \Phi_C(\mu^k)) \quad (1-4)$$

with $\mathbf{J}(\mu^k)$ being the sensitivity or Jacobian matrix whose elements are given by the partial derivatives $\mathbf{J}_{m,j}(\mu^k) = \partial\Phi_m / \partial\mu_j^k$, where m is the index for the particular source-detector measurement and j is the index for each node location. Because this problem involves the inversion of an ill-posed matrix $\mathbf{J}^T \mathbf{J}$, it is typically regularized by adding the identity matrix, \mathbf{I} , multiplied by an arbitrary parameter λ . In previous studies [65, 76], we found that different methods of determining the regularization parameter can change the reconstructed image significantly. In the study of the thesis, fixed decreasing method where

$$\lambda_{k+1} = \frac{\lambda_k}{C} \quad (C \text{ constant and } C = \sqrt{10} \text{ in this study}) \quad (1-5)$$

and Standard Levenberg-Marquardt algorithm where

$$\lambda_{k+1} = \begin{cases} \frac{\lambda_k}{\sqrt{10}} (\chi_k \leq \chi_{k-1}) \\ \lambda_k \times \sqrt{10} (\chi_k > \chi_{k-1}) \end{cases} \quad (1-6)$$

is used to efficiently reconstruct the tomographic images.

Alternatively, Equation (1-4) can be formally derived by adding a factor of $\lambda \|\mu - \mu^0\|^2$ to the initial objective function (Equation (1-3)) to yield an approximate form of this expression [70]. In practice, Equation (1-4) is typically solved with a homogeneous guess of the initial values for μ^0 , which can be derived from an overall average of the tissue properties derived from the measurement data [67].

Procedures for efficiently determining the regularization parameter have been studied in detail in theoretical situations, yet there is considerable freedom to arbitrarily set the value of this parameter during the reconstruction process. In previous studies a standard Levenberg-Marquardt algorithm has been used for stepping down the regularization parameter as the number of iterations progresses [65]. However, fixing the regularization parameter to a single value during the reconstruction has been found to increase the level of stability which can be achieved in the resulting images, especially when the algorithm is applied to multiple samples with varying levels of contrast. Empirical studies indicate that Levenberg-Marquardt style methods of lowering the regularization parameter based upon the convergence of the objective function can lead to images which are not linearly correlated to their true contrast value. Further, if the regularization gets too low, images become dominated by random property value variations. However, we have found that it is not possible to accurately recover objects with different sizes without varying the regularization

parameter. The process of varying the regularization parameter allows the inversion solution to contain spatial information which preserves different levels of resolution [76]. This tradeoff between statistical accuracy versus optimal resolution and contrast is important, and is investigated in chapter 2 of this paper through visualization and quantification of the image error in terms of statistical measures of bias and variance.

1.2.2 NIR Tomographic Imaging System

The frequency domain NIR tomographic imaging system at Dartmouth College consists of four parts: light delivery, detection array, fiber optic patient interface and computer control and electronics [46], as shown in Figure 1.2.

Six directly modulated lasers are used at wavelengths of 660nm, 761nm, 785nm, 808nm, 826nm and 849nm in the light delivery section. Modulation signal of 100 MHz from the signal generator and the DC laser drive current are combined with a bias-T and input into a 1x6 RF switch controlled by the computer. The 6 outputs of the switch are used to drive those 6 lasers, of which the output lights are combined into a single fiber by a 6x1 optical combiner, as illustrated in Figure 1.2.

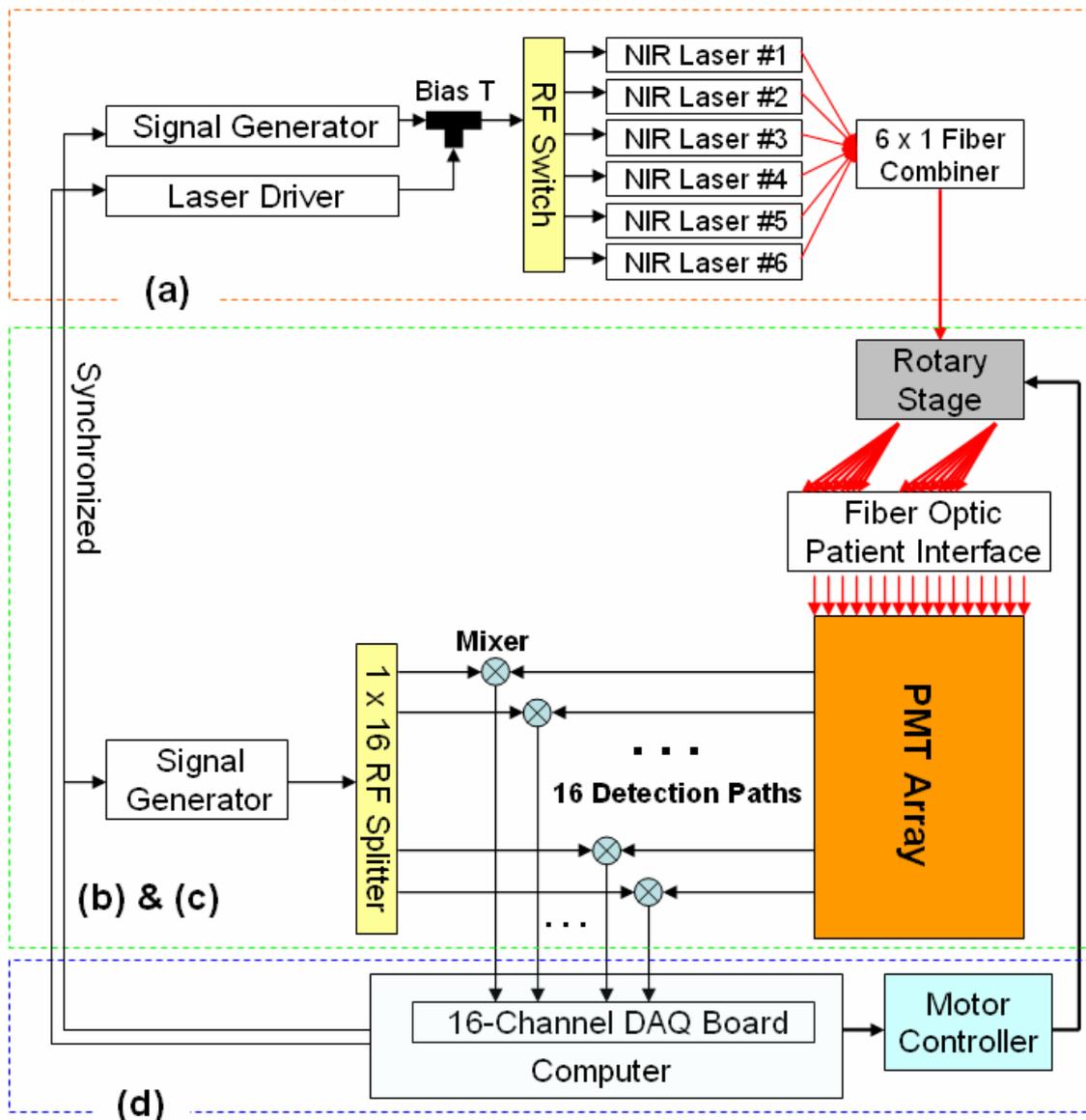


Figure 1.2 Schematic outline of NIR tomographic imaging system at Dartmouth College: (a) Light Delivery System, (b) Detection Array, (c) Fiber optic /Patient interface and (d) Computer Control System; Black arrow represents electrical signal paths and Red arrow represents optical signal paths transmitted in fiber optics.

In the detection array section, of which partial views are shown in Figure 1.3, sixteen Photo-Multiplier Tubes (PMTs) are used to detect the light from the fiber optic/patient

interface. Since the detected signal from the PMTs is at a high frequency (100 MHz), heterodyning of the signal is the common technology, as used in radio and wireless phones, to convert the MHz signal into a low frequency signal in the kHz range. For this purpose, a 2nd signal generator is used at a frequency of 100.0001 MHz and the signal is split and delivered to sixteen mixers for each detection channel. Two 16-way splitters are used to achieve multiplexing, one for splitting the DC power supply for the sixteen PMTs and the other one for splitting the local oscillator input to the sixteen RF mixers, as illustrated in Figures 1.2 and 1.3.

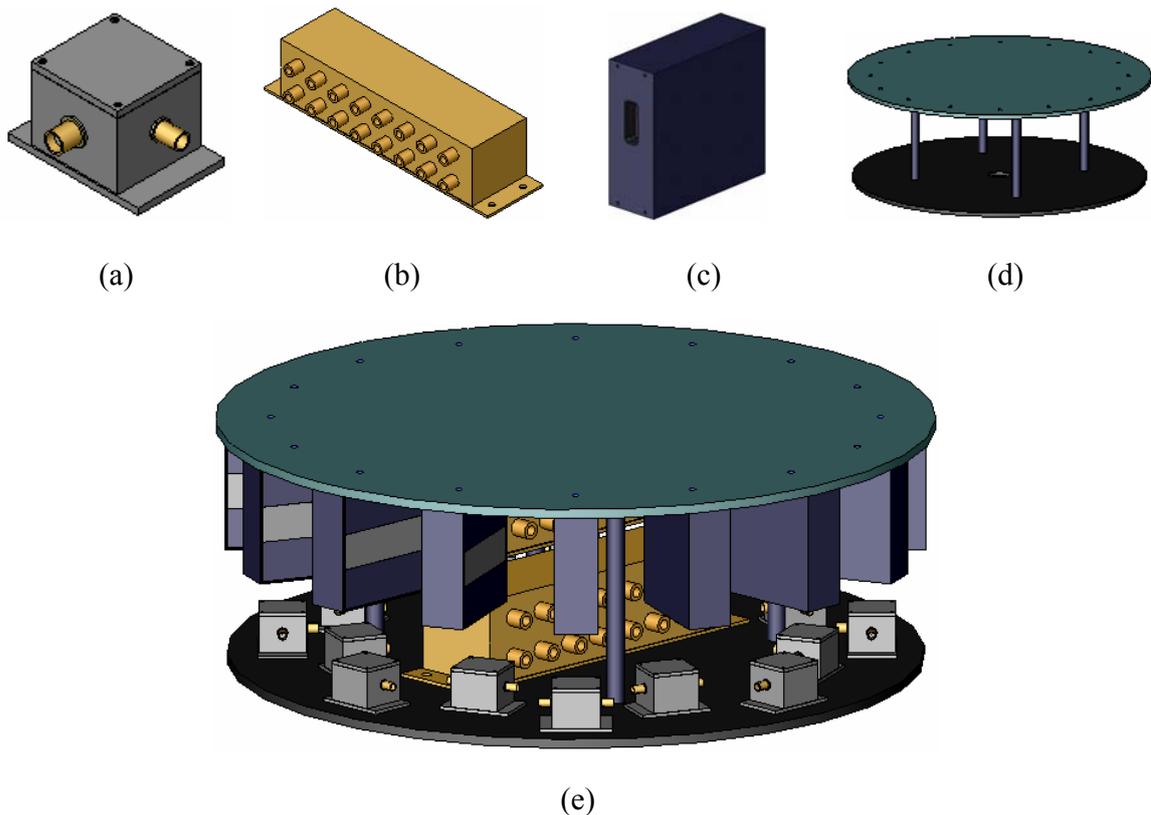
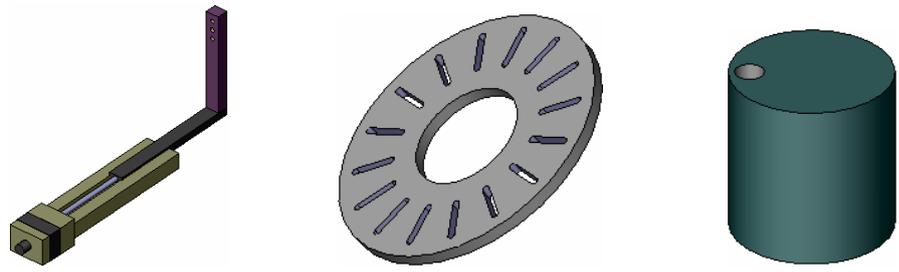
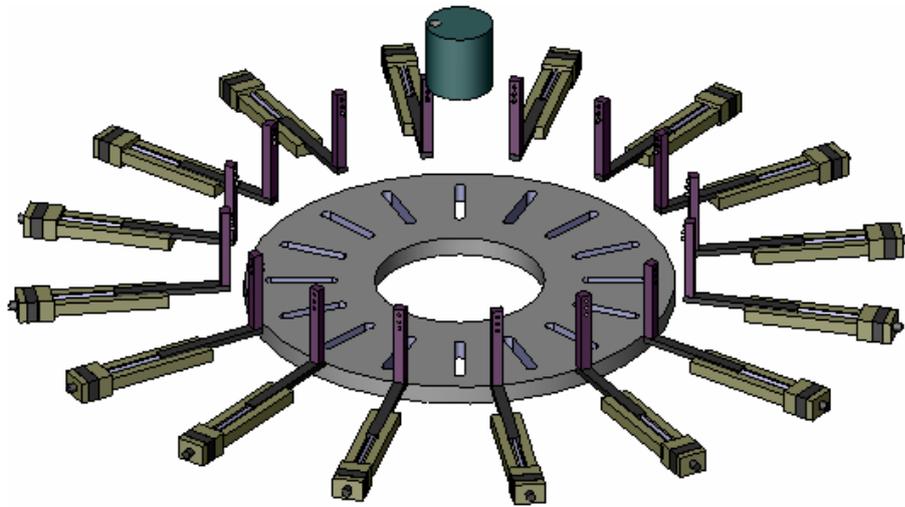


Figure 1.3 The rotary stage of the detection array (e). It consists of sixteen (a) mixers, (b) 1×16-way RF splitters, (c) PMTs (Photo-Multiplier Tube) and (d) a mounting disc.

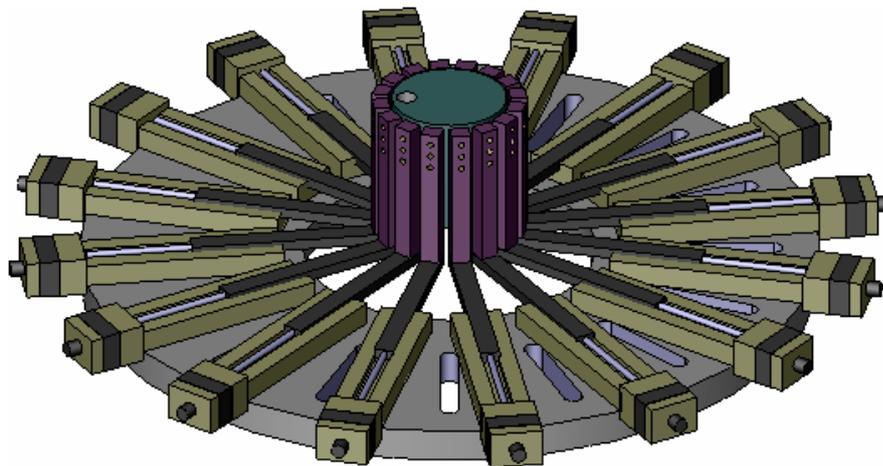
The fiber optic/patient interface is designed to be a circular geometry with adjustable diameter from 4 to 20 cm to fit different breast size. Direct fiber contact is used to deliver the light into the object under test. Sixteen motorized linear translation stages are arranged in a circle for radial positioning of the fiber array, as illustrated in Figure 1.4. Stepper motors are used to control the interface moving vertically and radially. Each radial translation bar is configured to have three fiber insert positions providing additional dimensional and object size choice. The whole fiber optic/patient interface section is placed under a patient bed with a hole through which the patient can place her breast into the circularly arranged interface bar array. In phantom studies, a fixed stand (not shown in Figure 1.4) is used to hold the phantom to be inserted into the patient interface. Figure 1.5 is the assembly view of the whole detection array and patient interface. In this thesis study, the system is configured to allow injection of a source of light into one of the sixteen fibers, while the other 15 bundles remain connected to PMTs for light detection. This unique arrangement is designed to allow the PMTs to measure from all 16 source positions without changing the gain on the detector, thereby minimizing the time for data acquisition. However for each breast or tissue phantom examined, the response of the PMTs at different gain levels requires calibration, and so this calibration is automatically controlled by the computer immediately prior to the imaging session [49].



(a)

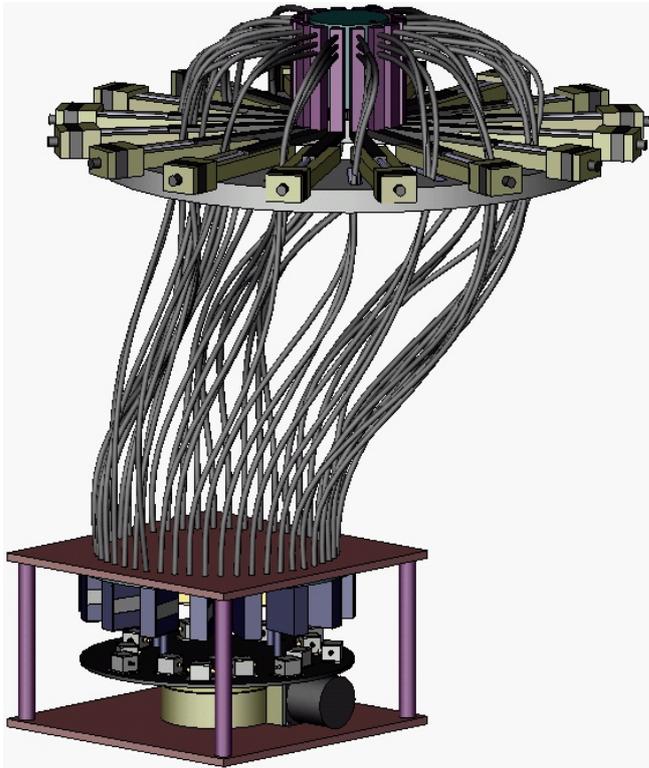


(b)



(c)

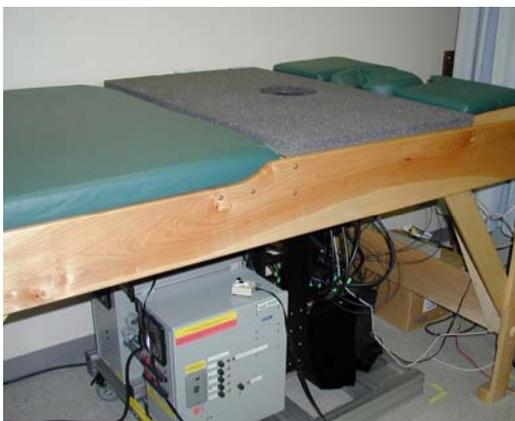
Figure 1.4 The fiber optic/patient interface. (a) one translation stage, mounting platform and phantom (b) Explode view of the radial translation stages on the circular platform (c) 3D assembly view of (b)



(a)



(b)



(c)



(d)

Figure 1.5 NIR tomographic imaging system at Dartmouth College. In (a), a schematic visualization of the system is shown for clarity. The lower source-detector array at the bottom has a fiber optic in-put for the source and has a fixed set of 15 PMTs that rotate around and align with each of the fibers. The photograph of the hardware in the lab is shown in (b), and in (c) placed underneath the patient table. In (d) a human subject having a test is shown on the patient bed [68]

The computer control system is an IBM type PC running Labview, the graphical programming development environment (National Instruments Corporation, Austin TX) It is programmed to fully operate the light generation and detection, including switching the different laser sources, rotating the detection array, calibrating and setting up the PMTs and fiber optic/patient interface and acquiring the data. Data is collected at all 16 measurement sites for each source position using a parallel acquisition strategy.

1.2.3 Experimental Research: Computer Simulation, Phantom and Patient Studies

The purposes of experimental research are to evaluate or optimize the reconstruction algorithm, and to assess the NIR tomographic imaging hardware system. In this thesis research, we designed three performance studies: computer simulations, phantom and patient studies. Circular geometry is used in computer simulation, as illustrated in Figure 1.6. The diameter of the inclusion object is ranged from 2 mm to 26 mm depending on the purpose of the studies.

The optical properties, i.e. the absorption coefficient μ_a and reduced scattering coefficient μ_s' , of the object inside the simulation field, are chosen to provide a certain optical contrast for different research purposes. Forward calculations from diffusion theory are used as the simulated data. Noise is included in the data with amplitude dependent upon the test, but in general the average percentage noise added to the amplitude is linked to the average degrees of noise added to the phase signal. For example, standard measurements with the

experimental system exhibit a random variation in amplitude near 1% and a corresponding random variation of 0.5 degrees standard deviation in the phase shift. These levels are taken to be characteristic of detector behavior and are used as representative noise levels in several of the simulation studies completed here.

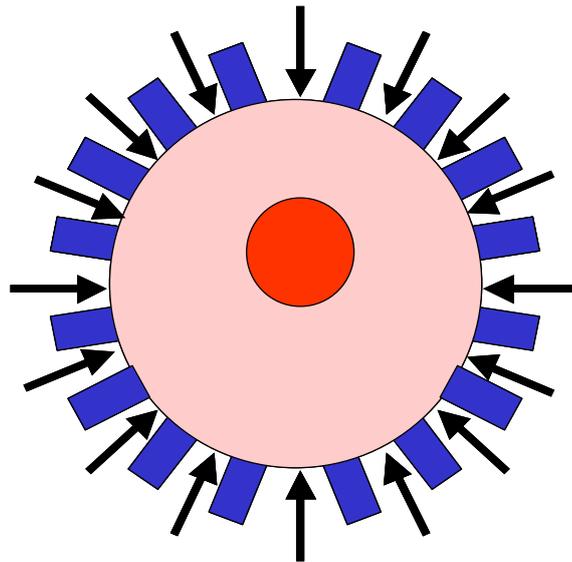


Figure 1.6 Illustration of the geometry used for the simulation studies, consisting of a circular object inclusion at a 2:1 contrast in the absorption coefficient. Orientation of the sources (arrows) and detectors (rectangles) are shown surrounding the circular object in an alternating pattern.

Gelatin phantoms are used to mimic the physical and optical properties of the human breast. The gelatin phantoms usually are placed in the imaging array to acquire tomographic data prior to cutting the hole. A homogenous fit to calculated data from diffusion theory was performed in this step to allow accurate prediction of the absorption and reduced scattering coefficient in the gelatin phantom.

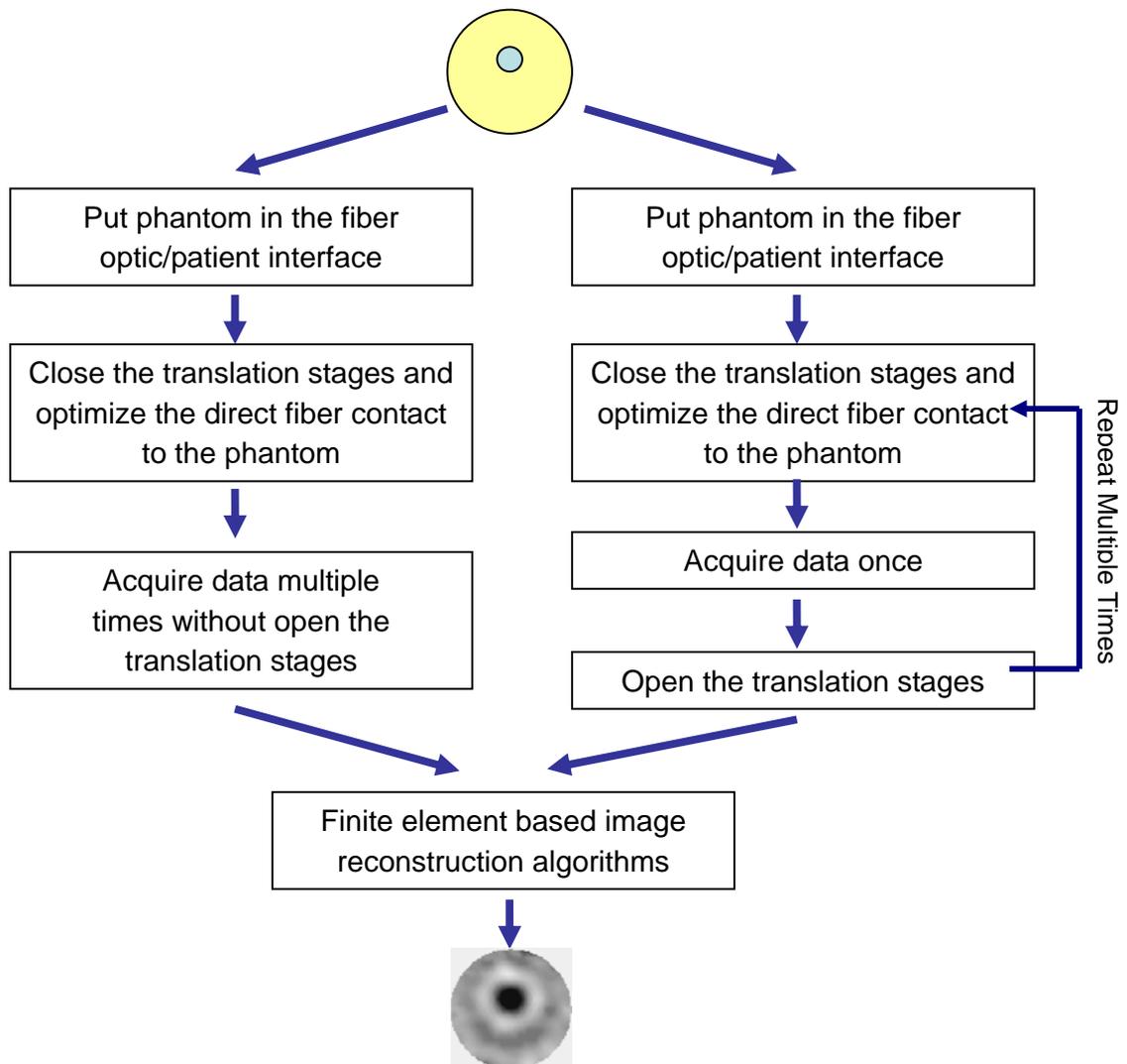


Figure 1.7 Illustration of the two procedures used in phantom study

The phantoms were imaged following the two procedures outlined in Figure 1.7. During procedure #1 (on the left in Figure 1.7), the fiber optic interface was fixed and multiple data acquisitions were performed sequentially, without moving either the fiber optic array or the phantom. In procedure #2 (on the right in Figure 1.7), the fiber optic interface, (i.e, the source and detection fiber) was repositioned after each data acquisition. During this

process, the linear translation stages which move the source and detection fibers were opened until there was no direct contact with the phantom. The array was then closed to re-establish a connection between the fibers and the phantom. As specified in [74], the total error consists of a sum of image bias and variance.

Each time the fiber optic interface was reset and the source and detection fiber created contact with the phantom, the positioning error of the fiber optic interface introduced another source of image bias in the reconstruction process. In procedure #1, the same image bias exists in each of the multiple sets of data because the standard deviation associated with the positioning of the fiber optic interface is zero, leaving only deviations due to the PMT detector random noise. In procedure #2, on the other hand, the image bias of each of the acquired data sets is randomly distributed thus introducing a standard deviation in the fiber optic interface position on top of the detector random noise.

Patient studies are considered to be the ultimate goal of any NIR tomographic imaging research. In the NIR tomographic imaging system assessment study presented in Chapter 4, a series of patient study are performed, where 95 female patients are tested in the system and a Receiver Operating Characteristic (ROC) analysis based observer performance studies are presented.

CHAPTER 2

Mean-Squared Error (MSE) based Statistical Analysis of Nonlinearly Reconstructed Near Infrared Tomographic Images

As discussed in Chapter 1, near-infrared tomographic (NIR) imaging often exploits an optical diffusion model-based image reconstruction algorithm to estimate spatial property values from measurements of the light flux at the surface of the tissue. In this chapter, mean-squared error (MSE) over the image is used to evaluate methods for regularizing the ill-posed inverse image reconstruction problem in NIR tomography. Estimates of image bias and image standard deviation were calculated based upon 100 repeated reconstructions of a test image with randomly distributed noise added to the light flux measurements.

In this chapter, we first carried out an image error analysis on simulated data of a diffuse near-infrared (NIR) tomography system. Then the same methodology was used in section 2 on experimental data acquired on phantoms with a prototype imaging system intended for characterizing breast tissue. The purpose of this study was to find out whether and how imaging performance is limited by random measurement error or by calibration issues. The

results indicate that the image error over the entire field of view is generally not minimized when an accurate homogeneous estimate of the phantom properties is available; however, local image error over a target region of interest (ROI) is reduced. The image reconstruction process which includes a Levenberg–Marquardt style regularization provides good minimization of the objective function, yet its reduction is not always correlated with an overall image error decrease. Minimization of the bias in an ROI which contains localized changes in the optical properties can be achieved through five to nine iterations of the algorithm. The studies also show that pre-calibration of the algorithm through statistical evaluation of phantom studies may provide a better measure of the image accuracy than that implied by minimization of the standard objective function. Furthermore, the test data indicates that the bias error dominates at high regularization parameter values while variance dominates as the algorithm is allowed to approach the optimal solution. This optimum does not necessarily correspond to the minimum projection error solution, but typically requires further iteration with a decreasing regularization parameter to reach the lowest image error. Increasing measurement noise causes a need to constrain the minimum regularization parameter to higher values in order to achieve a minimum in the overall image MSE.

2.1 Introduction

While both preclinical [35, 69, 75, 76] and clinical studies are underway [15, 27, 52, 73], the implementation of efficient and accurate image reconstruction algorithms is still

evolving. NIR reconstruction schemes typically involve a nonlinear inverse solution to an ill-posed problem that requires matrix regularization methods [2, 3, 37, 65]. While several promising approaches have been developed there has been comparatively little consideration given to the statistical accuracy of the resulting images [98]. This issue becomes particularly important for functional imaging that responds to a combination of physiological and anatomical factors that are otherwise difficult to predict in advance. Knowledge about the validity of the resulting image is important for producing information sufficiently reliable for clinical decision-making. To this end, we have applied standard concepts for calculating the mean error in an estimate, including bias and standard deviation, to analyze reconstructed NIR tomography images. This analysis is used to examine the effects of varying the reconstruction parameters on image accuracy under alternative assumptions about system measurement noise.

In previous studies [98], the concept of using maps of the surface of confidence ellipsoids for entire images was introduced as a method of visually displaying and analyzing the statistical uncertainty in our NIR reconstructions. The proposed methods rely on the theoretical properties of converged least squares estimates. However, because the problem of NIR image reconstruction from projection measurements is ill-posed, the inversion process seldom converges to a unique minimum, and different choices for structuring the minimization of the objective function can lead to different images. These numerical difficulties result from the incompleteness of the imaging data, both in terms of its precision and number of observations, and the nonlinear relationship between the image variable (i.e.,

optical properties) and the measured data set (i.e., signal amplitude and phase shift). Given these limitations, it is necessary to use empirical methods to study image accuracy in order to judge the validity of reconstructed results in typical clinical settings. Since an image can be considered as a large array of scalar data, we have examined the standard deviation and bias errors of all reconstructed values within the image, thereby creating spatial distributions of the total image error.

The major factors which affect image reconstruction are: 1) the method of regularization; 2) the regularization parameter; and 3) the type and level of noise in the data. In previous studies [40, 41, 72], regularization was achieved through an adaptation of the fixed-reduction (FR) method [77]. In this approach, the regularization parameter is decreased in successive iterations as long as the calculated objective function also decreased. During this process there is significant empirical flexibility in choosing the initial regularization parameter and its rate of descent. Further, it is not clear if decreasing this parameter beyond a certain level leads to reliable images [77]. In practice, constraints must be established for useful upper and lower bounds on the regularization parameter. For example, in some cases we have empirically found that more reliable images are obtained when this parameter remains fixed at its lowest value through all iterations [70]. It is also possible to consider reconstructing a fixed number of iterations which are chosen based on pre-calibration phantom studies. The advantage of the statistical methods we use here is that they provide a framework for quantitatively evaluating competing regularization options, at least under the controlled conditions of simulated data and laboratory phantom studies.

Constraints on the regularization parameter are related to noise in the measurements as well as the limitations of the projection data. In the model-based image reconstruction approach, there are several different forms of "noise" which we consider either as: 1) random noise (i.e., typically from stochastic detector fluctuations); 2) quasi-random noise (i.e., typically from systematic experimental variations such as changes in fiber positions); and 3) mismatch noise (i.e., typically from mismatch between the theoretical model and the physical light transport process). Each of these types of noise can independently alter the solution to the regularization inverse problem, and impact the creation of an optimal approach. Further, the latter two forms of noise may impact the objective function in a way which limits its accurate minimization.

On other hand, in the phantom and patient studies, the model-based image reconstruction approach has several forms of variation which we classify as: 1) purely random Gaussian distributed noise, typically from stochastic detector fluctuations; 2) detector calibration changes, model mismatches and quasi-random systematic experimental variations such as fiber positioning; and 3) physiologic noise due to movement, respiration and blood flow in the tissue being examined. In the current study, we limit our analysis to the first two "system" related noise sources leaving the physiologic factors for later investigation once the system response is better characterized. Inherent limitations in photomultiplier tubes generate the random light noise detection of approximately 1% in amplitude and about 0.5 in phase shift at 100-MHz signal frequency [49]. While it is important to assess the effects of these random variations in the resulting images, it is unlikely that the entire observed behavior can

attributed to random noise by itself. Here, we consider the relative contributions of random noise from the detectors as well as quasi-random variation from the fiber optic positioning system and model mismatches incurred from the two-dimensional diffusion equation approximation employed by the reconstruction algorithm.

In this chapter, we empirically explore the impact of noise on the resulting image error using statistical methods to guide the selection of a regularization strategy which limits this error. In section 2.2, we introduce the definition of mean-squared error (MSE) in NIR tomographic imaging. Section 2.3 reviews the important features of MSE based statistical analysis on simulated NIR tomography images. Simulation experiments are conducted which eliminate the quasi-random and mismatch noise components that exist in practice in order to first investigate the impact of random measurement error on choice of regularization parameter. In section 2.3, we extend these ideas to include images and data recorded during laboratory phantom experiments which contain all forms of noise inherent in our prototype NIR breast tomography system. Relationships between statistical properties, types of noise and regularization approaches that emerge from the simulation studies are compared to the analogous trends that result from these laboratory experiments.

2.2 Mean-Squared Error Definition

Evaluating the mean-squared error (MSE) in an entire image and appropriately

interpreting its meaning has not typically been considered in traditional medical imaging. The need for an error analysis is heightened when the image variable is expected to be interpreted quantitatively. It is also more important, and yet less well established, in nonlinear image reconstruction where the relationship between bias and variance can change depending upon the parameters used during the inversion process. In this regard, NIR imaging represents a case study where the MSE can be exploited to assess image accuracy as a vehicle for optimizing control of the image reconstruction process.

When the true image is known, such as in simulation or phantom studies, then spatial distributions or images of the bias and standard deviation, σ in the imaged parameter can be calculated for each pixel, j , according to:

$$\text{bias}_j = \frac{\sum_i^R [\mu_j - \mu_j^0]}{R} \quad (2-1)$$

$$\sigma_j = \sqrt{\frac{\sum_i^R [\mu_j - \bar{\mu}_j]^2}{R}} \quad (2-2)$$

where μ_j is the recovered interaction coefficient at pixel j , μ_j^0 is the true value of the interaction coefficient, and $\bar{\mu}_j$ is the average estimated value from a total of R reconstructed images. The mean square error is defined as $\text{MSE} = \text{bias}^2 + \sigma^2$, so it is important to evaluate both these terms to deduce which component dominates the image and to determine if the balance between these quantities changes with the parameters used in the reconstruction.

In this study we normalized the mean square error (MSE) contributions of bias and

standard deviation to be a percentage of the actual interaction coefficients. In addition to reporting the variance and bias as distributed over the entire image, we also computed average values as a way of characterizing and comparing reconstructions obtained with different parameters or successive image estimates recovered during a series of iterations. Summation over an entire image or a region of interest defines scalar values for bias and standard deviation which are averaged over the relevant group of pixels:

$$|\text{bias}| = \frac{\sum_j^N \sum_i^R |\mu_j - \mu_j^o|}{R N} \quad (2-3)$$

$$\sigma = \sum_j^N \sqrt{\frac{\sum_i^R [\mu_j - \mu_j^o]^2}{R N^2}} \quad (2-4)$$

2.3 MSE Statistical Analysis of Computer Simulated NIR Tomographic Images

2.3.1 NIR Tomography Simulation Setup

In order to evaluate the general performance of the image reconstruction algorithm, as briefly introduced in chapter 1, a regular geometry was chosen for simulation studies with an 86 mm diameter field having absorption coefficient $\mu_a = 0.01 \text{ mm}^{-1}$, and reduced scattering coefficient, $\mu_s' = 1.0 \text{ mm}^{-1}$. Within this field, a 26 mm diameter heterogeneity was located just off-center as depicted in Figure 2.1, with interaction coefficients $\mu_a = 0.02 \text{ mm}^{-1}$ and $\mu_s' = 1.0 \text{ mm}^{-1}$ to provide a 2:1 contrast in absorption but no scattering contrast. Forward calculations from diffusion theory were used as the simulated data.

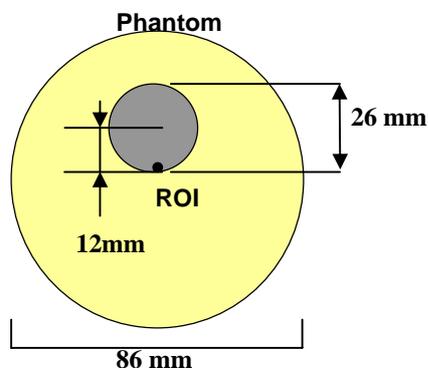


Figure 2.1 Illustration of the geometry used for the simulation studies of MSE analysis, consisting of a circular object of 86-mm-diameter having a 26-mm-diameter inclusion just above the center at a 2:1 contrast in the absorption coefficient.

Noise was included in the data with amplitude dependent upon the test, but in general the average percentage noise added to the amplitude was linked to the average degrees of noise

added to the phase signal. For example, standard measurements with the experimental system exhibit a random variation in amplitude of 1% and a corresponding random variation of 0.5 degrees in the phase shift. These levels were taken to be characteristic of detector behavior and were used as representative noise levels in several of the simulation studies completed here.

2.3.2 Images of Interaction Coefficient and Error

Image reconstruction with the FR approach of reducing the value of the regularization parameter at each step in the iteration process has been used. Recall in Chapter 1, we introduced the matrix solution derived from radiative transport equation for the image reconstruction problem. It has the form

$$\mu^{k+1} = \mu^k + [\mathbf{J}(\mu^k)^T \mathbf{J}(\mu^k) + \lambda \mathbf{I}]^{-1} [\mathbf{J}(\mu^k)]^T (\Phi_M - \Phi_C(\mu^k)) \quad (2-5)$$

and the objective function to be minimized the Newton-Raphson approach is:

$$\chi_k^2 = \frac{\|\Phi_M - \Phi_C(\mu^k)\|^2}{M\sigma^2} \quad (2-6)$$

At each iteration, the Hessian matrix ($\mathbf{J}^T \mathbf{J}$ in Equation (2-5)) is normalized such that the diagonal values become unity. The regularization parameter begins at a value of $\lambda = 100$, and decreases by a factor of $10^{0.5}$ in each iteration, such that after 5 steps it is equal to 1.0. Updates to the absorption and reduced scattering coefficients are estimated at each iteration, and the images of these quantities progressively improve over the first few cycles as long as the objective function (Equation (2-6)) is reduced. As the number of iterations progresses

and the regularization parameter decreases, random variations in the background optical properties can be observed on the order of the magnitude of the properties of the test object. Representative images from this process are shown in Figure 2.2 for both absorption coefficient (top row), and reduced scattering coefficient (bottom row), after 1, 2, 5 and 8 iterations, respectively.

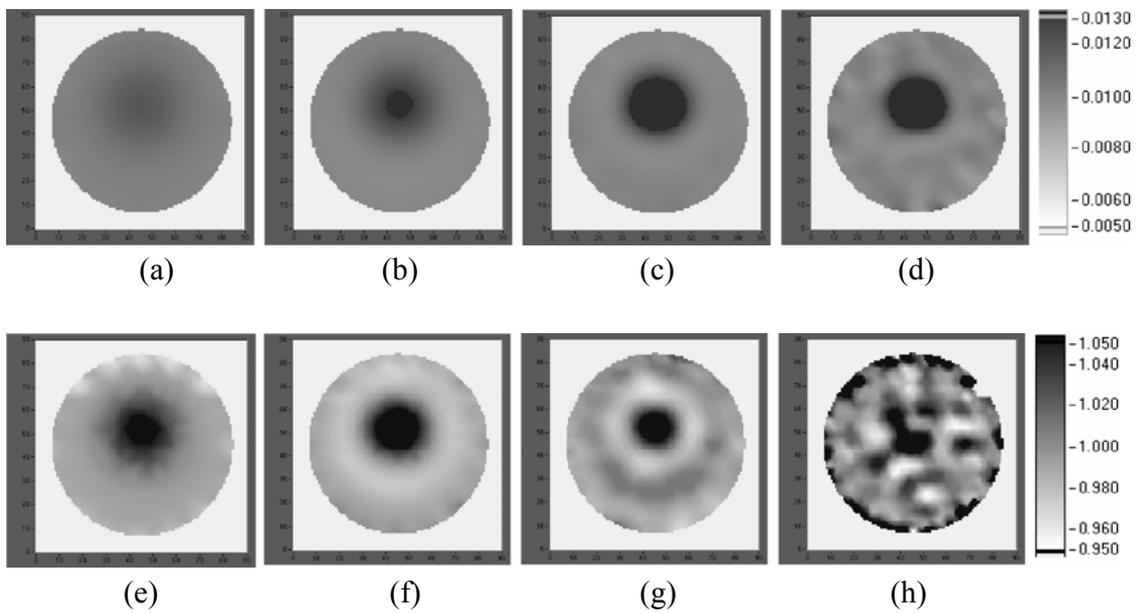


Figure 2.2 Reconstructed images of absorption coefficient (a - d) and reduced scattering coefficient (e - h) are shown after different numbers of iterations in the reconstruction process. The absorption and reduced scattering coefficient images were created in pairs with (a) and (e) resulting after 1 iteration, (b) and (f) after 2 iterations, (c) and (g) after 5 iterations, and (d) and (h) after 8 iterations of the algorithm. The units for the gray scale bars are mm^{-1} .

Using the same simulated data as the synthetic measurements in 100 repetitions of the reconstruction program, each with a different realization of 10% added noise. The average images in Figure 2.2 were formed by stopping the algorithm at iteration number 1, 2, 5 and 8. This high level of noise was added to visualize the spatial heterogeneity in the background

properties at higher numbers of iterations. The four sets of 100 images were then used to calculate images of bias and standard deviation at each of the four iteration endpoints, based upon Equations (2-1) and (2-2). These statistical images are shown in Figure 2.3 for the absorption coefficient as a function of iteration number on the same order as in Figure 2.2. The corresponding bias and standard deviations images for the reduced scattering coefficient are shown in Figure 2.4. Note that the reduced scattering coefficient was homogeneous in the test object, so that any spatially heterogeneous changes in this coefficient are due to bias or variance within the image.

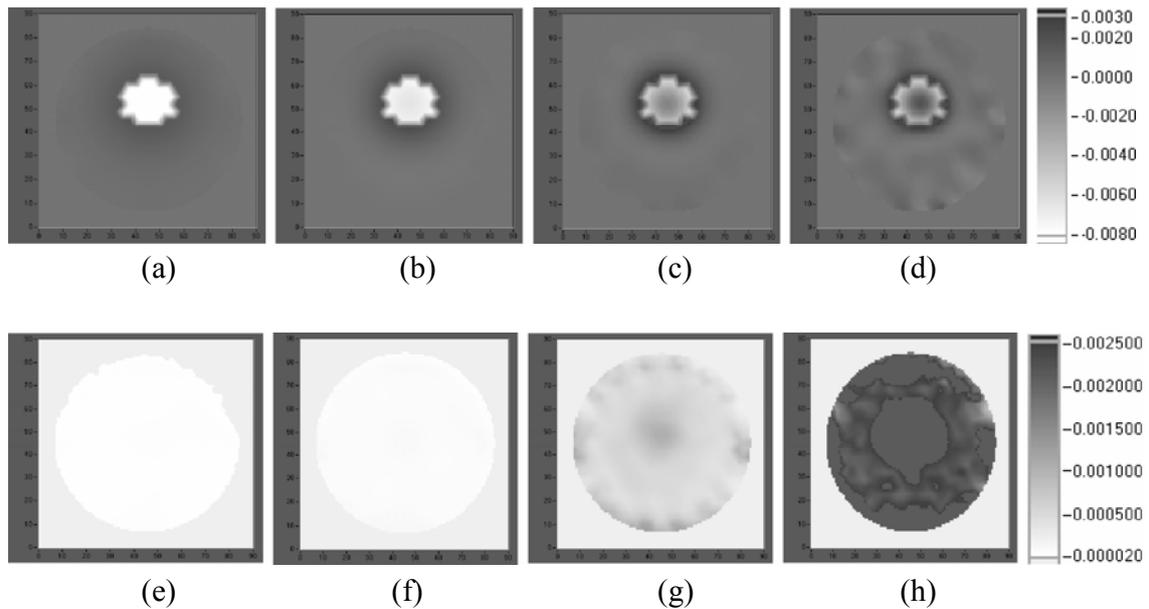


Figure 2.3 Calculated images of absorption coefficient bias error (a-d) and standard deviation error (e-h) from 100 repetitions of the reconstructions shown in Figure 2.2 with random noise added to the simulated measurement data. The units on the gray scale bars are in mm^{-1} .

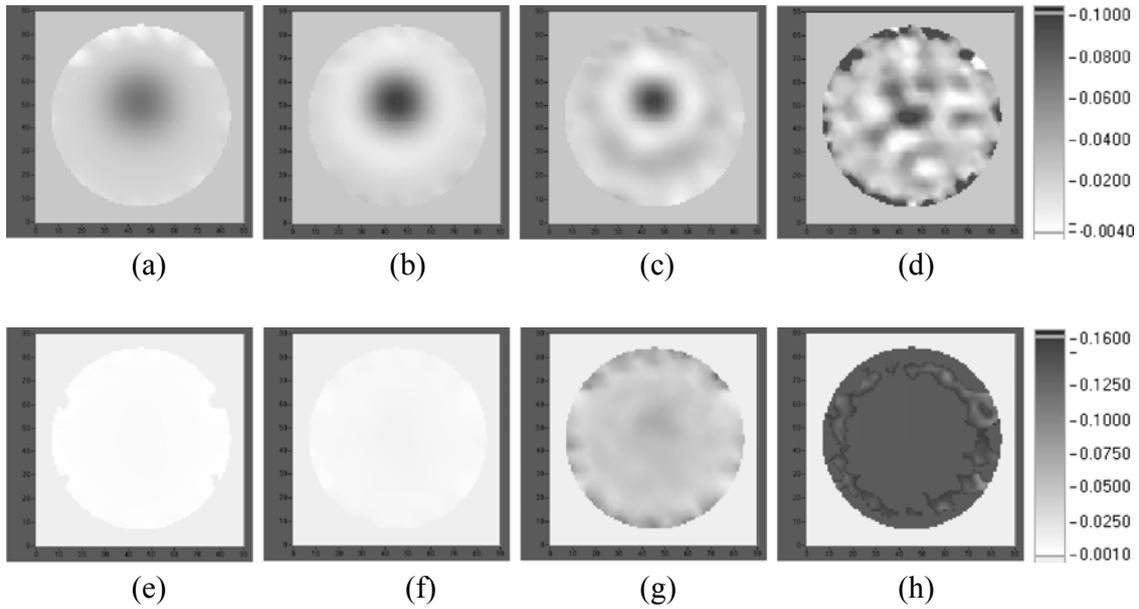


Figure 2.4. Calculated images of reduced scattering coefficient bias error (a-d) and standard deviation error (e-h) from 100 repetitions of the reconstructions shown in Figure 2.2 with random noise added to the simulated measurement data. The units on the gray scale bars are in mm^{-1} .

2.3.3 Fixed versus Levenberg-Marquardt Regularization

In order to quantitatively compare regularization options more easily, the standard deviation and absolute bias were averaged over all nodes within the image field using Equations (2-3) and (2-4). With a fixed regularization parameter of $\lambda = 100$, the same image field was reconstructed 100 times with a randomly varying noise level using standard deviation of 1% in amplitude and 1 degree in phase shift. The average standard deviation and absolute bias in nodal property values were calculated at each iteration from 1 to 10 using the ensemble of 100 images at each step. The same process was repeated with a fixed regularization parameters of $\lambda = 10$ and $\lambda = 1$, respectively. The results of these simulations

are plotted in Figure 2.5 with the top two graphs (for absorption and reduced scattering) reporting values for the entire image. The bottom two graphs show similar results where the averaging of standard deviation and bias was constrained to nodes within the region of interest (ROI) of the heterogeneity.

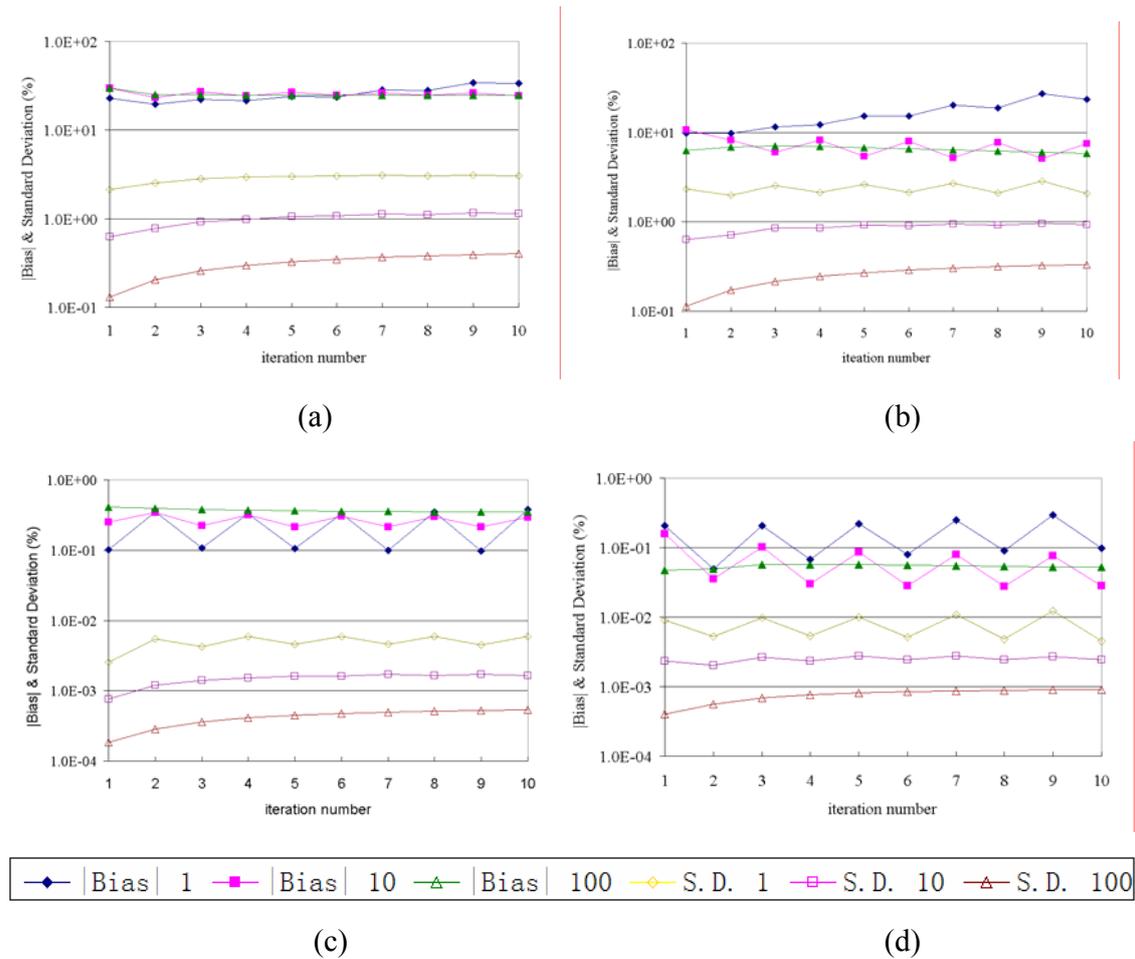


Figure 2.5 Absolute bias and standard deviation for the absorption coefficient of the entire image (a) and the scattering coefficient for the entire image (b). The same analysis for the region of interest (ROI) for the absorption coefficient (c) and the reduced scattering coefficient (d) are also shown. Each data set was completed with fixed levels of regularization parameter, $\lambda = 100, 10$ and 1 , as denoted in the legend. The errors are plotted as a percentage of the interaction coefficients in the background object.

The same series of calculations were performed for reconstruction with the Levenberg-Marquardt regularization approach where the regularization parameter was initiated at $\lambda = 100$ and decreased by a factor of $10^{0.5}$ during each successive iteration. In Figure 2.6, the absorption and reduced scattering coefficients error for the entire image and the ROI are presented.

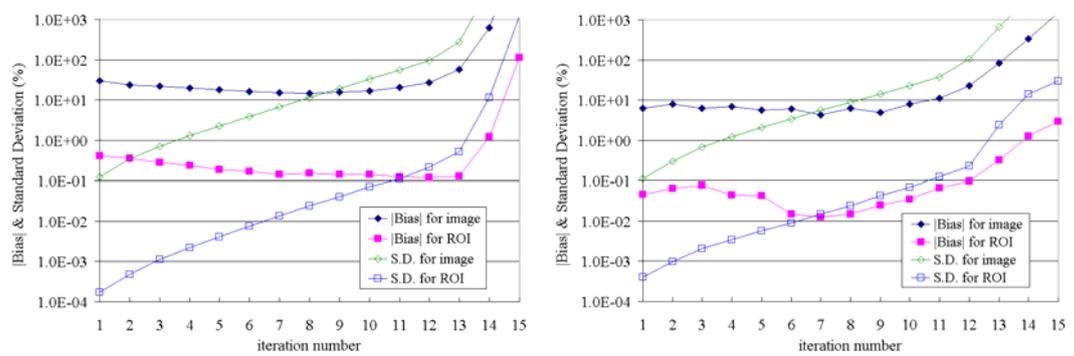


Figure 2.6 Absolute bias and standard deviation for the absorption coefficient (a) and reduced scattering coefficient (b) using the Levenberg-Marquardt scheme where the regularization parameter was initiated at 100 and decreased by a factor of $10^{0.5}$ after each iteration. The values for both the entire image and the region of interest (ROI) are shown in each graph as a percentage of the background interaction coefficient values within the object

An important measure of error in the reconstruction process is the squared projection error as calculated in Equation (1-4). The tabulated values for the projection error are plotted in Figure 2.7. The line labeled 0.5 corresponds to the reconstruction procedure where the λ value is decreased by a factor of $10^{0.5}$ at each iteration step, which is the condition used in the reconstruction data shown in Figure 2.6. In this line, the minimum projection error is obtained after the 5th iteration. The same test object was also reconstructed using reduction factors of $10^{0.25}$ and $10^{0.125}$ between iterations, and this data is

plotted in the same figure with labels 0.25 and 0.125, respectively. These results illustrate that as the step decreases; the minimum obtainable projection error is also decreased suggesting that a more optimal image is obtained for this objective function.

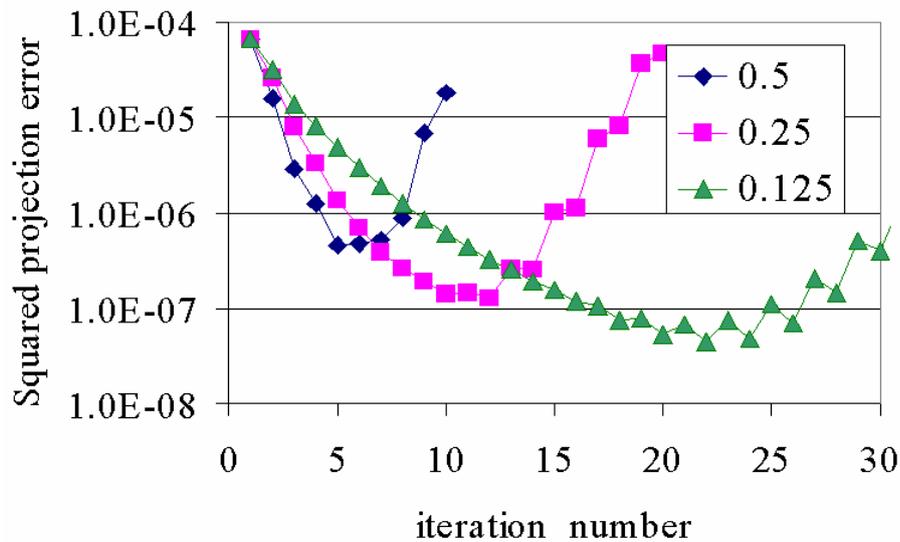


Figure 2.7 Projection error (or objective function values, from equation (2-6)) as a function of iteration number using the Levenberg-Marquardt scheme where the regularization parameter is decreased by a constant factor between iterations. The reduction factor had the values of $10^{0.5}$, $10^{0.25}$ and $10^{0.125}$, as denoted by the legend.

2.3.4 Image Quality

Reconstructed images of NIR interaction coefficients are readily obtained with the algorithm analyzed here, and the analysis examined here is an important tool for measuring the quality of the image recovery algorithms being developed. The images shown in Figure 2.2 are typical and illustrate many of the features of image reconstruction from diffuse projection measurements. Most notably, it is well known that the spatial resolution of NIR

tomographic images is limited and also spatially dependent [70]. The major consequence of having modest spatial resolution is that finely detailed objects appeared blurred. For example, the circular test heterogeneity with sharply defined edges (top-hat profile) that was used here appears as a circular blurred heterogeneity with a Gaussian-shaped profile. This change in shape will ultimately limit the accuracy of the reconstruction because there is an inherent bias in the image due to the loss of resolution. This loss of resolution can be minimized by inclusion of spatially varying regularization or through definition of alternative objective functions [25, 60, 66, 70]. There has been considerable study of how best to choose the regularization parameter constraints, however it is not well defined how this change in objective function will alter the resulting image variance under the conditions used here. The calculations shown in this paper present a first step towards this type of error analysis, testing first the impact of Levenberg-Marquardt style regularization upon the image error. The data presented in Figures 2.5 through 2.7 indicates that the Levenberg-Marquardt method indeed does minimize the mean square image error as well as the objective function. However there are features of the minimization algorithms which can be observed from the bias and standard deviation which are not apparent in the objective function values.

As the regularization parameter is reduced and the number of iterations progresses, the contrast and resolution of the image heterogeneity is increased as shown in Figure 2.2. Typically, as the contrast of the heterogeneity in the recovered image approaches the true value, the imaging objective function approaches a minimum. Since the required matrix inversion is ill-posed the solution becomes unstable as the regularization parameter is

reduced, which is manifested in the image as increased heterogeneity within the image field. It is not clear to what extent these spatial heterogeneities are correlated to the data noise, versus more systematic errors. When the noise in the measurements is random, then successive reconstructions from different data will produce images with different spatial noise. This random variance in the reconstructed image is what has been quantified here by the standard deviation. However in regions where the bias is significantly larger than the standard deviation (Figure 2.8(a)), it must be assumed that random variation is not highly correlated to image error. However after several iterations when the standard deviation significantly exceeds the bias error (Figure 2.8(d)), then there is correlation between the image heterogeneity and the data random noise.

A well known artifact of NIR imaging is the cross-talk between absorption and scattering coefficients [52]. While the measured phase and amplitude provide a unique set of data sufficient to accurately reconstruct the true absorption and scattering coefficients, it is well known that there can be approximately 1% cross talk from an absorbing object into the scattering image, or up to 30% cross talk from a scattering object into the absorption image. This artifact will increase the bias of the resulting image pair, and somewhat complicates the analysis used here.

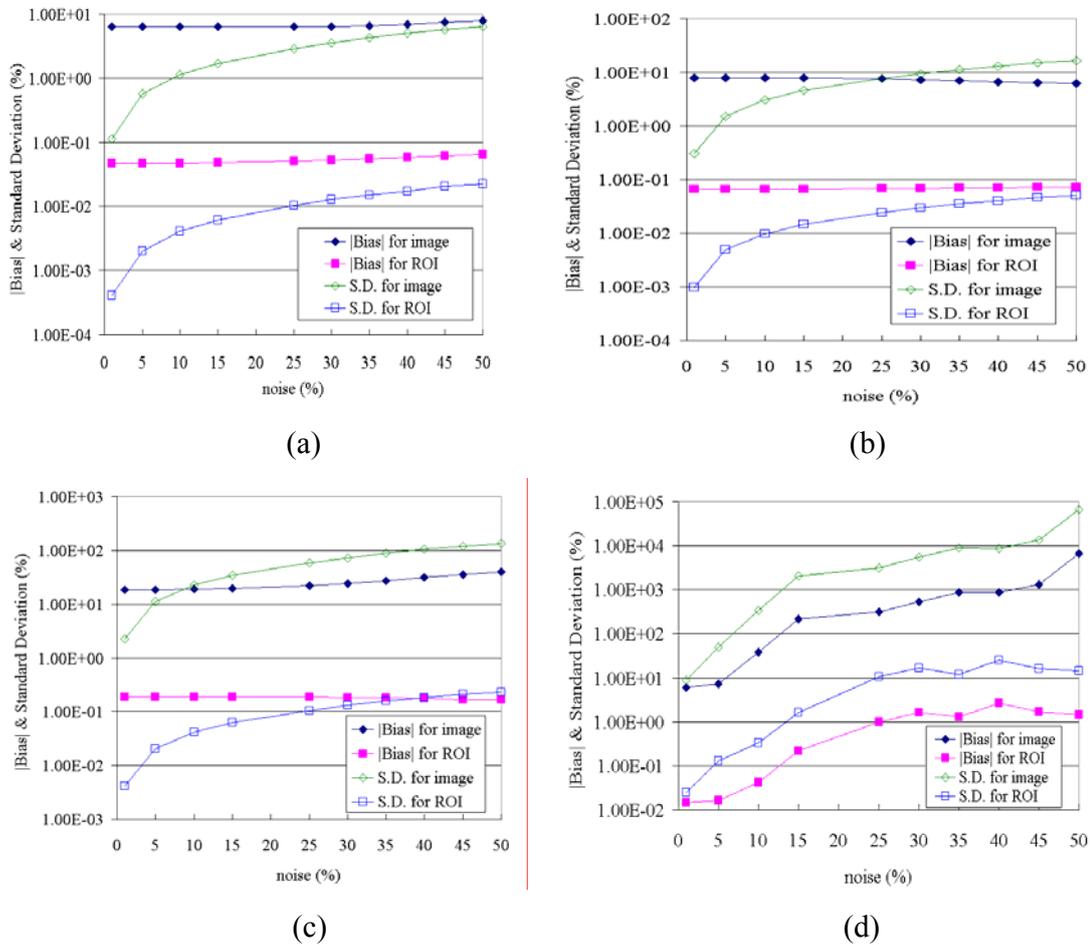


Figure 2.8. Absolute bias and standard deviation for the absorption coefficient (a) in the entire image and the ROI after 1 iteration. The same calculations are also shown after 2 iterations (b), 5 iterations (c) and 8 iterations (d). Each calculation was repeated with a range of noise levels between 1% and 50%.

In these studies we have separated the absorption and reduced scattering coefficient error analyses in order to individually observe the effect. As noted in the graphs of Figure 2.5, the bias and standard deviation errors are typically much lower in scattering coefficient than in absorption coefficient for the test object we studied. This is likely a result of choosing an absorbing object for the test field. However, this test object is representative of

many physiologically relevant tissue contrasts as well, so there is inherent practical value in this choice. Still, by separating the absorption and scattering contrast errors we are able to independently examine the reconstruction algorithm performance under the conditions used here for analyzing the relationship between noise and regularization method.

Analysis of the bias and standard deviation plotted in Figure 2.5 indicates that for the case of a fixed regularization parameter, the bias is almost always higher than the standard deviation. In general, the bias tends to decrease with successive iterations while the standard deviation increases. However, there are notable exceptions to this trend which occur at the lower regularization parameter values.

2.3.5 Regularization Method

When using the Levenberg-Marquardt approach to vary the regularization parameter there is a more efficient minimization of the bias error and the resulting image projection error is also minimized. The challenge in using this method is to optimally define the lower limit of the λ parameter such that the total error does not increase due to the matrix being ill-conditioned. Since the mean square error (MSE) of the estimate is the summation of the bias squared and the variance, the minimum MSE will coincide with the minimum of the standard deviation and bias errors. From the data in Figure 2.6, this will occur for the entire image near the 8th and 9th iterations, corresponding to λ near 0.01, for the case of 1% noise. However, if we are most concerned with minimizing the error in the ROI of the

heterogeneity, then the minimum occurs near the 11th iteration in absorption coefficient where λ is closer to 0.00033. There is a local minimum in the scattering coefficient error of the ROI near the 7th iteration; however, since the overall error in scattering is lower than absorption, the total error will be still be minimal near the 11th iteration.

Interestingly, the projection error plotted in Figure 2.7 for the same method of regularization (top line in this graph) shows a minimum between the 5th and 6th iterations, corresponding approximately to $\lambda = 0.1$. This observation suggests that the minimum projection error solution does not necessarily correspond to the minimum image error.

The effect of decreasing the reduction factor of λ between iterations from $10^{0.5}$, to $10^{0.25}$ or $10^{0.125}$, (as shown in Figure 2.7) decreases the overall projection error and the overall mean square error. This observation indicates that as the reduction in λ decreases, the projection error decreases, and the overall minimum in this objective function is lowered. This is an important observation since it indicates that the standard Levenberg-Marquardt method of using a factor of 10 between iterations is not the best way to minimize this problem. It is also an unfortunate observation because it suggests that more iterations at the lowest reduction factor is always the optimal way to minimize projection errors, thereby significantly increasing the computational time needed for the solution.

2.3.6 Noise Analysis

While the previous set of reconstructions were carried out with a fixed noise level, it is

also important to assess the influence of noise level on the regularization parameter. The Levenberg-Marquardt method with a reduction factor of $10^{0.5}$ was found to provide the minimum bias and standard deviation errors as well as the minimum projection error, so it was used exclusively in this analysis. The same image field was reconstructed from a fixed set of iterations with increasing levels of noise between 1 % and 50%. The graphs in Figure 2.8 show the calculated bias and standard deviation in the absorption coefficient as a function of noise for both the entire image and the ROI. The graph reports these values after 1, 2, 5 and 8 iterations.

The graphs of image error versus noise level shown in Figure 2.8 illustrate several important features about the trade-off between image bias and image standard deviation. In general, when the regularization is high (at low numbers of iterations), the image is dominated by bias error, and as the reconstruction progresses, the variance becomes the dominant factor. Overall the error in the entire image is typically more than two-orders of magnitude higher than in the ROI. As might be anticipated, the point at which the standard deviation and bias form a combined minima is the optimal image in terms of the mean image error. Interestingly, this optimal image occurs at a higher number of iterations for the ROI than for the entire image, suggesting that if quantification of a localized region is the most important consideration, than a larger number of iterations is required.

The noise in the measurement data determines the regularization level that is required for minimal image error. There is an inverse relationship between the noise level and the number of iterations needed to reach the minimal image error. For example, with 50% noise the

optimal image is achieved after approximate one iteration, whereas with 25% noise it is achieved with 2 iterations, with 8% noise it is achieved after 5 iterations, and with 1% noise it is achieved after 8 iterations. These iterations correspond to final regularization parameters of 100, 33.3, 1.0 and 0.0333, respectively. Thus, the noise level in the measurement data must be quantified in order to appropriately choose the number of iterations to be used in the algorithm and the lower limit on regularization parameter.

2.4 MSE Statistical Analysis of Phantom Study Images

In NIR, the presence of tissue–air interfaces and other boundaries plays a crucial role in the imaging problem. Because the light transport process is analogous to diffusion, the light signal significantly samples the tissue lateral to the line-of sight path between source and detector. Thus, small changes in fiber position can have a major effect on the resulting measured signal. Boas et. al, estimated that fiber positions need to be known to within 100 μ m in order to maintain a position-dependent variation of less than 2% [9]. In our current system, the positioning precision of an individual fiber is near this order of magnitude, although the inter-fiber repeatability over several repositioning placements is less accurate. Since these errors cannot be eliminated in the current realization, we have characterized system performance both with the fibers positioned once on a single phantom during repeated measurements, as well as with the fibers positioned on the same phantom

in-between repeated set of measurements. This comparison provides quantitative information on how the fiber positioning affects the images obtained.

Reconstruction algorithms are developed to accurately create images of the NIR absorption and reduced scattering of the region of interest (ROI) and previous studies have indicated that this can be achieved with modest spatial resolution [49]. However, determining the "accuracy" of the image reconstruction is difficult, and can only be reliably tested on tissue simulating phantoms where the optical interaction coefficients are known accurately. In section 2.2 of this chapter, where simulated data was used, the image reconstruction algorithm minimized the bias initially, while allowing the standard deviation to increase as the number of iterations progressed. The "optimal" image was obtained in a region when the bias and standard deviation were of comparable order of magnitude. In this paper, the same analysis is applied to real images recovered from experimental data collected from calibrated tissue phantoms.

Different methods of regularization have been applied in our optical diffusion model-based image reconstruction algorithm. The regularization parameter is a key factor which can modify the reconstructed images significantly [1, 3, 36, 40, 72]. The technique of developing the best method of regularization is problem-dependent [1, 36, 70]; hence, the type of analysis employed here provides an objective method with which to compare a variety of regularization schemes. Statistical analysis of the reconstructed NIR tomographic images efficiently evaluates the validity of competing regularization methods, and is potentially an alternative approach to understand algorithm convergence.

2.4.1. Phantom Studies Setup

In section 2.2, we introduced the concept of analyzing near-infrared (NIR) tomographic images by examination of bias and standard deviation maps of an ensemble of simulated reconstructions obtained from repeated sampling of synthetic measurement data. In this section, we will discuss the test results of phantom studies. Phantom studies provide a more practical way to evaluate the NIR tomographic image reconstruction method specified in former sections. Gelatin phantoms are used to mimic the physical and optical properties of the human breast. The phantoms demonstrated in Figure 1.5 and Figure 2.9 are gelatin phantoms with optical properties similar to breast tissue. Typical breast tissue is thought to have an absorption coefficient μ_a of 0.002 to 0.02 mm^{-1} and a reduced scattering coefficient μ_s' of 0.5 to 2.0 mm^{-1} [26, 46, 67, 73, 93, 99]. The gelatin phantoms were combined to achieve the required absorption coefficient, such that the absorption coefficient was determined predominantly by water and the concentration of India ink. Titanium dioxide is used in the resin phantom recipe to provide the appropriate reduced scattering coefficient. After being removed from the mold, the cylindrical gelatin phantom had a diameter of 86 mm and a height of 55 mm. A hole of 22 mm in diameter was cut in the phantom and filled with Intralipid and ink to produce absorption coefficient contrast in the phantom with a closely matched scattering coefficient to the background. As illustrated in Figure 2.9, the relative position of the hole and the phantom was well controlled. The only difference between phantom #1 and phantom #2 is the relative position of the location of the inner circle

creating heterogeneity within the region. The tumor-like circular object of 22 mm in diameter in phantom #2 was located closer to the edge of the phantom. These two phantoms were used to help evaluate the sensitivity to the object position in the imaging system.

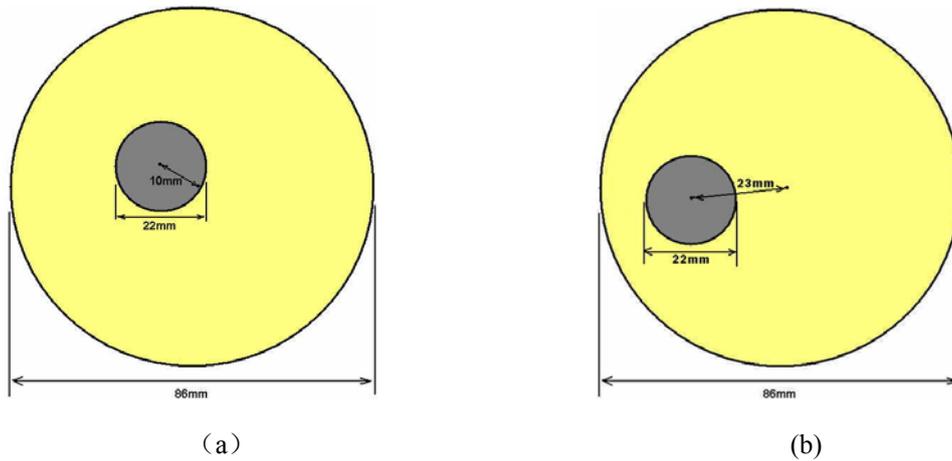


Figure 2.9. The geometries of the phantoms used in the studies for a 22 mm diameter circular object in each 86 mm diameter jelly phantom. (a) The absorption and scattering coefficients of the homogeneous phantom #1 are 0.00525 and 0.941 with the absorption and scattering of the heterogeneous object inside are 0.00930 and 1.02. (b) The absorption and scattering coefficients of the homogeneous phantom #2 are 0.00578 and 1.03 with the absorption and scattering of the heterogeneous object inside are 0.00951 and 0.973.

As described in chapter 1, newly made gelatin phantoms were first used in the imaging system to acquire tomographic data prior to cutting the hole. A homogenous fit to calculated data from diffusion theory was performed in this step to allow accurate prediction of the absorption and reduced scattering coefficient in the gelatin phantom. Once these background values were known, the holes are then created in the desired locations.

The two phantoms were tested each in two different procedures described in chapter 1 trying to minimize the fiber optic interface system error, as illustrated in Figure 1.6

2.4.2. Phantom Studies Results

2.4.2.1. Images of Interaction Coefficient and Errors

Initially, a decreasing regularization was used where the value was continuously dropped by a fixed amount (FR method). The initial regularization value (λ) was 100, and then it was reduced by a factor of 10 after each iteration. The effect of this procedure is to increase the contrast between the ROI and the background at the expense of introducing higher spatial frequencies in the recovered optical property at each iteration. Using the 20 sets of the

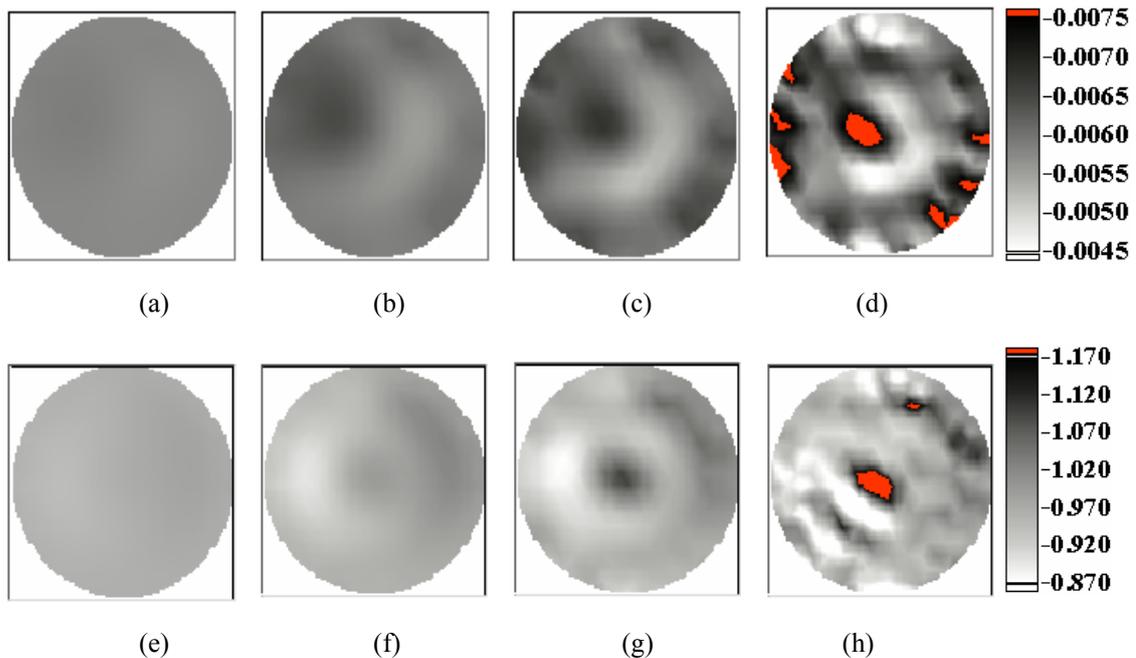


Figure 2.10 Reconstructed images of absorption coefficient (μ_a) and reduced scattering coefficient (μ_s') from 20 measurements by the fixed decreasing regularization method for the phantom #1 after 1 iteration (a, e), 3 iterations (b, f), 5 iterations (c, g), and 8 iterations (d, h).

Measurement data from the imaging instrument, the average images of the absorption (at top

row) and scattering (at bottom row) coefficients for phantom #1 are shown in Figure 2.10 after one, three, five, and eight iterations. The corresponding images for phantom #2 (not shown) were also obtained.

Estimation of the absolute bias and standard deviation errors was calculated for both of the absorption and scattering coefficients using (5) and (6). Each analysis consisted of 20 sets of measured data, producing 20 separate images from the reconstruction program. The mean images of the absolute bias (at top row) and standard deviation (at bottom row) errors are shown in Figure 2.11 for the absorption coefficients of phantom #1. Each column represents the same iteration endpoints presented in Figure 2.10. The equivalent images for the reduced scattering coefficient of the same phantom are shown in Figure 2.12.

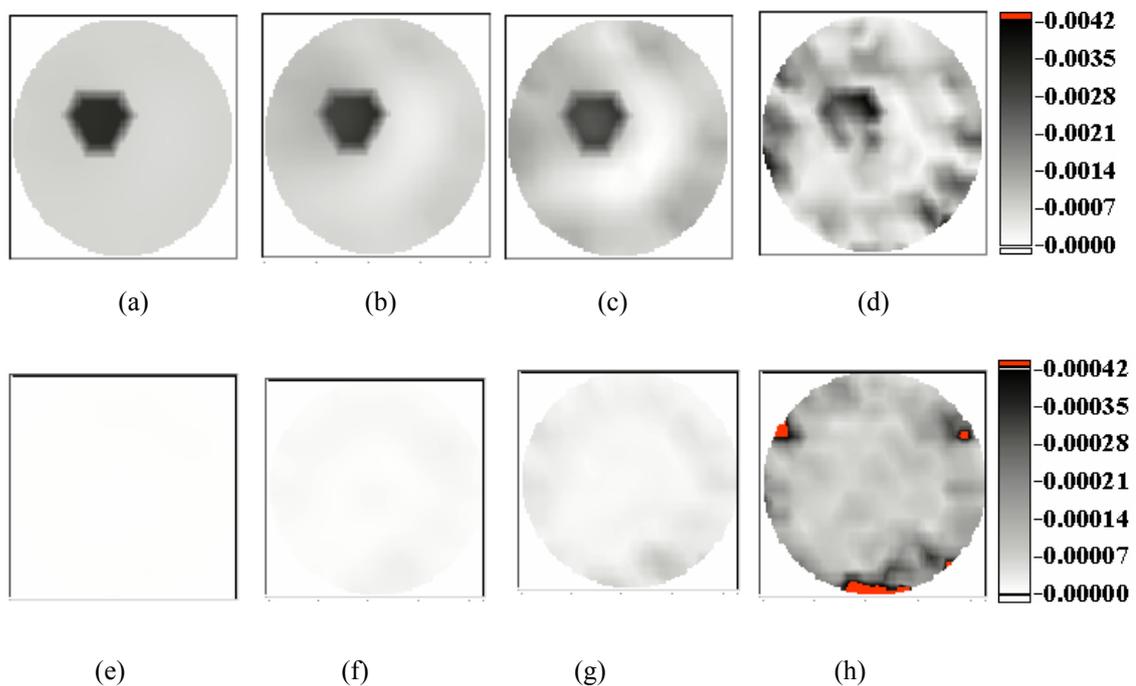


Figure 2.11 Calculated images of bias error (a-d) and standard deviation error (e-h) for absorption coefficient from 20 measurements after 1 iteration (a,e), 3 iterations (b,f), 5 iterations (c,g), and 8 iterations (d,h) by the fixed decreasing regularization method for the phantom #1.

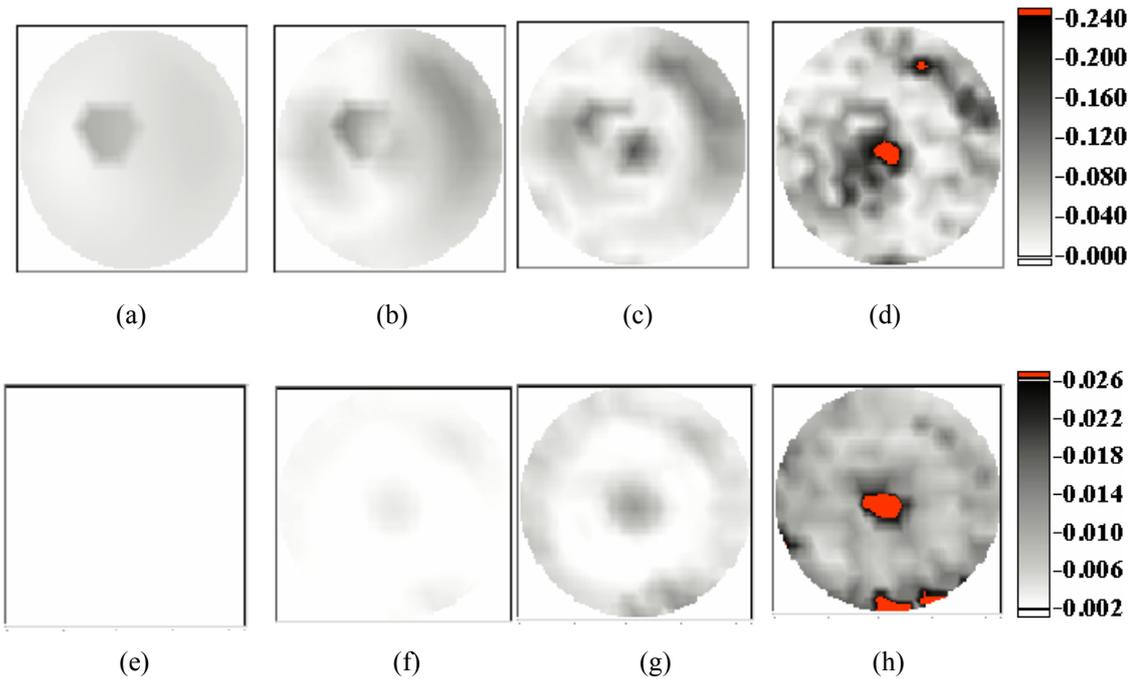


Figure 2.12. Calculated images of bias error (a-d) and standard deviation error (e-h) for reduced scattering coefficient from 20 measurements after 1 iteration (a, e), 3 iterations (b, f), 5 iterations (c, g), and 8 iterations (d, h) by the fixed decreasing regularization method for the phantom #1.

To avoid the reconstructed image accuracy decreasing as the regularization parameter reduces below a certain level, the Levenberg-Marquardt regularization approach is used to control the regularization value based on the projection error, which was described in Equation (2-6). Like the fixed regularization method, the initial regularization value (λ) is 100, and it is reduced a factor of $10^{0.5}$ after the first iteration. Afterwards, the projection errors at the previous step and the current step are compared. If the previous projection error is greater or equal the current one, and then the regularization is abated as the same of the fixed decreasing method; but, if the previous one is less than the current one, the regularization increases by the factor of $10^{0.5}$, so that the regularization will be kept in a

range. The average images of the absorption and scattering coefficients calculated using the Levenberg-Marquardt method are shown in Figure 2.13, which is similar to Figure 2.10 where the fixed decreasing approach was used; also, Figure 2.17 is similar to Figure 2.16 for phantom #2 by these two different approaches. Compare the images in Figures 2.11 and 2.12 with the images in Figures 2.14 and 2.15 for the absorption and reduced scattering coefficients respectively. These represent the absolute bias and standard deviation errors by the Levenberg-Marquardt approach.

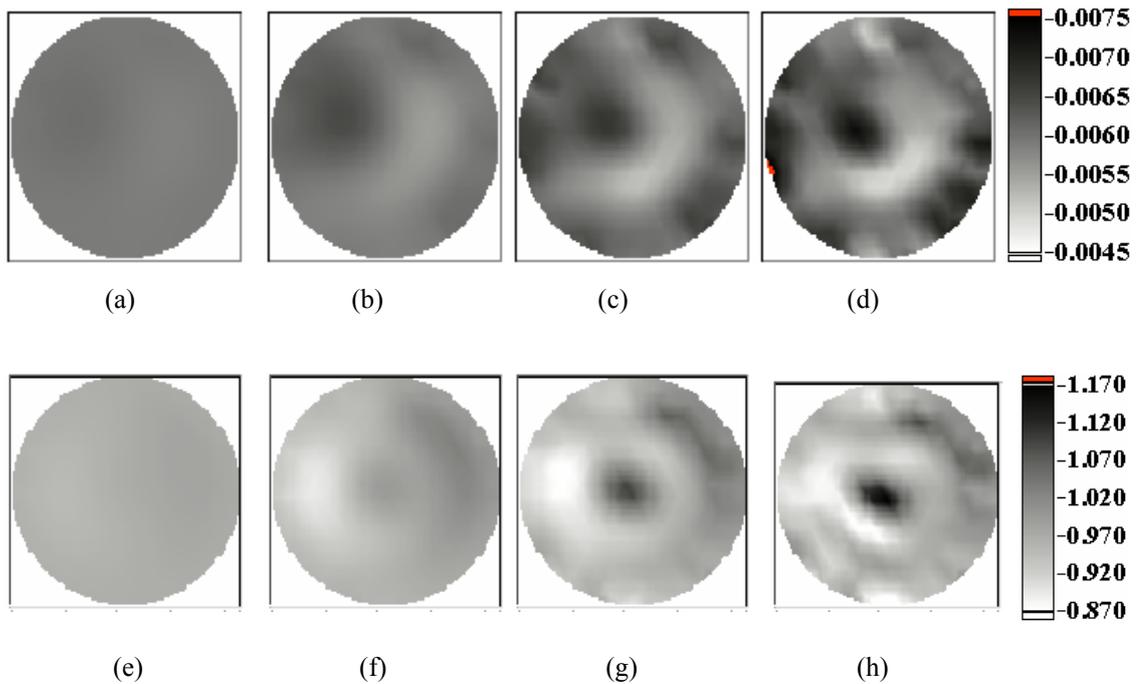


Figure 2.13. Reconstructed images of absorption coefficient (μ_a) and reduced scattering coefficient (μ_s') from 20 measurement by the Levenberg-Marquardt regularization method for the phantom #1 after 1 iteration (a, e), 3 iterations (b,f), 5 iterations (c,g), and 8 iterations (d,h).

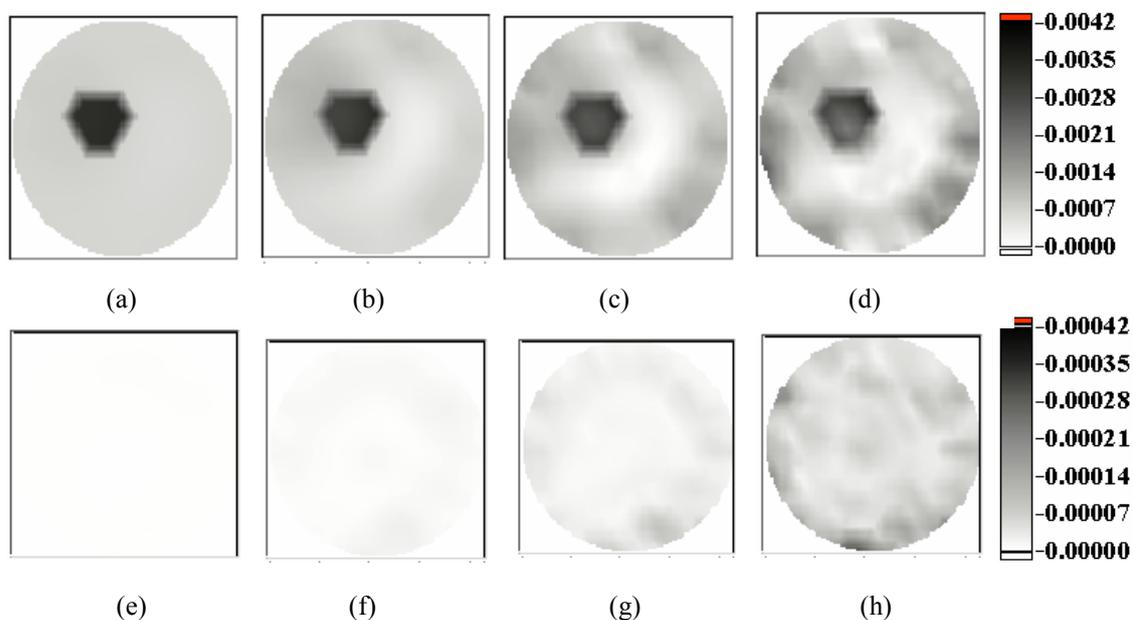


Figure 2.14. Calculated images of bias error (a-d) and standard deviation error (e-h) for absorption coefficient from 20 measurements after 1 iteration (a, e), 3 iterations (b, f), 5 iterations (c, g), and 8 iterations (d, h) by the Levenberg-Marquardt regularization method for the phantom #1.

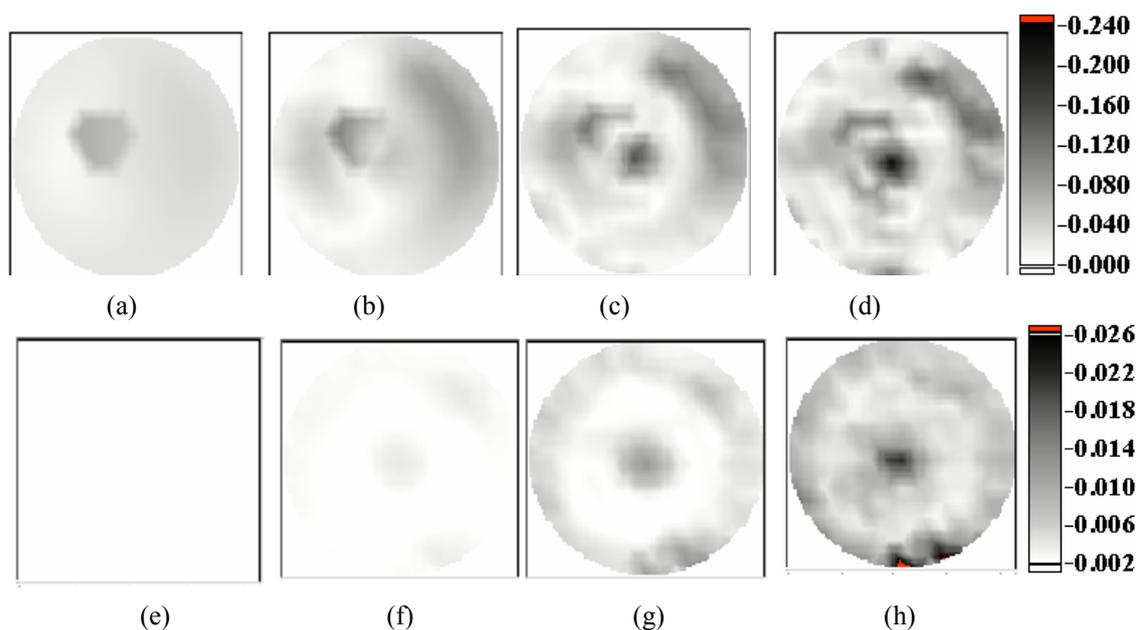


Figure 2.15. Calculated images of bias error (a-d) and standard deviation error (e-h) for reduced scattering coefficient from 20 measurements after 1 iteration (a,e), 3 iterations (b,f), 5 iterations (c,g), and 8 iterations (d,h) by the Levenberg-Marquardt regularization method for the phantom #1.

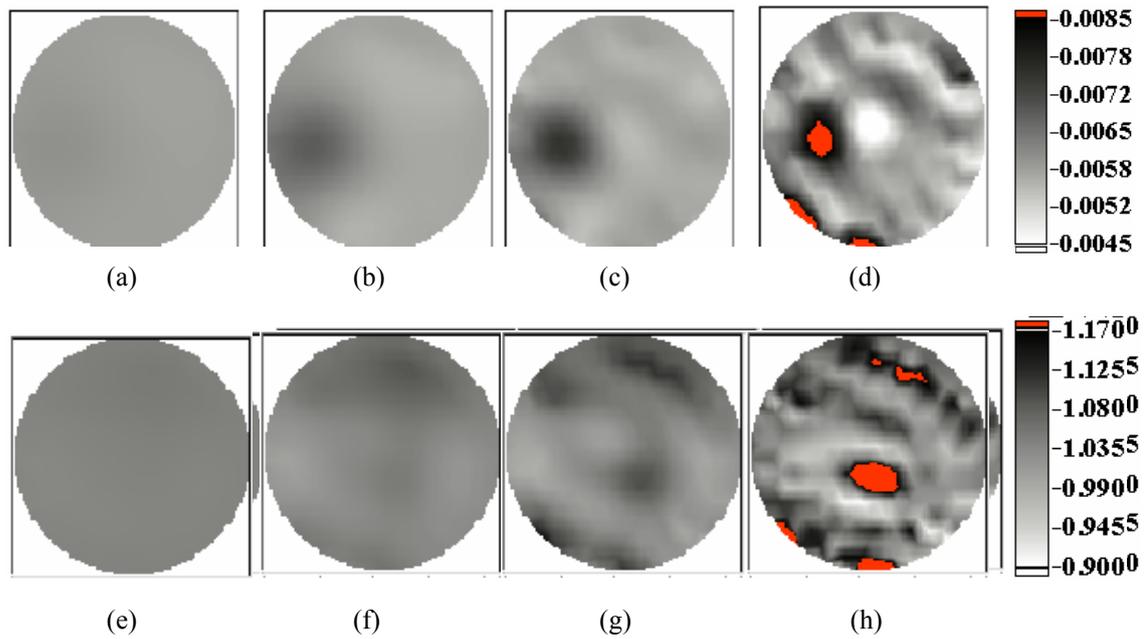


Figure 2.16. Reconstructed images of absorption coefficient (μ_a) and reduced scattering coefficient (μ_s') from 20 measurements by the fixed decreasing regularization method for the phantom #2 after 1 iteration (a, e), 3 iterations (b, f), 5 iterations (c, g), and 8 iterations (d, h).

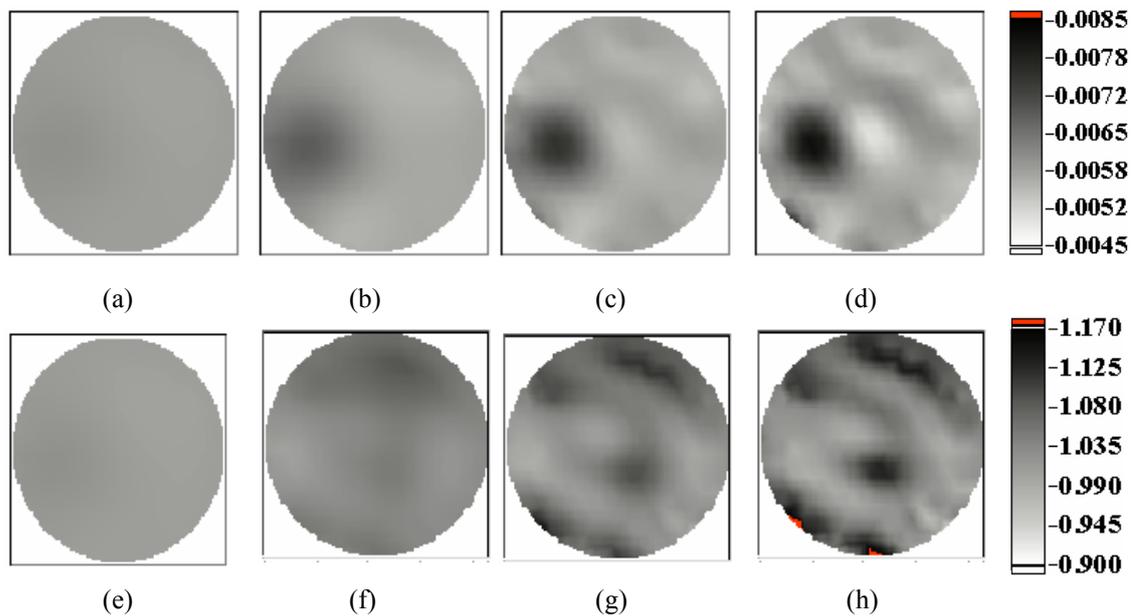


Figure 2.17 Reconstructed images of absorption coefficient (μ_a) and reduced scattering coefficient (μ_s') from 20 measurements by the Levenberg-Marquardt regularization method for the phantom #2 after 1 iteration (a, e), 3 iterations (b, f), 5 iterations (c, g), and 8 iterations (d, h).

In contrast to the standard practice of minimizing the objective function, it is also possible to use the images mean squared error that should be minimized at the same time. In the phantom studies used here, the contrast of the image heterogeneity and the accuracy of the data acquisition system are the most important factors that determine the quality of the images reconstructed from the algorithm analyzed here. The low spatial resolution of near-infrared tomographic images limits the quality of the reconstructed images.

As we can see from Figures 2.10, 2.13, 2.16 and 2.17, the circular heterogeneities with sharply defined edges, as shown in Figure 2.9, appear as blurred objects with Gaussian-shaped profiles. At the same time, the images can be dominated by fluctuations in the properties near the surface. It is often these highly sensitive regions near the sources or detectors which dominate the overall image quality, and the level at which the matrix can be regularized. A striking observation from these calculations, is that the image error (as defined by bias and standard deviation together) never decreases in our situation, as can be seen in Figures 2.18 through 2.21. This fact is likely due to errors in the periphery of the image dominating the overall image solution. Thus, while we would like to obtain the most accurate image, as measured by mean squared error; this is not achieved by our current reconstruction algorithm. However, at the same time, the random variation between images is negligible, indicating that the repeatability of the system is quite high. This latter observation can be seen in all the graphs in Figures 2.18 through 18, where the standard deviation is never significant in comparison to the bias error. Thus, variation between images should not be an important aspect of the imaging system performance, other than

those changes which are due to calibration changes.

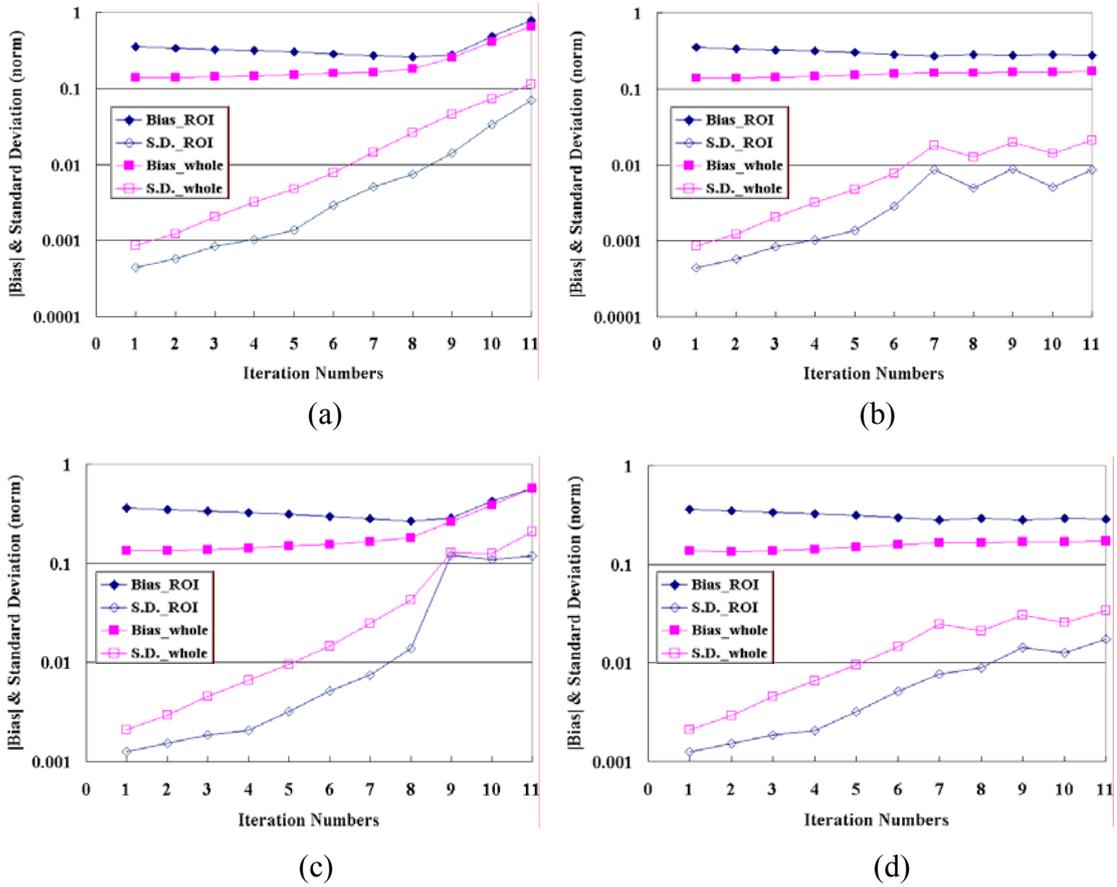


Figure 2.18. Calculations of the absolute bias and standard deviation for absorption coefficient in the whole image and the region of interest (ROI) of the phantom #1 by the fixed decreasing regularization method (a,c) and Levenberg-Marquardt regularization method (b,d). The fiber optic interface was fixed during the data collecting process for the graphs at the top row (a,b); and, the imaging system was readjusted after each data acquisition to make the sources and detectors touch well with the boundary of the phantom for the graphs at the bottom row (c,d).

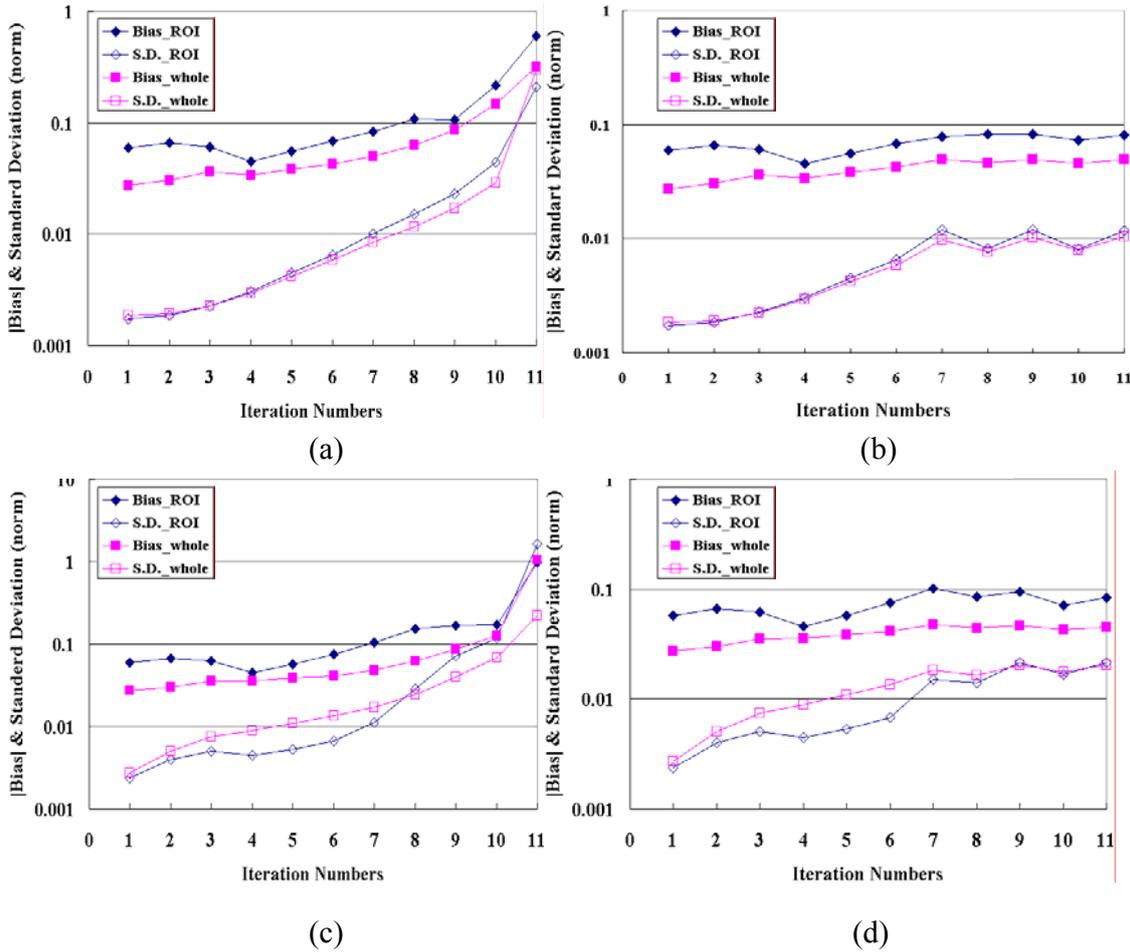


Figure 2.19. Calculations of the absolute bias and standard deviation for reduced scattering coefficient in the whole image and the region of interest (ROI) of the phantom #1 by the fixed decreasing (a,c) and Levenberg-Marquardt (b,d) approaches. The fiber optic interface was fixed during the data collecting process for the graphs a and b; and, the imaging system was readjusted after each data for the graphs c and d.

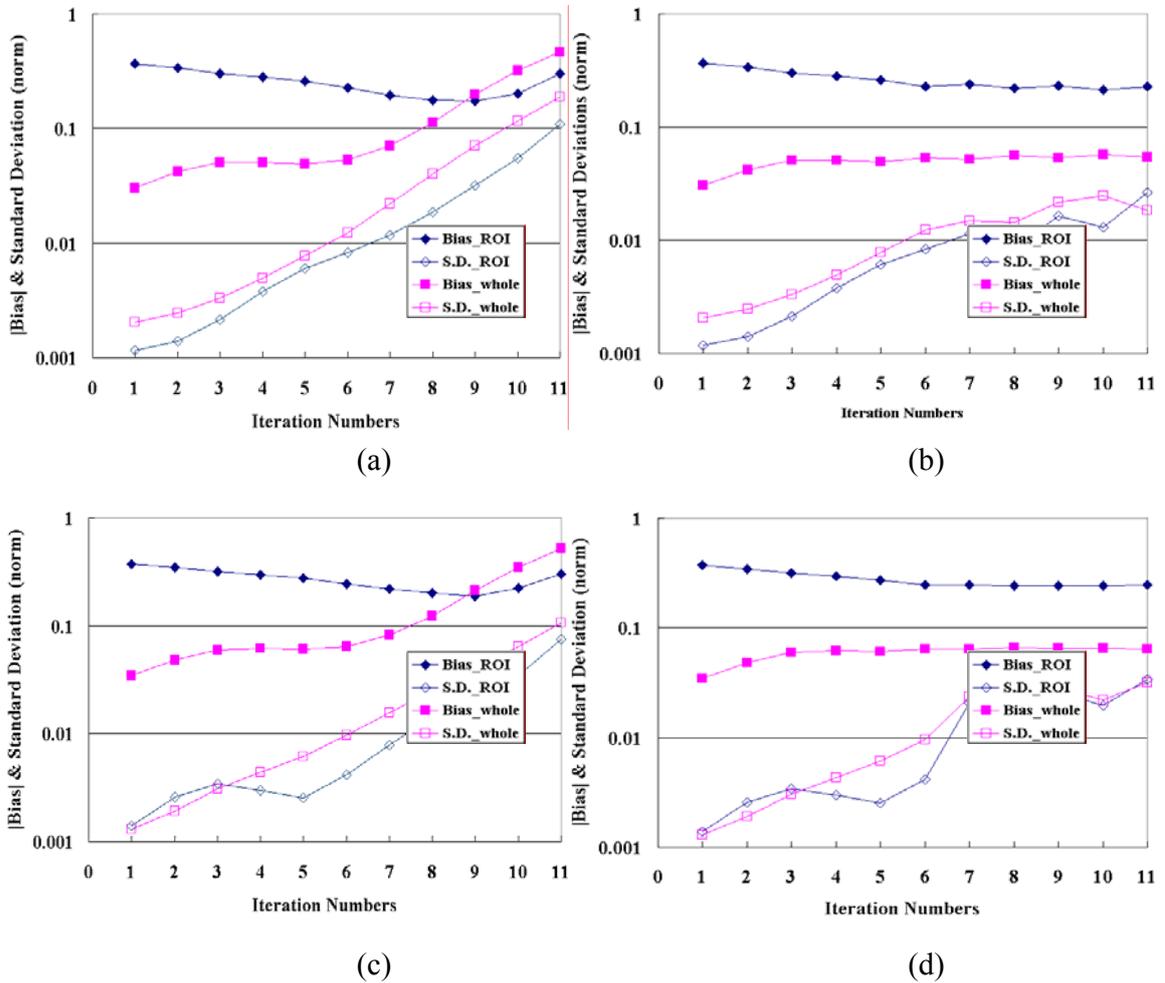


Figure 2.20. Calculations of the absolute bias and standard deviation for absorption coefficient in the whole image and the region of interest (ROI) of the phantom #2 by the fixed decreasing (a,c) and Levenberg-Marquardt (b,d) approaches. The fiber optic interface was fixed during the data collecting process for the graphs a and b; and, the imaging system was readjusted after each data for the graphs c and d.

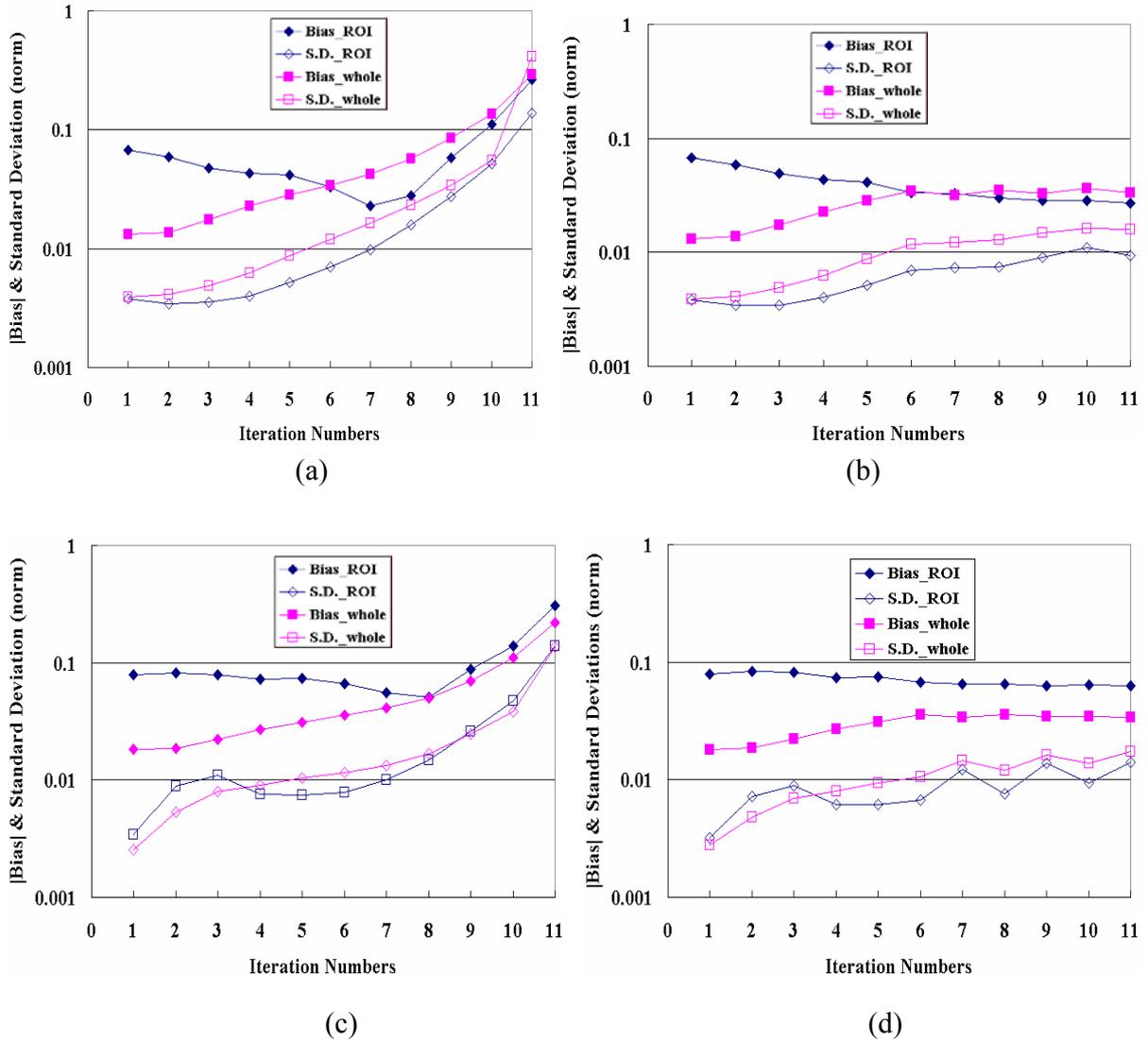


Figure 2.21. Calculations of the absolute bias and standard deviation for reduced scattering coefficient in the whole image and the region of interest (ROI) of the phantom #2 by the fixed decreasing (a,c) and Levenberg-Marquardt (b,d) approaches. The fiber optic interface was fixed during the data collecting process for the graphs a and b; and, the imaging system was readjusted after each data for the graphs c and d.

Interestingly, the increasing error observed in Figures 2.18 through 2.21 is opposite to the situation observed in the simulation part, where the bias error could be minimized somewhat, when theoretical data was used. These theoretical simulations contained random noise in a Gaussian distribution, mimicking that of the spontaneous detector fluctuations. Thus, we must conclude that the limitation in our current detection system is not random noise, but rather systematic errors, likely induced by the calibration procedure. This systematic error level likely limits our ability to produce more accurate images, and ultimately limits the ability to resolve the heterogeneities with high resolution

2.4.2.2. Projection Error

The squared projection error is a very important estimate in the reconstruction course to acquire the optimal images by minimizing it. Moreover, it is a flag to control the value of the regularization parameter (λ) in the Levenberg-Marquardt regularization approach. The graphs in Figure 2.22 show the average projection errors of 20 reconstruction data sets, with both the fixed decreasing and Levenberg-Marquardt regularization approaches for both phantom #1 (at top row) and phantom #2 (at bottom row). Graphs (c) and (d) were generated from data sets where the interface was opened and closed repeatedly between each imaging acquisition, as described in section 2.2. Graphs (a) and (b) were generated from data sets with the interface fixed for all acquisitions, thereby eliminating positioning errors. The data in these graphs allows us to determine the minimum regularization level required to invert the

matrix at each iteration, and in all cases this occurs in the 5th iteration ($\lambda = 1.0$), with the error

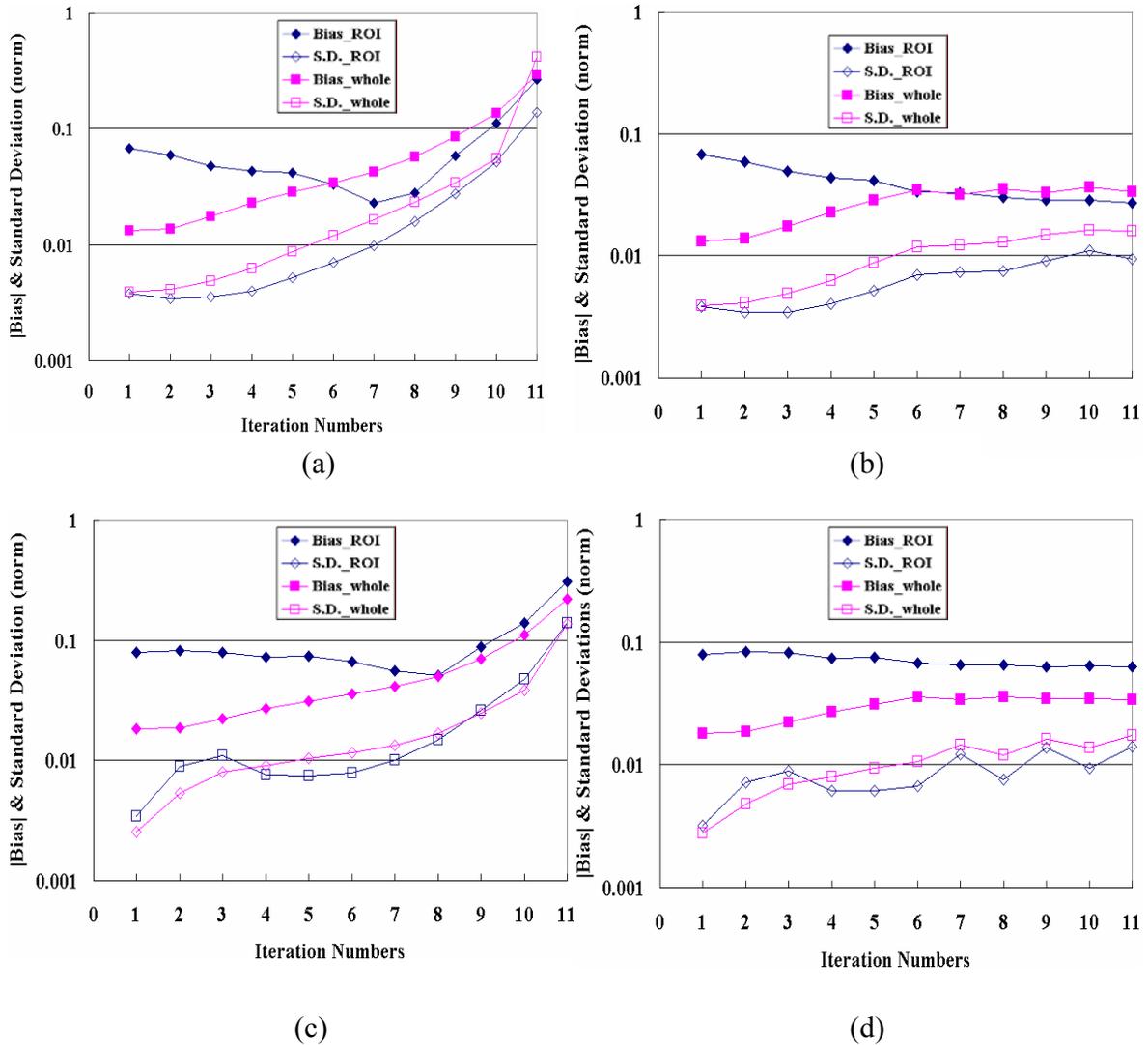


Figure 2.22. Calculated the average projection errors from 20 measurements for each of the phantom #1 (a,b) and the phantom #2 (c,d) by fixed decreasing regularization method and Levenberg-Marquardt regularization method. The fiber optic interface was fixed during the data collecting process for the graphs at the top row (a,b); and, the imaging system was readjusted after each data acquisition to make the sources and detectors touch well with the boundary of the phantom for the graphs at the bottom row (c,d).

increasing in the 6th iteration ($\lambda < 1.0$) if the Levenberg-Marquardt regularization method is not used. Continuation of the Levenberg-Marquardt method produces a typical oscillation in the projection error with minimal further reduction in the error.

As shown in Figure 2.22, the ranges of the projection error at the identical iteration step are the same, whether for either of the two phantoms or by two different procedures described in the previous section, except the errors at the first two iterations in Figure 2.22(a) and (b), where the interface was moved between imaging acquisitions, the mean projection error starts at a value which is actually slightly larger than when the interface was not moved, as shown in Figure 2.22(c) and (d). This would indicate that the majority of the error is not from positioning errors, but from other sources of variability, such as detector calibration or spontaneous fluctuations. This is also in agreement with the raw data measured of phase and amplitude, which indicates that the variation observed when the phantom is repositioned, is comparable to the variation observed when the phantom is not moved in the positioning array.

During the reconstruction, the squared projection errors in Figure 2.22 are the same at the first six iterations in all of the four graphs for the two methods of regularization parameter change. The minimum errors are at the 5th iteration when the regularization parameter (λ) is equal 1.0. It is at the 6th iteration, where the Levenberg-Marquardt method is required to maintain a converging solution. This level of regularization indicates that the minimum parameter value is equal to the normalized diagonal value, and below this the ill posed nature of the matrix dominates the inversion, thereby causing divergence of our initial

objective function shown in Equation (2-6), (i.e. the projection error). Hence, if the objective is to minimize this objective function then there is a clear need to implement the Levenberg-Marquardt method for regularization. This provides a stable and controlled convergence to the minimum solution.

2.4.2.3. Region of Interest (ROI) Analysis versus the Entire Image Field

For quantification of the reconstructed images and results, the absolute bias and standard deviation errors from 20 measurement data sets were calculated using equations (2-3) and (2-4). These were computed at each iteration step by both the regularization methods introduced in the previous section. In order to compare the reconstructed differences between in the region of interest (ROI) and the entire field of view, the absolute bias and standard deviation errors were averaged at all of the pixels within the ROI as well as in the whole image field. In the remaining figures, the absolute bias and standard deviation error for both the ROI and the entire field of view are displayed. The left column graphs in Figure 2.18 show the errors for the absorption coefficient of phantom #1 by using the fixed decreasing regularization approach; meanwhile, the right column graphs uses the Levenberg-Marquardt regularization approach as the only difference. The top row plots in Figure 2.18 follow the procedure #1, in which the fiber optic interface was not reset between data acquisitions; and the bottom plots are from the procedure #2, in which the phantom was

re-placed in the array between each imaging acquisition. The corresponding graphs for the scattering coefficient errors are shown in Figure 2.19 for these same phantom measurements. In Figure 2.20, the absorption coefficient errors calculated from imaging phantom #2 are shown, and in Figure 2.21 the reduced scattering coefficients are shown.

In summary, the averages of the absolute bias errors of both the ROI and the entire field of view are always greater than the standard deviation errors at the same iteration step. The absolute bias in the image and in the ROI always converges when using the Levenberg-Marquardt regularization method, although simply using a fixed decreasing regularization method can often lead to a lower bias error in the ROI (for example, see Figures 2.20 (a) versus (b), and (c) versus (d), in the black circle cases). In no cases does the standard deviation become greater than the bias error.

In spite of our observations that the image error does not decrease over the entire imaging field, the algorithm does recover heterogeneities with reasonable position and peak height values. Thus, we must conclude that the observations of image error in the region of interest would be different than the observations averaged over the entire field. The data plotted in Figures 2.18 through 2.21 for the ROI where the heterogeneity was located indicate that this is true. In contrast to the entire image, the bias error can always be decreased in the ROI, in the first several iterations of the reconstruction process. Interestingly, the Levenberg-Marquardt regularization method does not always provide the minimal level of

ROI bias, and using a continuous decreasing regularization pattern can reduce the bias to lower level (as seen in iteration number 9, in Figure 2.20). This observation is in spite of the fact that the projection error is no longer being minimized at these later iterations. Thus, even the projection error is reporting more information about the image field near the periphery, and not reflective of the ROI minimization.

The position of the ROI also appears to have significant bearing on the overall minimization, since the data from phantom #1, plotted in Figures 2.18 and 2.19, indicate that the bias of the ROI cannot be minimized as well as the data from phantom #2, plotted in Figures 2.20 and 2.21. This observation makes sense, since the object in phantom #2 is closer to the exterior wall, where the sensitivity of the response appears to be much greater. This effect is well known in diffusion imaging, yet the observation that the projection error minimization is not reflective of the ROI minimization indicates that alternative strategies for the image reconstruction algorithm might be considered. In particular, it might be more accurate to calibrate the reconstruction algorithm based upon phantom studies, such as these, rather than rely upon the projection error minimization, as defined by the Levenberg-Marquardt procedure.

2.5 Conclusions

This study illustrates the ability to examine image accuracy using analysis of the entire image mean square error, taken to be the summation of image variance and bias. By analyzing images of bias and standard deviation we obtain qualitative information about how the reconstruction program works, and by summing the values over all pixels quantitative values of image accuracy can be analyzed. Interestingly, in this type of ill-posed problem, where significant regularization is required to solve the inverse problem, the image stability is very good. As the reconstruction algorithm reduces the regularization level, the image bias is reduced and the image standard deviation increases. The optimal image is obtained after several iterations when a minimum in the summation of bias and standard deviation is reached. This minimum does not always correlate to the minimum of the projection error (i.e. objective function). Increasing measurement error levels causes a need to constrain the minimum regularization parameter to higher values in order to minimize the overall image mean square error, following a typical Levenberg-Marquardt style approach. Further analysis of the algorithm performance with measured data is examined in the second part to this study.

In summary, we have observed five important facts about the algorithm performance:

- (I) Bias error dominates the image and the average bias is not minimized in the reconstruction, due to dominant errors near the sources and detectors.
- (II) Random fluctuations from the detectors are not significant, indicating that

repeatability of the system is high.

- (III) The system performance is currently limited by systematic calibration accuracy, rather than random noise.
- (IV) Minimization of the objective function does not always directly correlate to minimization of the ROI bias error.
- (V) Levenberg-Marquardt style adjustment of the regularization parameter based upon the objective function provides a well-controlled image, but does not always minimize the ROI as much as alternative strategies.

These observations are important as the current system is evaluated in a clinical study for breast tumor characterization. In particular as the system is designed to quantitatively image hemoglobin concentration, oxygen saturation and water concentration, and so understanding the algorithm performance is fundamental to the success of this study. In general, since the reconstruction algorithm is initiated with an accurate value of the average absorption and scattering coefficients, the ability to quantify concentrations of these chromophores is fairly good. In particular, this analysis indicates that quantization of ROIs can be best achieved after 5 to 9 iterations, using the procedure outlined here. We expect a bias error of less than 20% in the ability to quantify absorption coefficients over regions larger than 1 cm. Further analysis of the effects of ROI position and size are ongoing with this procedure.

Beyond our current study, there has been significant interest in alternative regularization schemes that might be used to maximize the image accuracy, possibly improving the high

spatial-frequency information available [25, 60, 66, 70]. Yet, to use these newer methods, the measure of image performance must be shifted away from objective function minimization, towards image accuracy (i.e. mean square error). This change in the analysis allows a better examination of how changes in the regularization method affect the actual image accuracy, rather than the observed projection error. The analysis presented here is an appropriate tool to evaluate changes in the regularization strategy. Future studies will use this to examine ways to better quantify the region of interest in a manner that preserves the blinded manner in which clinical images must be reconstructed.

CHAPTER 3

Contrast-to-noise ratio based Contrast-Detail Analysis of System Performance

3.1 Introduction

Near-infrared (NIR) diffuse tomography can non-invasively quantify oxygen saturation, hemoglobin concentration, water concentration, scattering, and potentially exogenous chromophores, therefore it may provide an effective diagnostic tool for early detection of breast cancer [27, 62, 73, 90, 83, 99]. Tumors have a higher level of vascularity due to hyperactive angiogenesis leading to optical contrast to regular tissues in the NIR spectrum or light between 650 and 950 nm [11, 85]. Detecting these changes with optical absorption and scattering based tomography requires implementation of accurate and fast image reconstruction algorithms. Because the reconstruction process results in moderate-resolution images, it is becoming increasingly important to develop tools to accurately interpret the reconstructed image and promptly find the shape, size, and location of the tumor or the region of interest (ROI) [74, 89, 98]. Even more objectively, it would be desirable to automatically find regions of increased contrast and quantify their values. However, in most

of current preclinical and clinical NIR tomography studies, the shape and location of the ROI in the reconstructed image are decided by a radiologist projecting what is observed in the x-ray mammograms into the circular tomographic coordinate system of the NIR tomography images. Although this approach is practical and works reasonably well, it has some subjectivity associated with it as well as problems associated with tissue deformation. This approach becomes especially problematic when one considers that the image reconstruction process for NIR tomography is nonlinear and has its own unique set of artifacts and a spatially dependent contrast detail response [70, 76]. As NIR tomography transitions into a tool for imaging patients, a quantitative computational tool is necessary for consistent and fast ROI detection and delineation to eliminate use of subjective ROI detection. Furthermore, because evaluating a new imaging modality requires knowledge of the validity of the reconstructed images, this type of computational tool can be used to generate contrast-detail curves that can be used to objectively determine the system performance and allow accurate comparison with other imaging systems.

To develop an accurate and efficient computational tool for automatically and efficiently segmenting out the ROI in a laboratory or clinical NIR tomography imaging system, an automated algorithm to calculate the contrast-to-noise ratio (CNR) is used. Similar to the signal-to-noise ratio in digital signal-processing theory, the CNR is defined as the difference between the ROI and the background region values of the optical properties, i.e. the absorption and reduced scattering coefficients, divided by the average variation in the background, providing a relative difference property value. Although there is only one choice

for the ROI location, there are different choices for the background that can be used for the analysis. Ideally, the property distribution of the reconstructed image should be exactly the same as the real image whose background (or regular tissue) is uniform and identical everywhere. It is reasonable to randomly choose a ROI within the whole image. However, the intrinsic diffusion characteristic of the light traveling in the tissue makes every spot of the reconstructed image dependent on other locations of the same image, and generally there is a radial variation in the modulation transfer response function [70]. Thus it is equally justified for one to consider taking background values from the same radial locations of the ROI to avoid biasing the analysis to regions with a different spatial-frequency response. We need to take this into account and find the best solution for calculating the background properties. In this chapter, we investigate and compare three methods of choosing the background values, and the results are presented in this chapter.

A contrast-detail curve, where the line of minimum-detected contrast is plotted for various sized inclusions located within a tissue, is an accurate and objective way to evaluate the spatial resolution of a medical image system, especially in computed tomography and ultrasound [24, 38, 76, 78, 79]. It indicates the transition from a detectable heterogeneity to a statistically insignificant undetectable heterogeneity, and the shape of the curve provides a quantitative description of the system's performance and is particularly relevant where the clinical role of the system is in the regime of a low-contrast or low resolution modality. Traditionally, contrast-detail curves are derived subjectively by human inspection [5], but it is reasonable to expect some inter-observer variability in the development of these curves. Thus,

as a first attempt in this direction, we used an objective threshold as the measure of detection rather than using repeated observers. Given a chosen threshold value of CNR, then it is fairly straightforward to accurately calculate the CNR values of reconstructed images and find those images in which objects appear to be detectable. Thus, based on whether the object presence is above or below the threshold, a contrast-detail curve can be created that shows the minimum contrast required to detect objects at each size (i.e. detail is used for size in this types of resolution test), with the assumption that a CNR of 4 is required for detection of the object. The justification for this threshold level and the implications for clinical utility are discussed.

3.2 Theory and Method

3.2.1 Contrast-Detail Curve

Contrast-Detail (C-D) curve was developed as an experimental technique based on the Rose model to evaluate object detectability at the threshold of human visibility in a medical image [42]. Figure 3.1 describes the general procedure of obtaining a C-D curve in X-Ray medical imaging. It is easy to see that small objects require a high contrast to be visualized while small objects can be visualized at a low contrast. Therefore, if the threshold contrast is displayed on the vertical axis and detail on the horizontal axis, the contrast detail curve starts at the upper left corner (high contrast, small detail) and declines asymptotically toward the lower right corner (low contrast, large detail). The C-D curve indicates the transition from

objects that can be seen to those that can not. Contrast-detail curve is commonly used in CT, but more and more investigators use it to evaluate MRI, ultrasound, NIR tomography and other medical imaging systems.

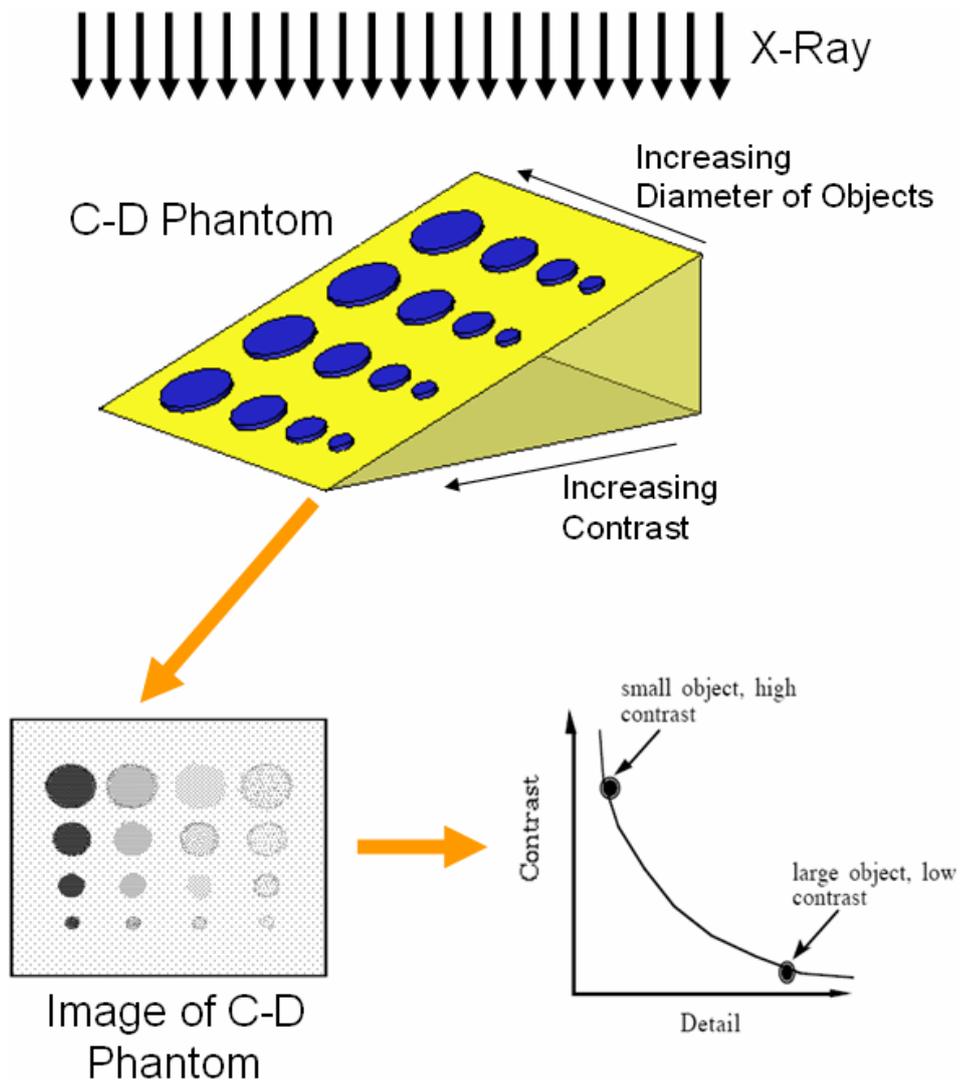


Figure 3.1 Formation of Contrast-Detail curve. Contrast detail curves are obtained by having observers detect circular targets in a radiograph of a C-D phantom as shown above. The contrast detail curve relates the size of the minimally perceptible object as a function of object contrast for different levels of the signal-to-noise ratio.

In NIR contrast-detail curve, where the line of minimum-detected contrast is plotted for various sized inclusions located within a tissue, is an accurate and objective way to evaluate the spatial resolution of a medical image system, especially in computed tomography and ultrasound. [[524, 38, 76, 78, 79] It indicates the transition from a detectable heterogeneity to a statistically insignificant undetectable heterogeneity, and the shape of the curve provides a quantitative description of the system's performance and is particularly relevant where the clinical role of the system is in the regime of a low-contrast or lower-resolution modality. Traditionally, contrast-detail curves are derived subjectively by human inspection, [5] but it is reasonable to expect some inter-observer variability in the development of these curves. Thus, as a first attempt in this direction, we used an objective threshold as the measure of detection rather than using repeated observers. Given a chosen threshold value of CNR, then it is fairly straightforward to accurately calculate the CNR values of reconstructed images and find those images in which objects appear to be detectable. Thus, based on whether the object presence is above or below the threshold, a contrast-detail curve can be created that shows the minimum contrast required to detect objects at each size (i.e., detail is used for size in this types of resolution test), with the assumption that a CNR of 4 is required for detection of the object. The justification for this threshold level and the implications for clinical utility are discussed.

3.2.2 Contrast-to-Noise Definition of NIR Tomographic Images

CNR is defined as the difference between the averaged optical coefficient within the ROI and the differences within the background region, divided by the averaged optical coefficient variation in the background. Because the background profile of the ROI is unknown after forward calculation as well as in real experimentally acquired data, one of the goals of this study is to find the proper way to calculate the background profile. Same as the method we introduced in Chapter 1 and 2, a circular geometry of 86mm in diameter for the simulation studies. As shown in Figure 3.2, the background has an absorption coefficient $\mu_a = 0.005 \text{ mm}^{-1}$ and a reduced scattering coefficient $\mu_s' = 1.0 \text{ mm}^{-1}$. Within this field, a spherical heterogeneity with a variable diameter was located 3 mm away from the edge as depicted in Figure 3.2, with a fixed $\mu_s' = 1.0 \text{ mm}^{-1}$ and a variable μ_a to provide the required contrast in absorption from 1.2:1 to 5.0:1. We accomplished the forward calculations of diffusion theory with a finite-element numerical solution using a circular mesh with a set of regularly spaced node positions. Noise can be added to either the simulated measured data or the optical property distribution. In this study, zero-mean Gaussian noise of 1% in amplitude and 1 deg in phase shift were added to the calculated boundary data to simulate the realistic condition area in our NIR system.

We developed three methods of choosing the relevant background values to calculate the CNR and construct the contrast-detail curves in this study. The three choices were (i) backgrounds sampled at the same radial distance in the image, (ii)

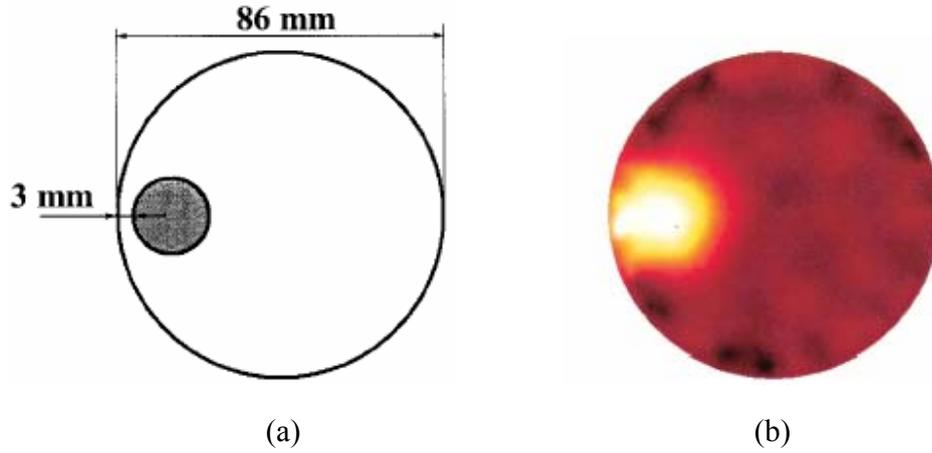


Figure 3.2 (a) Geometry of the simulation for optical property distribution, and (b) a typical reconstructed absorption coefficient image. The image in (b) was generated with simulated forward data with a 1% noise in amplitude and a 1° noise phase shift by use of zero-mean Gaussian distribution.

backgrounds of randomly sampled ROIs of the same size as the original, and (iii) all the available background. First, the background profile we chose in the image field had the same radial size and similar location of the target, and they did not overlap each other as shown in Figure 3.3(a); thus they had independent mean values and noises. As many non-overlapping ROI-shaped areas as possible were chosen to minimize the displacement error.

We can define the contrast in the reconstructed image as the difference of the mean values in the target and background; CNR can be defined as

$$CNR_I = \frac{\mu_{ROI} - \mu_{mean}}{\sqrt{\frac{\sigma_{ROI}^2 + \sigma_{mean}^2}{2}}} \quad (3-1)$$

where μ_{ROI} is the mean of the node values in the target; μ_{mean} is the mean value of the averaged node values in the total ROI-shaped background areas; and σ_{ROI} and σ_{mean}

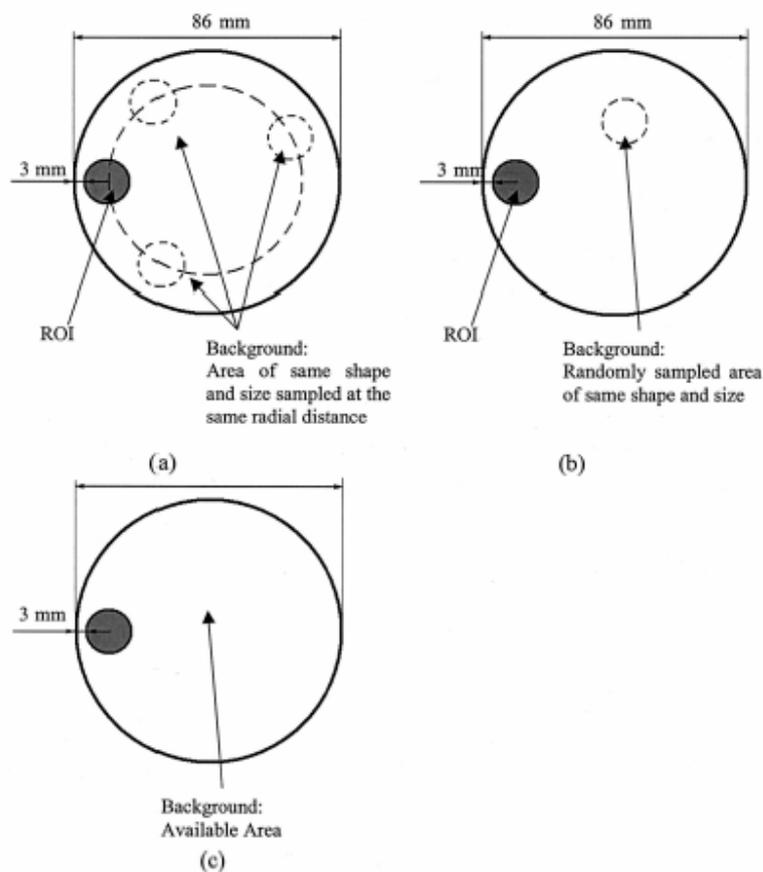


Figure 3.3 Schematic showing the three possible choices of backgrounds for calculating CNR: (a) sampled at the same radial location as the lesion object, (b) randomly sampled, or (c) sampled by use of the entire region outside of the target lesion object.

are the standard deviations of the collectivities of the target and the mean of the background areas, respectively. Because the circular simulated target and background areas have the same size, it is reasonable to assume that the target and background areas have the same noise weights.

However, the noise from the nodes that are closest to the field edge is higher because of the nonlinear image reconstruction process used here and because the weight functions (i.e.,

Jacobian matrices) have much higher contributions nearest the source and detector locations. Overall, this contributes to higher system noise in regions nearest the edge of the field and lower values toward the center of the field.

This effect is not accounted for in the above definition of CNR. As shown in Figure 3.3(b), to make sure that every node in the background has an equal chance to contribute to the background noise value, we introduce the second definition of CNR as

$$CNR_{II} = \frac{\mu_{ROI} - \mu_{random}}{\sqrt{\frac{\sigma_{ROI}^2 + \sigma_{random}^2}{2}}} \quad (3-2)$$

where μ_{random} is the mean value of the nodes within a region having a location that is randomly chosen in the variable background and σ_{random} is the standard deviation of these same nodes. Only one location of the ROI-shaped area was chosen in one calculation, such that the ROI background profile is truly random.

Although this definition has a good basis in statistical probability theory, the shortcoming of this definition is that the CNR depends on randomly chosen nodes in the whole image; therefore CNR is not exactly identical among different calculations.

In the first two definitions of CNR, ROI and the testing background have same area; thus a noise weight, which presents the contribution of a certain area noise to the whole area noise, is not necessary.

But if we use the whole available area as the testing background, as shown in Figure 3.3(c), we must include a noise weight into the CNR definition because the background and ROI have different contributions to the whole noise. The CNR than becomes

$$CNR_{III} = \frac{\mu_{ROI} - \mu_{background}}{\sqrt{\omega_{ROI}\sigma_{ROI}^2 + \omega_{background}\sigma_{background}^2}} \quad (3-3)$$

where $\mu_{background}$ is the mean value over the variable background and $\omega_{ROI} = AREA_{ROI} / (AREA_{ROI} + AREA_{background})$, and $\omega_{background} = AREA_{background} / (AREA_{ROI} + AREA_{background})$ are the noise weights.

We use this simple linear form of noise weight because the exact noise weight of different areas in the field is hard to determine because of the nonlinearity of the inverse problem.

3.3 Contrast-to-Noise Ratio Analysis of NIR Tomographic Simulation Images

3.3.1 Evaluating the Three Methods of Calculating the CNR

Assuming that the real diameter and location of the target are given, we calculated the CNR by using the three methods described above as we systematically varied the absorption contrast and the ROI size. Figure 3.4 shows the difference in the CNR values of the three methods in the simulation study. Figure 3.4(a), in which we used the first method of calculating the ROI background profile, is the CNR distribution for different ROI sizes and absorption contrasts; for each ROI diameter value, the CNR was calculated and assigned to the related spot in the CNR profile as the absorption contrast varies from 0 to 5. The resolution is 1 mm for the horizontal axis and 0.2 for the vertical axis. Similarly, the CNR

distribution profiles of the second and third background calculating methods are plotted in Figure 3.4(b) and 3.4(c). Theoretically, if we apply a practical detection threshold with a CNR of 4, which indicates that any reconstructed image with a CNR less than 4 is not detectable to the human eye, we can find the transition from a detectable ROI to a nondetectable ROI on any of the profiles plotted in Figure 3.4(a)–3.4(c). In practice, the transition boundary points around where the CNR transits from smaller than 4 to larger than 4 were calculated and served as the construction points to plot the contrast-detail curve. The interpolation method was used to find and smooth the transition boundary. As we can see from Figure 3.4(a)–3.4(c), the CNR profile calculated by the third method is apparently the best profile to use to find the transition boundary, as shown in Figure 3.4(d). The contrast-detail curve in Figure 3.4(d) shows the spatial resolution of the NIR breast imaging system used in our study.

Our choice of a CNR of 4 as the detection threshold was based on the simulation study performed on different ROI size and contrast combinations. As shown in Figure 3.5, both the target with a large size (23 mm in diameter) and low contrast (1.4 in absorption) and the target with the small size (3 mm in diameter) and high contrast (3.6 in absorption) can be observed in the reconstructed images if the CNR is greater than 4 for all three calculation methods. Thus it is reasonable to set the CNR to 4 as the detection threshold, although clearly a more in-depth study should be undertaken to determine which CNR value would correspond best to that used by humans. Nonetheless for this study we used a CNR of 4 as a possible threshold. The actual CNR values are plotted for all sizes and absorption contrast, as

shown in Figure 3.4(a)–3.4(c) as well as the system contrast-detail curve Figure 3.4(d) assuming a decision threshold CNR of 4.

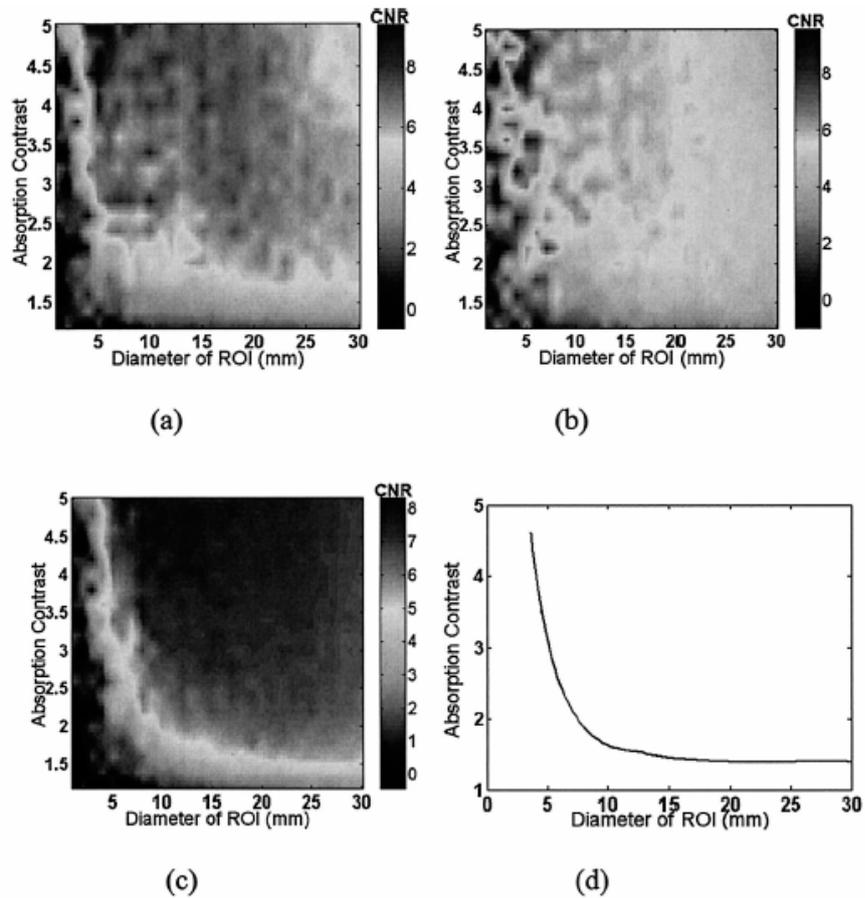


Figure 3.4. Graphs of CNR distribution (the color bar represents CNR values) obtained by the three different methods of choosing the background areas including (a) CNR_I, (b) CNR_{II}, and (c) CNR_{III}. (d) Based on the method shown in (c), the contrast-detail curve for the images are plotted in by a detection threshold with a CNR of 4.

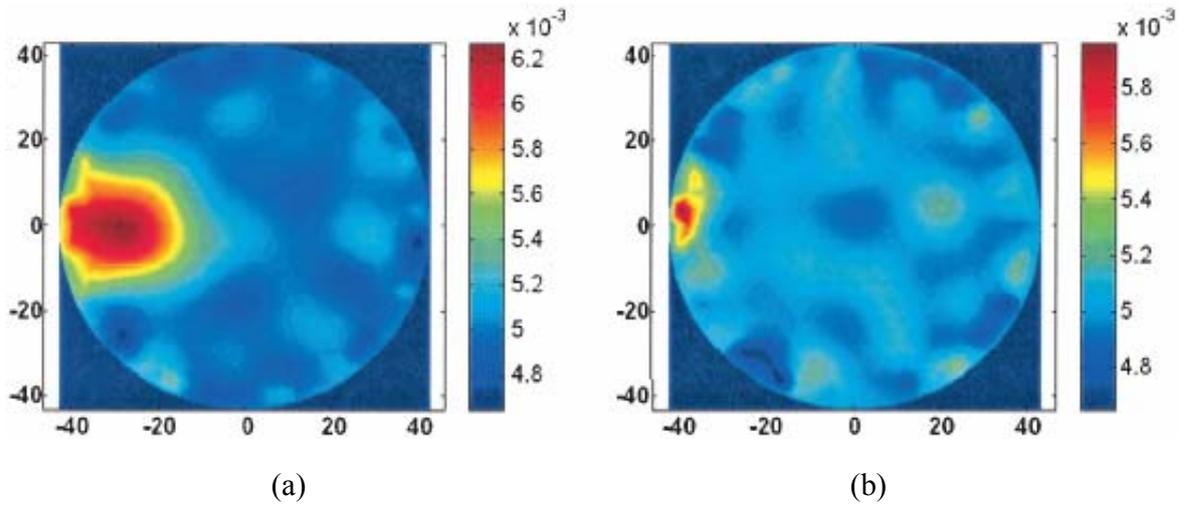


Figure 3.5. Reconstructed absorption images for (a) target with large size (23mm diameter) and low contrast (1.4) and (b) target with small size (3 mm in diameter) and high contrast (3.6) in which CNR is greater than 4 by all three calculation methods.

3.3.2 Comparison of ROI Location Detection: Optimized CNR Method vs. 2-D Convolution Method

Estimates of the locations of the ROI in the reconstructed images are necessary in our simulation and experimental studies. Two methods were used to detect the location of the ROI with different sizes and absorption contrast values where the step size for scanning was 0 to 5 mm. In the first method, it was assumed that the CNR should be the maximum when the suppositional region was at the location of the ROI, as shown in Figure 3.6(a).

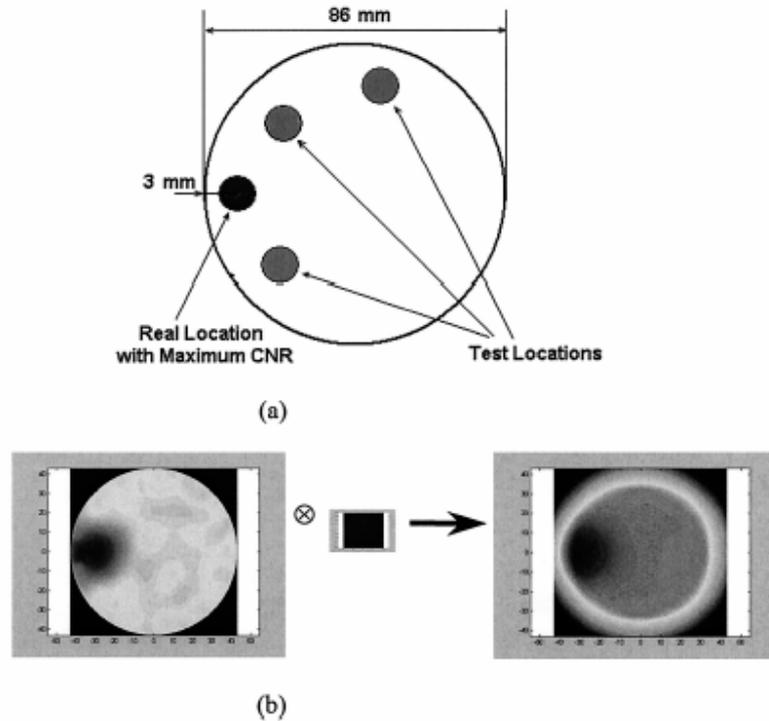


Figure 3.6 Two methods of ROI detection: (a) Detection by a search for the maximum CNR from all the possible locations in the image field where alternate possible locations are randomly tested and the sizes of these regions are varied. This approach is used to determine when the CNR calculation is maximal and yields an estimate of the size and location of the ROI but requires a significant amount of computational time to achieve the result. (b) 2-D linear convolution method where \otimes denotes the convolution process.

In this method, CNR was calculated by the third method defined in Equation 3-3, which includes all the available background as the background area. However, this algorithm requires a significant period of time to complete the calculations for all possible locations and sizes of the ROI within the field. Thus a 2-D linear convolution method was introduced to detect the ROI. In this method, a ROI pattern that has the same optical coefficient profile as the simulated lesion in computer simulations or a phantom study was created, and we developed the correlation image by convolving it with the reconstructed image in which the

highest correlation location corresponds to the maximum CNR, as depicted in Figure 3.6(b).

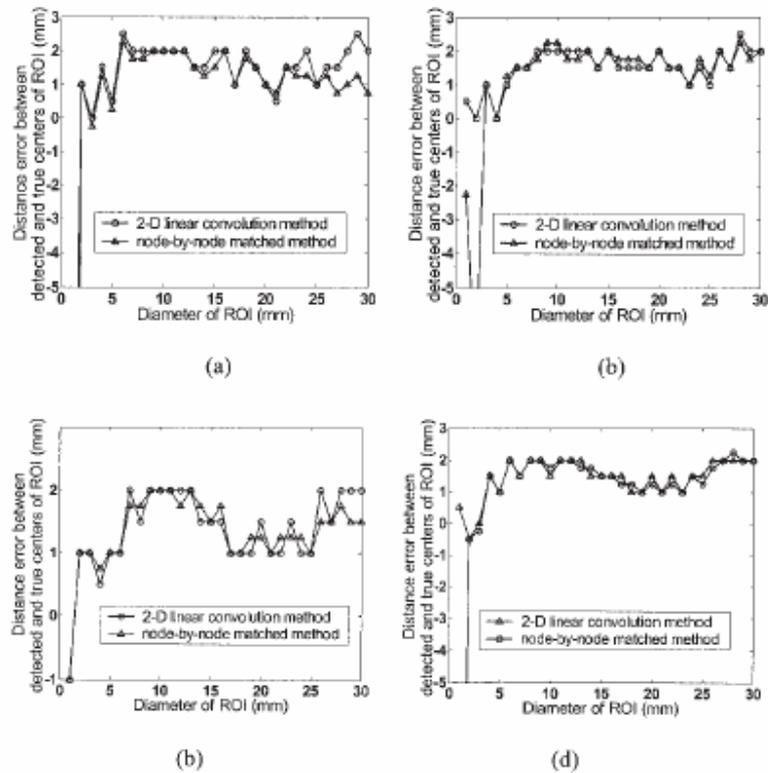


Figure 3.7 Distance error between the detected center of the ROI in the reconstructed image relative to the real location of the circular ROI center is plotted as a function of the diameter of the ROI by use of absorption contrast values of (a) 2.0, (b) 3.0, (c) 4.0, and (d) 5.0 in the target relative to the background.

Comparing this approach with the previous method, the advantage of the 2-D linear convolution method is that it decreases the calculation time from several minutes to a few seconds for processing one image.

As discussed above we used both the node-by-node method, in which the maximum CNR from all the suppositional regions in the image field is searched, and the 2-D linear

convolution method, in which the maximum correlation location between the ROI pattern and the reconstructed image profile was detected. In both methods we calculated the CNR by Equation 3-3, and the differences between the true location of the ROI and its detected location were compared at different contrast values, as shown in Figure 3.7.

Figure 3.7 shows that the error between the detected and the true locations of the ROI is even larger than the size of the ROI in both detection methods if the diameter of the target is less than 3 mm and the contrast is low [i.e., equal to 2 in Figure 3.7(a)]. As the contrast increases [Figure 3.7(b)-(d)], the location error from the node-by-node matching method is less than 1 mm, but it requires a maximum CNR much lower than 4 - the CNR threshold we assumed above. The error from the 2-D linear convolution method is also large when the diameter of the ROI is small. Thus neither of the two detection methods can find the location of the ROI if the diameter of heterogeneity is less than 3 mm. As the diameter of the ROI becomes greater than 4 mm, the location errors from both of the methods are around 1-2.5 mm. Although the location error from the convolution method is slightly larger, it is more than ten times faster.

3.4 Contrast-to-Noise Ratio Analysis of NIR Tomographic Phantom Study Images

To test the spatial resolution of our NIR imaging system, a solid phantom was used to mimic the optical properties that are described in Section 4. The detail of the NIR

tomographic imaging system at Dartmouth College and the phantom study setup and procedures have been described in detail in previous chapters.

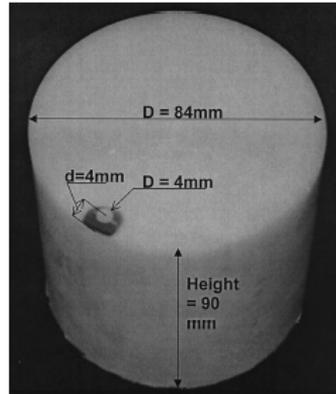


Figure 3.8 Photograph of the phantom used in our study. The 4-mm-diameter hole is filled with an Intralipid and ink solution with a variable absorption contrast from 2:1 to 9:1 and a similar scattering coefficient. Note that the 4-mm hole is outlined with dark ink to make it visible, but this ink on the upper surface of the phantom does not affect the image quality of the interior of the phantom.

The photograph of the actual phantom we used in our study is shown in Figure 3.8. The 4-mm-diameter hole was filled with an Intralipid and ink solution with a variable absorption contrast from 2:1 to 9:1 and a closely reduced scattering coefficient with the phantom. Figure 3.9 shows the reconstructed absorption images with the absorption contrast values of 2.0, 4.0, 6.0, and 9.0, respectively, and their maximum CNR values are 2.9, 3.8, 4.9, and 5.8 calculated by the third definition of CNR. The noise in the experiment includes not only the random noise but also the systematic noise. (Pogue, 2002) The random noise in the experimental measurement is typically from the stochastic detector fluctuations, which is almost equal to or a little less than 1% of the noise level that we added to the calculated boundary data in our simulation study.

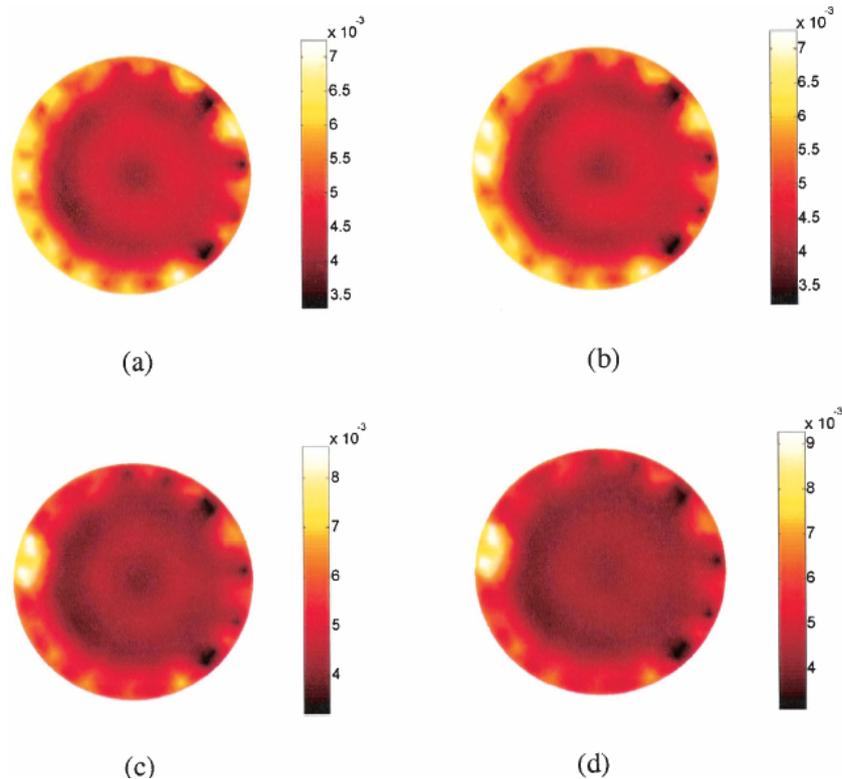


Figure 3.9 Reconstructed images of the absorption coefficient (μ_a), with the object size of 4 mm in diameter and the absorption contrast equal to (a) 2:1, (b) 4:1, (c) 6:1, and (d) 9:1.

As an indication, a comparison of these results with the results in the simulation, we can see that there are more obvious artifacts in the reconstructed images of the phantom study, especially at the edge of the field. However, we still can observe that the reconstructed positions of the ROI shift around 2 mm toward the edges of the field, as discussed above. Also we can see that, if the CNR values are higher than 4, the 4-mm-diameter object is clearly visible on top of the noisy background (detectable by observers or by the computer) as shown in Figure 3.9(c) and 3.9(d). But if the CNR value is lower than 3.0, as indicated in Figure 3.9(a), even the computer cannot detect the ROI location by any of our location

detection methods. Thus the detectable size of our NIR imaging system is at least 4 mm, providing a good match to the results we achieved in our simulation.

3.5 Discussion

Comparing our three CNR calculation methods of the graphs in Figure 3.4 illustrates that the choice of background value can significantly affect the decision criteria of whether the region is detected. The first definition of CNR that was based on the matched filter method did not include information about the nodes closest to the edge of the image field. This contributed to a higher noise because of the effect of spatially dependent noise in NIR tomographic images, where most high-frequency noise is located in the periphery of the image and lower-frequency noise is located in the interior of the image. In the second method for calculation of the CNR, we used the background nodes randomly in the image field; thus all the nodes in the background had the same chance to be sampled, but it was not stable when the target was small. Again, this is likely due to problems with a spatially dependent frequency response. Finally, the third method of calculating CNR appeared to be the best, as it included all the nodes in the background image outside of the ROI and corrected the noise weights based on the size ratio of the ROI and the whole image field. This third approach is also easy to implement and provides a reasonable approach for NIR image analysis.

Based on Figure 3.4(c) from the third method, we created Figure 3.4(d) using the threshold value of 4 for the lower limit on CNR. The blue regions represent the undetectable conditions (CNR less than 4) and the red regions represent the detectable conditions (CNR

greater than 4). From the data in Figure 3.4(c) we established that the spatial resolution of the NIR imaging system was 4 mm in diameter when the contrast is high (i.e., greater than or equal to a contrast of 4.4), and this was established by theoretical simulation and experimental verification. This is interesting in that the system noise level did not appear to have much of an affect on this detection limit, indicating that the lower size limit is perhaps more of a function of the physical phenomenon of how light is diffused out in a highly scattering medium such as tissue. Therefore, if the contrast between the object and the background is high enough, the minimum-detectable size of a target object depends on the physical characteristics of the imaging system and the reconstruction method, but not on the noise of the imaging system or the object characteristics. Also, the contrast resolution (i.e., contrast required for detection at low resolution) was 1.4 when the diameter of the target was large (i.e., greater than or equal to 17 mm) with 1% zero-mean Gaussian noise and 1° phase shift added. This latter observation would indicate that lower-contrast regions that are large would be difficult to detect when the contrast is less than 1.4 relative to the background; however, this is likely the regime where this CNR threshold of 4 is breaking down. Further analysis of detectability of large objects with a CNR lower than 4 should be performed with multiple observers to determine what the lower limit should really be in this region. We also note that a similar reconstructed ROI image shift of 1-2.5 mm toward the edge of the field occurred in both the simulation and the phantom studies, which is likely an artifact of the strong radially dependent sensitivity that occurs in this type of imaging.

In Figure 3.4(d) the contrast-detail curve indicates that the system resolution is limited

by both the ROI size and the system noise characteristics. As the contrast increases, the upper left corner of the curve shows the smallest object size that can be detected by the reconstruction system or the spatial resolution of the system. Then, as the ROI size increases, the lower right corner of the curve indicates the smallest contrast in object-to-background levels that can be detected by the system. This latter definition is often called the contrast resolution of the system. The spatial resolution and contrast resolution are dependent on the experimental setting of the detection threshold. In our study, the contrast-detail curves are all developed based on the assumption that the detection threshold of the CNR is a value of 4.0. Again, although this number is a useful threshold value and agrees with our experience in small objects, it is likely that it could be considerably lower when one is working on the lower right portion of the contrast-detail curve, where objects are large and the contrast is low. In fact, this CNR value is likely a nonlinear function in itself, which is dependent on object size, contrast, and features of the image reconstruction algorithm. Thus, to complete this study, we chose a CNR value of 4 as a practical working number and have begun a follow-up study to further define how the CNR varied with object and image reconstruction characteristics.

As shown in Figure 3.4, the node-by-node matched and 2-D linear convolution methods have similar results for the detection of the ROI location in the reconstructed image. Neither of the two methods was able to detect the ROI location when the diameter of the object being imaged was less than 3 mm. While the diameter of the target increases, the reconstructed positional error range was near 1-2 mm toward the field edge. Meanwhile, the location errors

have no obvious change as the absorption contrast changes from 1.4 (the contrast resolution of our imaging system) to 5 if the diameter is the same. The 2-D linear convolution method appears optimal for the purposes of our system, mainly because it is significantly faster than a random or systematic search algorithm, especially when the size of the target is small.

The lower spatial resolution limit at high contrast levels was determined to be near 4 mm, and we established this by both theoretical simulation and experimental verification using detectability of round objects. This is interesting in that the system noise level did not appear to have much of effect on this detection limit, indicating that the lower size limit is perhaps more of a function of the physics of how light is blurred out in a diffusing medium such as this. Therefore, if the contrast between the object and the background is high enough, the minimum-detectable size of a target object depends on only the physical characteristics of the imaging system and the reconstructed method, such as the width of the laser Gaussian beam and the mesh used for reconstruction, but not the noise of the imaging system or measured object. Also, we note that a similar reconstructed ROI image shift of 1-2.5 mm toward the edge of the field occurred in both the simulation and the phantom studies (see Figure 3.7), which is likely an artifact of the strong radially dependent sensitivity that occurs in this type of imaging. The amount and direction of the object shifts in the reconstructed images depend on the true position of the target. If the heterogeneity is placed near the edge of the field, the ROI in the reconstructed images is close to the edge; on the other hand, the ROI shifts toward the center if the target is near the center.

3.6 Conclusions

In this chapter, we calculated the CNR by choosing the background areas including (i) ROIs sampled at the same radial distance in the image, (ii) randomly sampled ROIs of the same size as the original, and (iii) all the available background. According to the comparison, the third method appears to be optimal for NIR tomography because it includes the nodes closest to the field edge. This approach contributed higher background noise and is more stable than the other two methods, especially when the target is small. On the basis of the CNR defined by the third calculation method and a system decision threshold with a CNR of 4, we find that the NIR imaging system spatial resolution is 4 mm in diameter and the contrast resolution is 1.4.

Two approaches were tested to find the location and size of the object based on maximal CNR, including a random node-by-node matching process and a 2-D linear convolution method. Although the results from these two methods are similar, the convolution method is faster than the node-by-node matching process.

Finally, the experimental study on a phantom with similar size and optical properties as in our simulation indicated a series of similar reconstructed images and system spatial resolution. Future plans will involve implementation of this ROI detection tool with our NIR prototype imaging system. The assessment and validation of the threshold CNR value of 4.0 needs to be addressed through multiple observer sessions where images of different size, and contrast levels are examined. Furthermore, because only a single anomaly was used in this

study, we would like to upgrade this tool set to diagnose tissue with multiple anomalies. It is also useful to investigate simulated or experimental objects with highly heterogeneous backgrounds such as breast tissue. Future studies should help address these various complex issues associated with NIR as a new and promising imaging modality for physiology and pathology.

CHAPTER 4

Observer Performance Studies with Receiver Operating Characteristic (ROC) Analysis

4.1 Introduction

The receiver operating characteristic (ROC) methodology has been widely used to address the clinical efficacy of medical imaging systems [5, 12, 54, 106]. In an ROC study, the reader views images, some of which contain single or multiple abnormalities, and the rest of which are normal. The reader then assigns numeric (typically five) ratings to each image as an indication of their confidence level that the image is abnormal. The resulting rating data are then plotted on an ROC curve, which entails the true-positive fraction plotted against the false-positive fraction. Summary measures of the curve are typically used as an objective measure of the ability of the reader to detect objects in the images, representing the image quality of the medical imaging modality when applied in a human diagnosis task. These summary measures include the area under the entire curve and the partial area under the curve in a particular region of interest. In standard ROC methods, the major focus is to assess diagnostic utility of the medical images, where the complexity of the target object location is often eliminated by clearly specifying the ROI (region of interest) in the images. However, in

some complicated medical imaging applications where the expected image resolution is spatially dependent, the applicability of the standard ROC analysis is very limited. Recent developments in localization-response ROC (LROC) analysis statistically offer more understanding of medical imaging methodology, in terms of measuring the conjoint ability of detecting and correctly localizing the actual targets in medical images. These developments include simultaneous ROC/LROC fitting [94] and alternative free-response ROC (AFROC) analysis [12, 13]. The LROC plots the probability of successfully detecting as well as locating objects in abnormalities images versus the probability of falsely detecting objects in normal images, as the detection criteria are varied. Several models have been proposed to be used in LROC analysis, including the discrete-location models [92] and the general detection-localization model [94], however, both models hold their particular assumptions. In this thesis, both human readers with or without medical imaging background experience and a “computational reader” were asked to specify the location of the suspicious area of abnormalities, which makes both ROC and LROC techniques applicable.

Despite the essential simplicity of the fundamental concepts of ROC analysis in medical imaging, professionals who design and perform ROC studies often find that a lot of subtle issues related to experimental design and data analysis must be confronted in practice. Such issues include:

- Case selection, collection and presentation
- Reader selection, grouping and characterization
- Dealing of pooled data, e.g. differences between single reader and multi readers data
- System error and tolerance analysis;
- Localization of data

- Strategies for data analysis and curve fitting
- Confidence of test results and conclusions

This work attempts to cover many of these issues both prior to the observer performance studies and, design a fast, full-scale and informative medical imaging methodology evaluation process for NIR tomography, as discussed in detail in the following sections of this chapter.

4.2 ROC and Localization-Response ROC Analysis in Medical Imaging

4.2.1. Sensitivity and Specificity of Diagnostic Tests

ROC analysis is widely employed to evaluate observer diagnostic performance in "two alternatives" classification of tasks in which both the stimulus (patients' real condition) and response (radiologists' diagnosis) have only 2 alternatives: "normal" or "abnormal". Consider a diagnostic test of N patients of which n_1 patients are "diseased" or "abnormal" and n_0 patients are "non-diseased" or "normal", the so-called "2 x 2 diagnostic table" completely described the observer's diagnostic performance on the number N of patient images as shown in Figure 4.1 (a).

A 2 by 2 table is the listing of all the possible combinations for a pair of binary variables. The data is laid out in a grid and the numbers in the grid represent the number of occurrences of certain combinations of the two variables. In here, the two variables are the patient's real condition and the observer's diagnosis. The observer gives m_1 positive and m_0 negative diagnosis responses to N patients of which n_1 are truly diseased and n_0 are truly non-diseased (see Table 4.1). "TP" (True Positives) and "FP" (False Positives) represent, respectively, the number of diseased and non-diseased patients who are diagnosed positive, while "TN" (True Negatives) and "FN" (False Negatives) are the number of diseased and non-diseased patients who are diagnosed negative. Figure 4.1 (b) is a numeric example of a diagnostic study. If we

divide each entry in the 2 x 2 diagnostic table by N, the total sample size, we get cell probabilities, as shown in Figure 4.1 (c).

		Disease Information		
		Present (D ⁺)	Absent (D ⁻)	Total
Reader's Diagnosis	Positive (T ⁺)	True Positives (TP)	False Positives (FP)	m ₁
	Negative (T ⁻)	False Negatives (FN)	True Negatives (TN)	m ₀
	Total	n ₁	n ₀	N

(a)

		Disease Information		
		Present (D ⁺)	Absent (D ⁻)	Total
Reader's Diagnosis	Positive (T ⁺)	17	25	42
	Negative (T ⁻)	9	29	38
	Total	26	54	80

(b)

		Disease Information		
		Present (D ⁺)	Absent (D ⁻)	Total
Reader's Diagnosis	Positive (T ⁺)	21.25%	31.25%	52.50%
	Negative (T ⁻)	11.25%	36.25%	47.50%
	Total	32.50%	67.50%	100.00%

(c)

Figure 4.1 The 2x2 Diagnosis Table. (a) The observer gives m₁ positive and m₀ negative diagnosis to N patient images of which n₁ are diseased and n₀ are non-diseased. (b) An example of diagnostic test of 80 patients and (c) the 2x2 cell probabilities table

From the "2 x 2 diagnostic table" we can also calculate the "Sensitivity" and "Specificity" commonly used in medical literature as an indication of discriminatory accuracy of the diagnostic study. To understand the physics and meaning of sensitivity and specificity,

we first need to understand the meaning of “probability” and “conditional probability”. As discussed in detail in “Appendix: Statistics and Probability Theory and Terms”, the word “probability” is used to mean the chance that a particular event (or set of events) will occur expressed on a linear scale from 0 (impossibility) to 1 (certainty), also expressed as a percentage between 0 and 100%. A conditional probability is the probability of an event given that another event has occurred. According to the axioms of probability and the definition of conditional probability specified in “Appendix: Statistics and Probability Theory and Terms”, we can define the sensitivity as the conditional probability of positive diagnosis given that the patient is in fact diseased or abnormal, i.e.:

$$\text{Sensitivity} = P(T+ | D+) = \frac{TP}{n_1} \quad (4-1)$$

Sensitivity represents the proportion of truly diseased persons in a screened population who are identified as being diseased by the test. It is a measure of the probability of correctly diagnosing a condition. In the same manner, we define the specificity as the conditional probability of negative diagnosis given that the patient is in fact non-diseased or normal:

$$\text{Specificity} = P(T- | D-) = \frac{TN}{n_0} \quad (4-2)$$

Specificity is the proportion of truly non-diseased persons who are identified by the screening test. It is a measure of the probability of correctly identifying a non-diseased person. More straightforwardly, in some medical literature, the sensitivity and specificity are also called “true positive rate” and “true negative rate”.

4.2.2. Concept of ROC Curve

Sensitivity and specificity usually appear together because use of the sensitivity or specificity alone is inadequate and is not a clear measure of a system utility or performance [86].

In practice, however, diagnostic discrimination capacity of an observer in a specific diagnostic test is usually not perfect because some diagnoses are made from various states of symptom or evidence. In other words, the diagnosis depends on the confidence level of diagnostic evidence, i.e. the confidence threshold. Thus it is more informative and meaningful to design the diagnostic tests on a “confidence rating” scale -- either on a fixed number of discrete response categories or a continual test variable-- and then calculate different sensitivity and specificity pairs, which finally lead to the generation of Receiver Operating Characteristic (ROC) curve for that particular diagnostic.

Defined in “Appendix: Statistics and Probability Theory and Terms”, the probability density distribution function of a radiologist’s confidence in a positive diagnosis for particular diagnostic task are shown schematically in Figure 4.2. The degree of overlap of the diseased and non-diseased distribution function completely determines the ability of the test to distinguish a diseased patient from a non-diseased one. As shown in Figure 4.2, for a specific decision or confidence threshold value x_c we can calculate a (sensitivity and specificity) pair. As x_c increases, specificity increases at the expense of sensitivity.

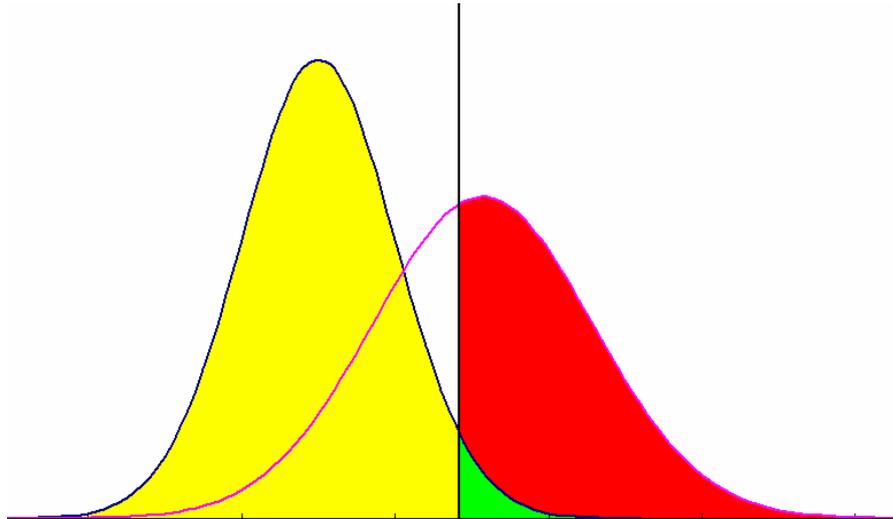


Figure 4.2 Example of probability density distributions of an observer's confidence in a diagnostic test. The observer's diagnosis is made according to the confidence threshold, which varies to generate different (sensitivity, specificity) pairs and thus an ROC curve.

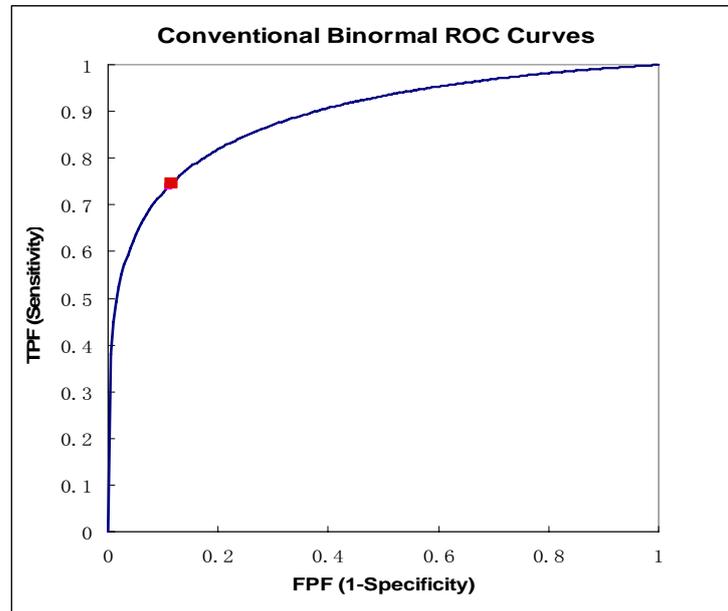


Figure 4.3 The ROC curve achieved from Figure 4.2 as the observer's confidence threshold varies. The square point on the curve corresponds to the observer's confidence threshold line shown in Figure 4.2

To graphically present the relationship of sensitivity and specificity, people generates the Receiver Operating Characteristic (ROC) curve which plots sensitivity, also known as True Positive Rate (TPR) versus (1-specificity), also known as False Positive Rate (FPR), as shown in Figure 4.3.

4.2.3. ROC Curve Fitting

ROC curves are usually acquired by proper modeling and fitting to well-designed experimental trial data for the probability density distribution functions. There are several methods that have been published to plot an ROC curve based on discrete or continuous test data, of which there are 2 basic categories: non-parametric or parametric - depending on if the approach is to assume a parametric model for the probability density function or not.

The empirical non-parametric approach is to calculate the ROC using the empirical histogram distributions in which there is no need for structural assumptions and parameters to modeling or fitting the density function. Though the empirical non-parametric method is easy and robust in general cases, it is not a smooth fitted curve and no statistical confidence level is available for evaluation. A great improvement to the empirical non-parametric method is the so-called non-parametric kernel smoothing technique [106, 108]. In this method, a kernel function estimating the densities of the distribution functions in the diseased and non-diseased populations and a bandwidth are applied and optimized to numerically

represent the distribution functions, thus resulting in a plot, which is smooth and therefore an optimal ROC curve.

More objective and specific case oriented methods require that some assumption be made regarding the functional form of the ROC curve. Several functional models have been proposed since the ROC analysis was first developed in signal detection theory [97]. One of the most successful models, the binormal ROC method, which assumes that a pair of latent normal decision-variable distributions underlies ROC data (similar to that depicted in Figure 4.2), has been widely used for ROC curve fitting [54, 95, 96]. According to this model, which includes two parameters, each ROC curve is assumed to have the same functional form as that implied by two Gaussian decision variable distribution functions. With the binormal model, the task of curve fitting becomes one of choosing numerical values for the two parameter pair (i.e. mean and variance) to best represent the measured data in observer performance studies.

4.2.3.1. Non-parametric Kernel Estimation of ROC

In the kernel estimation method, the diagnostic probability density distribution functions of normal and diseased cases are estimated to be in the form of

$$\hat{f}(t) = \frac{1}{n_0 h} \sum_{i=1}^{n_0} K\left(\frac{t - x_i}{h}\right) \quad (4-3)$$

$$\hat{g}(t) = \frac{1}{n_1 h} \sum_{i=1}^{n_1} K\left(\frac{t - y_i}{h}\right) \quad (4-4)$$

where x_i and y_i are the empirical distribution points of the diagnostic data for normal and

diseased cases with a total sample size of n_0 and n_1 , respectively. $K = \int k$ is a mean 0 kernel density function, and h is the so-called “bandwidth” which controls the degree of smoothing. This method gives a smoothed version of the empirical non-parametric probability density distribution functions, i.e. the relative frequency histograms, thus providing a practical method to generate a smoother non-parametric ROC curve, as shown in Figure 4.8. It is considered to better match the real ROC curve in the case of parametric model distribution assumptions are not valid and lack-of-fit occurs for parametric or semi-parametric ROC methods. The disadvantages of kernel estimation of ROC, however, are also obvious because it may give inaccurate estimation when the diagnostic density distribution function is partially near zero. Non-parametric kernel ROC analysis was used for Local ROC curve

4.2.3.2. Maximum-Likelihood Estimation of Parametric and Semi-parametric ROC

In 4.2.3.1, we introduced a Kernel function based ROC analysis technique, which one of the most popular non-parametric ROC estimation methods. In general, however, the non-parametric ROC fitting methods are unable to provide a real smooth ROC curve due to limited test data. In parametric ROC analysis, some particular assumptions concerning the mathematical form of the probability density distribution function of an observer’s confidence in a diagnostic test are necessary for real parametric smooth ROC curve fitting. Once a parametric model is set up, one can use advanced mathematical technique to fit the model to the test data, and then construct smooth ROC curves and perform error analysis thereafter. In some studies, parametric model based ROC estimation is divided into 2 groups:

parametric approaches and semi-parametric approaches depending on whether the measured data are re-grouped before the fitting process [86].

The maximum-likelihood method is one of the most successful mathematical techniques to fit parametric ROC models. Theoretically, the maximum-likelihood method is independent of the model-based assumptions underling parametric ROC analysis. But its most successful application is in parametric model, such as binormal model [20], logistic model [7] and Gaussian rating model [102]. In following paragraphs, the maximum-likelihood method is briefly reviewed as well as gives some results of its application in the binormal model of ROC analysis used in this thesis.

We introduced the probability density distributions of an observer's confidence in a diagnostic test in section 4.2.2. Suppose the outcome of a diagnostic test results in a continuous measurement t with diagnosis probability distribution function of $F_N(t) = P(T/D+)$ and $F_D(t) = P(T/D-)$, where N and D denotes actual normal and diseased condition. In a typical parametric model, $F_N(t)$ and $F_D(t)$ are of the same form and are fully defined by two parameters, i.e. $F_N(t) = F(t)$ and $F_D(t) = F(bt - a)$. According to the definition of the cumulative distribution function F , defined in "Appendix: Statistics and Probability Theory and Terms", for a given threshold value t_c whereby observer diagnoses a subject as diseased if $t > t_c$ and normal if $t \leq t_c$, we have

$$\text{FPR} = 1 - \text{specificity} = 1 - P(T/D-) = 1 - F(t_c) \quad (4-5)$$

$$\text{TPR} = \text{sensitivity} = 1 - P(T/D+) = 1 - F(bt_c - a) \quad (4-6)$$

If we know the form of F - for instance, in the binormal model, it is the standard-normal

cumulative distribution function Φ - and the value of a and b, we can plot the ROC curve by varying t_c . In practice, the decision axis t is partitioned into n categories by $n-1$ criteria, $t_i, i = 1, 2, \dots, n-1$ so that the observer makes a diagnosis R_i if the decision variable falls in the interval:

$$t_{i-1} \leq t \leq t_i, \quad i = 1, 2, \dots, n+1, \quad (4-7)$$

here $t_0 = -\infty$ and $t_{n+1} = +\infty$.

To simplify the formula system, we denote D- as S_1 and D+ as S_2 , and let $P(R_j/S_i) = P_{ij}$. If every trial of the diagnostic test is independent of each other, these trials will follow independent multinomial distributions. According to the “maximum-likelihood” estimation theory [80], if we take n statistically independent samples from the population $\{X_1, X_2, \dots, X_n\}$ and let the conditional probability density function on the i th sample given the parameter be denoted by $f_i(x_i|S)$, the joint conditional probability density function of all samples is given by

$$L(S) = f_n(x_1, x_2, \dots, x_n | S) = \prod_{i=1}^n f_i(x_i | S) \quad (4-8)$$

Thus, if the observer assigns R_j to S_i a total of r_{ij} times during a diagnosis test, the probability of obtaining the result data, i.e. the likelihood function is

$$L(S) = P_{11}^{r_{11}} \cdot P_{12}^{r_{12}} \dots P_{1n+1}^{r_{1n+1}} \cdot P_{21}^{r_{21}} \cdot P_{22}^{r_{22}} \dots P_{2,n+1}^{r_{2,n+1}} = \left(\prod_{j=1}^{n+1} P_{1j}^{r_{1j}} \right) \cdot \left(\prod_{j=1}^{n+1} P_{2j}^{r_{2j}} \right) \quad (4-9)$$

A condition for finding the maximum likelihood estimate is obtained by differentiating the likelihood function, given by (4-6), setting the resulting equation equal to zero and solving for S . It is often easier to differentiate the natural logarithm of the likelihood function, and

this is permissible since the logarithm is a monotonically increasing function of its argument: the value of the independent variable maximizing the logarithm of a function will also maximize the function itself. Thus a condition that the maximum likelihood estimate of the test result, i.e. the parameter S , must satisfy is

$$\frac{d}{dS} \ln L(S) = \frac{d}{dS} \ln \left(\left(\prod_{j=1}^{n-1} P_{1j}^{r_{1j}} \right) \cdot \left(\prod_{j=1}^{n-1} P_{2j}^{r_{2j}} \right) \right) = \frac{d}{dS} \left(\sum_{i=1}^2 \sum_{j=1}^{n-1} r_{ij} \ln P_{ij} \right) = 0 \quad (4-10)$$

From the definition of P_{ij} and (4-3, 4, 5), we know

$$P_{1j} = F(t_j) - F(t_{j-1}) \quad (4-11)$$

$$P_{2j} = F(bt_j - a) - F(bt_{j-1} - a) \quad (4-12)$$

Given a specific ROC model assumption, Solving (4-7, 8, 9) will give a fixed form of F , i.e. fixed value of a and b , and $n-1$ cut points $t_1, t_2, t_3 \dots t_{n-1}$. In practice, a direct solution to Equation (4-7, 8, 9) is impossible, but it can be iteratively solved by suitable numerical techniques, such as the “method of scoring” [20, 21, 30, 96].

The maximum-likelihood fitting process explained above requires ordinal category data. In many diagnosis tests, however, the observers give continuous data rather than 4-5 fixed ratings. In this case, pre-process steps are required to convert the continuously-distributed data into ordinal category data before applying the maximum-likelihood estimation theory based numerical ROC curve fitting process [55]. This process assumes that maximum-likelihood estimation of an ROC curve from continuously-distributed data is equivalent to maximum likelihood estimation from ordinal category data if truth state runs in the rank-ordered data, a unique data conserving only the sequence but ignoring the numerical

value of the continuously distributed data, are interpreted as categorical data. This process, in combination with the maximum-likelihood fitting algorithm, is often referred as the semi-parametric approach because the parametric model is required only after a monotone transformation from continuously-distributed data to grouped data [86].

Table 4.1 Comparison of Parametric and Semi/Non Parametric ROC Fitting Approaches

	Parametric Approaches	Semi-Parametric Approaches	Kernel Density Estimation Based Non-Parametric Approaches
Assumption	<ul style="list-style-type: none"> The cumulative distribution functions for normal and diseased populations follow a parametric family, e.g. normal or lognormal. 	<ul style="list-style-type: none"> The underlying distributions of the grouped normal and diseased test measurements can be transformed to normal distributions by an unspecified monotone transformation of the test measurement axis. 	<ul style="list-style-type: none"> Free of any distributional assumptions in that it depends only on the ranks of the observations in the combined sample
Advantages	<ul style="list-style-type: none"> Yield a smooth ROC curve completely specified by a small number of parameters Hypothesis tests, estimates, and confidence intervals for the parametric binormal ROC curve and its parameters are well developed 	<ul style="list-style-type: none"> Less strict assumption than that of fully parametric approaches which requires explicit distributional form for continuous measurements 	<ul style="list-style-type: none"> Closely follows the details of the original data Free of any parametric assumptions Universal fitting approaches for any observer operating characterization experiments
Disadvantages	<ul style="list-style-type: none"> Very sensitive to departures from distributional assumptions The fitting algorithm needs to take into account the errors in both FPR and TPR 	<ul style="list-style-type: none"> Still assume that a single monotone transformation exists that would make both latent distributions normal simultaneously Still sensitive to departures from distributional assumptions 	<ul style="list-style-type: none"> The resulting ROC curve is a series of horizontal and vertical steps not smooth curves It is unreliable at the ends of the ROC curve, requires some ad hoc experimentation with choice of bandwidth

To summarize, Table 4.1 lists the model assumption, advantages and disadvantages of the parametric, semi-parametric and kernel density estimation based non-parametric approaches of ROC curve fitting.

4.2.4. Multiple-Reader Multiple-Case ROC and Location ROC

One practical issue people often face in ROC analysis in a typical observer performance studies is how to deal with the variation of skill level among observers. Although the basic concept of ROC analysis had been understood in principle since the early 1980's [96], the available analytical techniques or research tools had limited practical applicability. This situation changed in the early 1990's with the introduction of the so-called Multiple-Reader Multiple-case (MRMC) ROC paradigm first published by Dorfman, Berbaum and Metz [22]. The MRMC ROC paradigms used a certain theoretical model and proposed applying procedures, such as jackknife method (jackknife readers or jackknife cases), bootstrap method [104] to the area under the ROC curve obtained for each reader, followed by use of the resulting pseudo values in a mixed-effects analysis of variance.

Another issue we need to deal with in ROC analysis is the localization of signals. In most clinical practice, a decision-maker usually does not have a precise knowledge of the location of a possible tumor, which leads to the effect of spatial uncertainty. As a result, some true positive responses may be due to the observer failing to detect the actual signal while at the same time falsely detecting a signal elsewhere in the image. C. Metz et al. introduced the

three-dimensional generalized ROC curve for a detection and localization task [56]. In this method, 5 types of decisions and associated conditional probabilities exist: “true positive, correct location”, “true positive, incorrect location”, “false positive”, “true negative”, and “false negative”. The degree of freedom for these 5 decisions is three because of the intrinsic property of a made decision. Thus, a particular combination of decision frequencies or probabilities resulting from a detection and localization experiment can be represented as a point in three-dimensional space. In practice, the conventional ROC curve and the LROC curve, which is introduced in separate paper [92], may be used to reconstruct the three-dimensional generalized ROC curve.

4.2.5. Summary Measures of the ROC Curve

After ROC curves are constructed, we need to statistically evaluate and understand the difference among the curves. Many methods are used for interpreting ROC curves and evaluating performance of different ROC curves [33, 95, 96]. The most popular method is the “Area Test” method in which the area under the curves is calculated and compared. It ranges from 0.5 for accuracy no better than a random binary chance to 1.0 for perfect accuracy. However, this global index can be misleading when curves crossover each other or there is only one ROI. Other ROC-based indices of performance are available, such as partial area below or to the right/left of a segment of the ROC curve, TPF at fixed FPF position or

vice-versa, or particular utility at optimal operating point, etc.

In the binormal model where FPR and TPR are fully defined by (4-6) and (4-7) and F is of the form of standard-normal cumulative distribution function Φ , the area under the ROC curve A_z is theoretically calculated to be

$$A_z = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right) \quad (4-13)$$

Furthermore, because of the intrinsic relationship between the estimation of a, b and the test data sample means (μ_{normal} and μ_{diseased}) and standard errors (σ_{normal} and σ_{diseased}) [55, 103], the area under curve can be estimated directly by substituting the μ and σ into

$$A_z = \Phi\left(\frac{\mu_1 - \mu_0}{\sqrt{\sigma_0^2 + \sigma_1^2}}\right) \quad (4-14)$$

In non-parametric model, the area under curve can be calculated analytically by numerical integration of the trapezoidal areas.

Other popular statistical indices used to describe ROC curves include “slope-related and intercept-related indices” and “partial area under the curve”. The slope-related and intercept-related indices are often used in Gaussian distribution model where the expected ROC points should follow a straight line when plotted on a binormal coordinate paper [95]. The partial area under the curve is usually used if only a particular range of test data on specificity and sensitivity graph is valid or is of interest [53].

4.3 ROC Analysis: Software Platform Development at Dartmouth

As addressed in 4.2, there are several free or commercial software algorithms involving parametric ROC curve fitting to different type of observer performance studies. There is also some other available research grade or commercial computer software involving one or several specific functionalities of ROC curve fitting, as summarized in Table 4.2.

Table 4.2 Summary of ROC Software

Software	Functionality	Developer
ROCFIT	Fit a single ROC using the binormal model	University of Chicago
INDROC	Tests difference between independent ROC estimates	University of Chicago
CORROC2	Test difference between correlated ROCs	University of Chicago
CLABROC	Test difference between correlated ROCs	University of Chicago
ROCKIT	Integrates and extends the 4 programs above	University of Chicago
PROPROC	Fits a single ROC using the "proper" binormal model	University of Chicago
LABMRMC	Does a jackknife-ANOVA test for difference in AUC (Data collected on continuous and/or discrete scale)	University of Chicago
RSCORE2	Fit a single ROC using the binormal model	University of Iowa
RSCORE4	Fit a single ROC using the binormal model	University of Iowa
MRMC	Jackknife-ANOVA test for difference in AUC (discrete scale only)	University of Iowa
Unnamed	Unified general model and maximum-likelihood procedure based ROC and LROC fitting program	Richard G. Swensson Karen H. Chan
Analyse-it	Statistical software add-in for Microsoft Excel with basic ROC analysis capability	Analyse-It Software, Ltd., U.K.
MedCalc	Commercial Statistical Program with basic ROC analysis capability	MedCalc Software Belgium
ROC-&-ROL	Non-Parametric Kernel Smoothing of ROC Curves	University of Rochester
S-Plus	Non-Parametric Kernel Smoothing of ROC Curves with different bandwidth choice methods	University of Rochester

Most of these software are coded in FORTRAN scientific programming language, which has limited graphical user interface and is not as user-friendly as most Microsoft Windows or Linux System based applications today. On the other hand, these software are separate mathematic tools with totally different I/O formats and result presentation styles, which make it hard for users switching from one program to another. This issue becomes more problematic when the complexity, i.e. the number of people and cases evolved, increases.

In order to implement our study of NIR tomography images with a generalized functionality program which could be expandable, a multifunctional MATLAB based program was developed and used in the ROC/LROC curve analysis tasks performed in this thesis. Binormal model based categorical and continuous data ROC/LROC fitting methods discussed in section 4.2.1 to 4.2.3 were embedded in this software platform. The general fitting and analysis algorithms are described in Figure 4.2. Some of the post-curve-fitting statistical measurements discussed in section 4.2.5, including area under the curve, standard error, data correlation, confidence intervals and confidence interval boundary estimations are also included. Figure 4.5 is a screen capture of the program user interface.

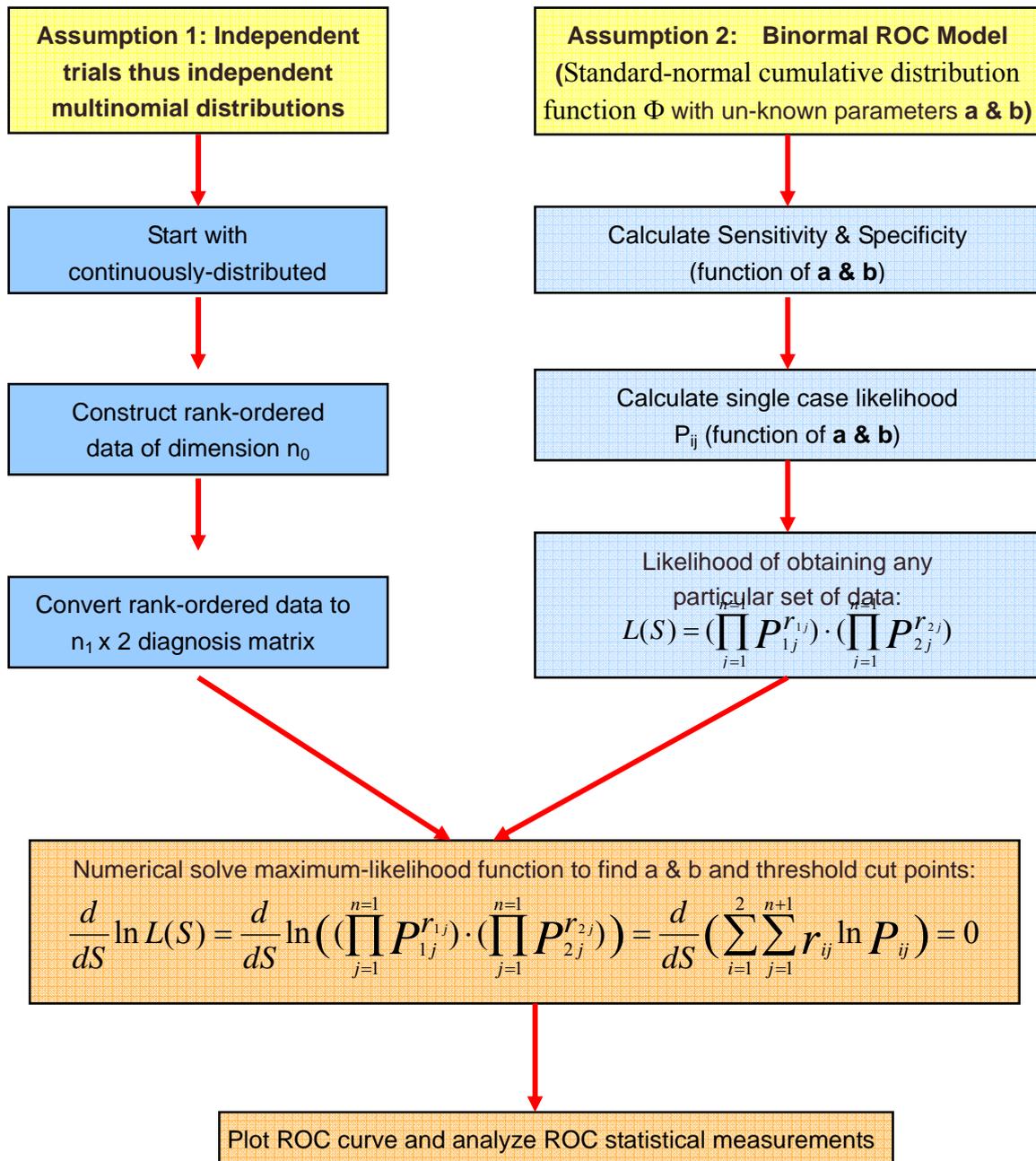


Figure 4.4 Procedures of Maximum-Likelihood Estimation Method Receiver Operating Characteristic Curve Fitting.

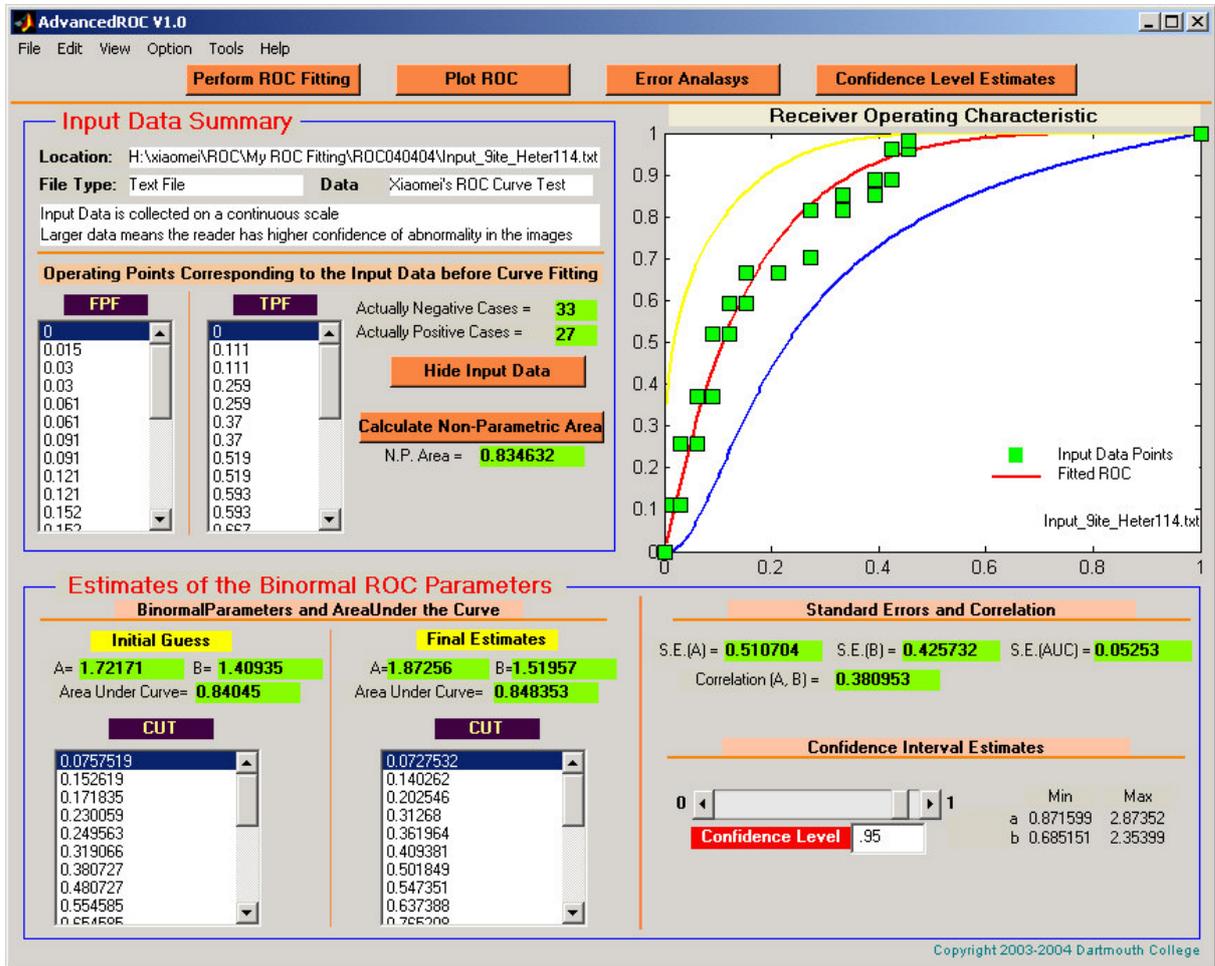


Figure 4.5 Screen Capture of the MATLAB GUI based ROC/LROC fitting platform, including: Main menu options, Input Data Information, ROC Analysis Output, Confidence Interval Boundary, Input Data And Fitted ROC Plots, Statistical Tests Results of Standard Errors, Data Correlation and Confidence Intervals.

4.4 Assessment of Near-Infrared Tomography Imaging System

In this section, Receiver operating characteristic (ROC) analysis based human observer and contrast-to-noise ratio (CNR) based computational assessment performance studies were performed and analyzed on tomography images produced at Dartmouth College near-infrared tomography imaging system. Parametric and nonparametric approaches have been applied for estimating the ROC curves. Computational CNR and human observers are then compared in the capability to accurately localize the heterogeneity under high heterogeneity contrast level and the results are discussed thereafter. The effects of iteration and algorithm on the detectability of objects in NIR tomography images for both human observers and in the computed CNR value are also investigated in the following sections.

4.4.1. Computer Simulations

4.4.1.1. NIR Tomography Simulation and Image Preparation

As discussed in chapter 1, near-infrared frequency domain absorption and scatter tomography modeling and image reconstruction has been partially developed based upon a Finite Element forward simulation of diffusion theory, and an iterative non-linear reconstruction algorithm [12, 94]. This system can produce 2D images of absorption and scatter coefficients of the objects examined by the non-invasive imaging array.

A circular 86mm diameter field was chosen as the background in which to place objects

and test the methodology based upon computed images. This field had absorption coefficient $\mu_a = 0.005 \text{ mm}^{-1}$, and a reduced scattering coefficient $\mu_s' = 1.0 \text{ mm}^{-1}$, and within the field a cylindrical object with fixed $\mu_s' = 1.0 \text{ mm}^{-1}$ and a variable μ_a was placed to provide a localized heterogeneity. The value of the anomaly was varied to simulate changes in absorption contrast, with varying values from 1.1:1 (10%) up to 2.0:1 (100%). The size of the anomaly ranged from 4mm to 16mm in diameter. Forward calculations based on this specific NIR tomography system were used along with zero-mean Gaussian noise of 1% in amplitude and 1 degree in phase shift, to create simulated measurement data. A non-linear reconstruction algorithm mentioned previously was then applied to generate the 2D absorption and scattering coefficient images. In the same manner, thousands of reconstructed images were automatically created, having the same noise level and size of heterogeneities but with different locations of the ROI and different contrasts. Similar images were created with no objects inside, to simulate normal tissue, or the "control" images.

We carefully chose the amount and constitution of the reconstructed image pool for several different studies presented here, and these contained images of different ROI (region of interest) size (4mm, 6mm, 10mm and 16mm) or different reconstruction algorithms, i.e. different iteration number of the reconstruction program or different contrast level between the ROI and the background.

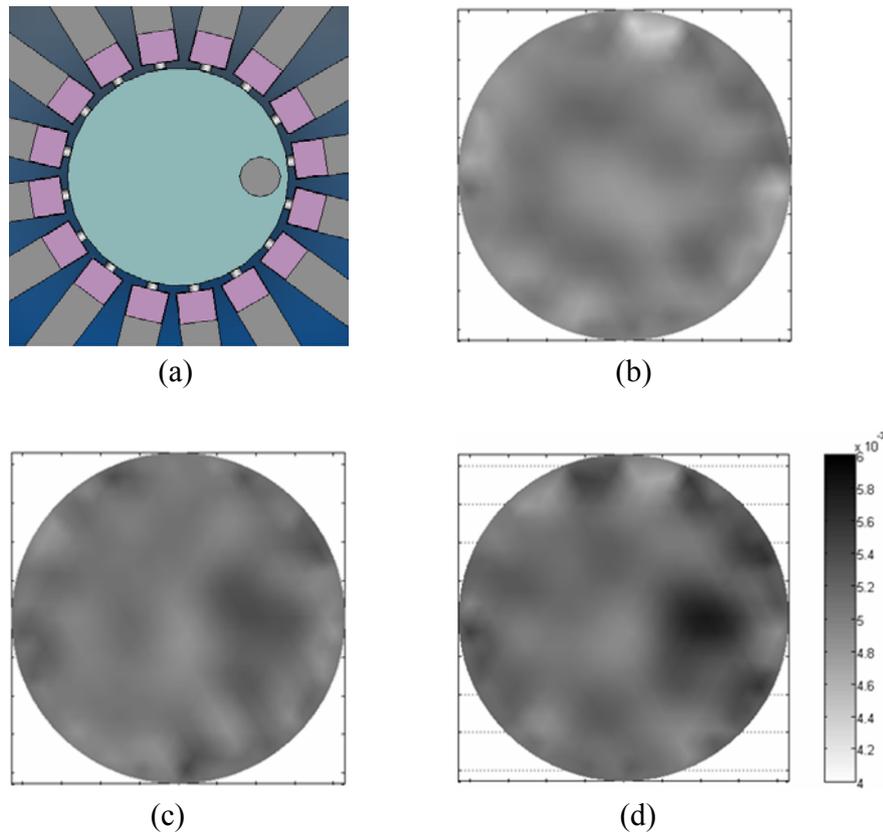


Figure 4.6 The NIR tomographic imaging system is shown along with a typical reconstructed image. (a) is the top view of the phantom and fiber interface; (c)-(d) are typical reconstructed NIR tomography images with the same size, and location of ROI, iteration number, and reconstruction algorithm but different contrast levels: (b) $C = 1.1$; (c) $C = 1.4$, and (d) $C = 2.0$.

4.4.1.2. Detection and Localization Task

Four human observers participated in this study of which 2 observers had NIR tomography imaging experience and the other 2 observers did not. A MATLAB program developed with a graphical user interface was used as the main test program, and is shown in Figure 4.6.

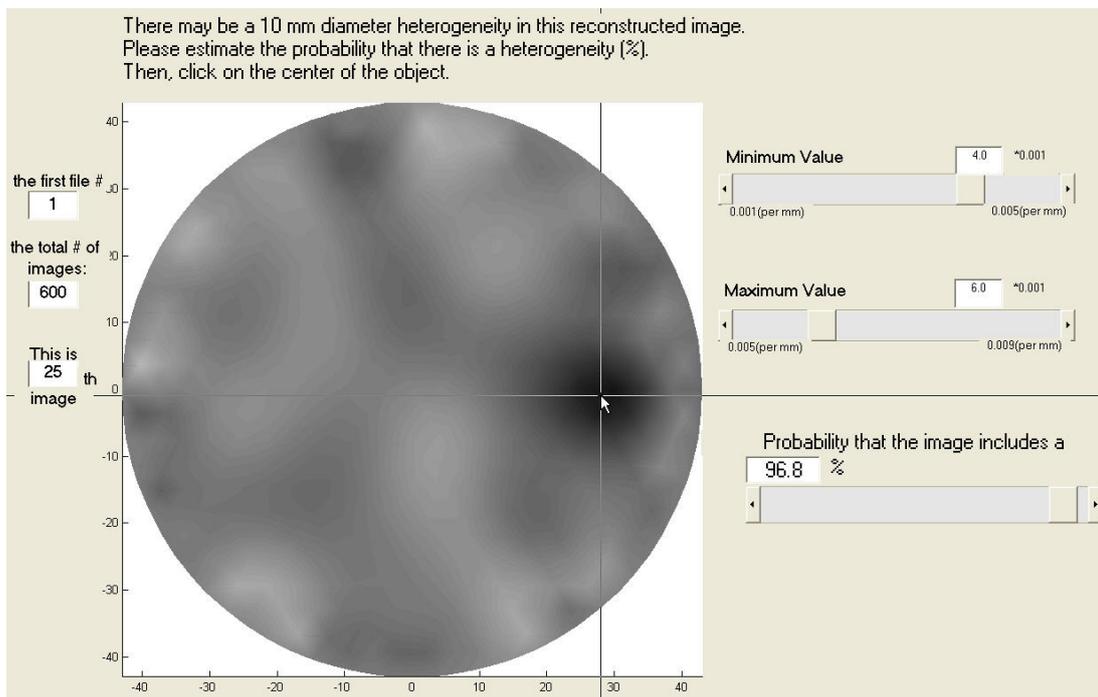


Figure 4.7: The MATLAB image detection program is shown. Observers can adjust the image gray scale map with the top two slide bars, and point out the estimated center location of the region with the computer mouse

This program controlled the data collection stage of the observer performance studies. Displaying a fixed number of reconstructed homogeneous or heterogeneous images in random order completed the ROC experiment. Initially a computer controlled training session is used in which the observer was presented sample images and information was asked of the observed as to whether the image contained a heterogeneity and the possible region size. The observers were asked to view 600 different reconstructed images in one ROC experiment session. Instead of just deciding whether the viewed image was homogeneous or heterogeneous, the observers were asked to rate a probability that the image was heterogeneous, using a continuous scale value between 0 and 1, where 0 represented

“absolutely homogeneous”, 1 represented “absolutely heterogeneous” and “0.5” was a guess of “it not being known”. The observer was also asked to point to a possible location of the ROI center, by clicking the centroid of the region with a cursor. The observer was able to control the image contrast and brightness by adjusting the minimum and maximum display parameters of the Matlab image profile. For every viewed image, the program recorded the history of all the information about the image and the decision of the observer.

4.4.1.3. Computational Assessment

To develop a fast computational metric, which may serve as a surrogate measure, an automated and efficient algorithm is needed. The Contrast-to-Noise Ratio (CNR) is the basis for the computational metric used throughout this study. Similar to the signal-to-noise ratio in digital signal processing theory, the CNR is defined as the relative difference between the ROI and the background region values of the property, divided by the average variation in the background [45, 81, 101]. There are different choices for the background [88], and here the entire region outside of the target ROI was used as the background, and thus the CNR is defined as [88]

$$CNR = \frac{\mu_{ROI} - \mu_{background}}{\sqrt{w_{ROI} \sigma_{ROI}^2 + w_{background} \sigma_{background}^2}} \quad (4-15)$$

where μ_{ROI} is the mean of the value of in the target; $\mu_{background}$ is the mean value over the variable background; σ_{ROI} and $\sigma_{background}$ the standard deviations of the target and the whole background areas, respectively; w_{ROI} and $w_{background}$ are the weights in the target and

background, which are defined as the fractional size of the target and background in the image field. Given the ROI size, CNR was calculated at each point within the domain and the maximum CNR and its correlated location was considered as the real ROI CNR and location. Using the images generated in Section 2.2 and given the real diameter and location of the target, we were able to calculate CNR as the absorption contrast and the ROI size were systematically varied.

4.4.1.4. Results

Four human observers were used, of which two have a background in medical imaging. Also, the computational CNR based ‘synthetic’ observer was used in the studies as a comparison. Human observer and computational observer performance are reported in terms of ROC and LROC curves and the related area-under-curve value with error correction. The result of human observers with or without medical imaging background were compared to each other and then compared with the CNR based computational observer. As discussed in Section 4.4.1.1, given a fixed system noise level, three key parameters are thought to determine the quality of a reconstructed image: (1) the size of the heterogeneities, (2) the absorption and scattering coefficients contrast between the heterogeneities and the homogeneous tissue, (3) the number of the reconstruction iterations used in the image formation and (4) the location of the region of interest (ROI) in heterogeneous images.. In other words, the detectable object size and object to background contrast level need to be determined in this approach, and the reconstruction process efficacy needs to be analyzed. In

this section, the results of a number of studies designed to evaluate the impacts of these three parameters are reported for a standard version of the NIR tomography reconstruction algorithm. All the images used in these studies were generated by the procedure described in Section 4.4.1.

In the LROC fitting process, observer performance information for reconstructed images consists of two parts: (1) a continuous confidence rating of possible heterogeneity presence and (2) coordinates representing the possible object location. The one-stage “plug-in” method with the Epanechnikov kernel function [106] was used to generate the distribution bandwidth, and then applied for fitting of the LROC curve, as illustrated in Figure 4.8.

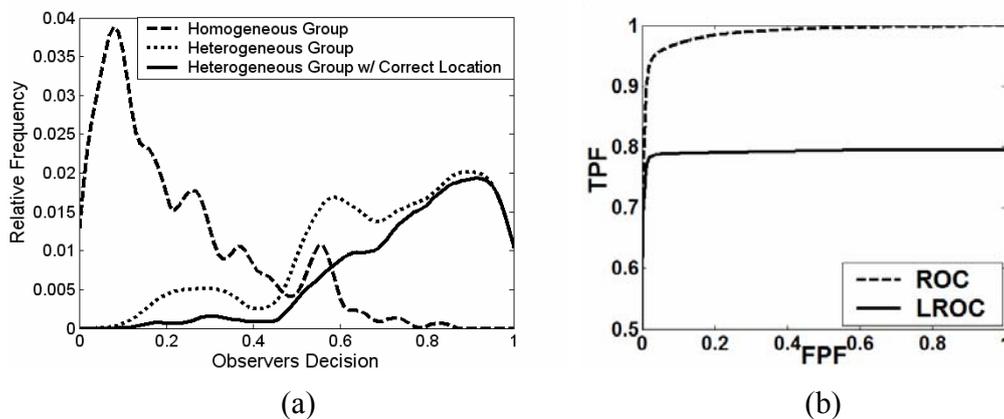


Figure 4.8 (a) The distribution sample of human observer response is shown, which was fitted by kernel density estimation. The window width was chosen from the direct ‘plug-in’ method and the Epanechnikov kernel function was used (b) ROC and LROC curve for (a)

A. Human Observers Performance Comparisons

In trying to minimize the observer background factor, a training session was used for all

participants prior to the final image presentation and detection process. In the training session, the trainee was asked to review sample images with the same reconstruction condition and was notified if the images contained heterogeneity and if so, what the location of the heterogeneity was. The trainee is allowed to review and analyze as many as 200 images before the real detection task began. Table 4.3 summarizes the area under curve value and associated standard error for both ROC and LROC curves for different observers in the three heterogeneity-size studies.

Table 4.3 Human Observers Performance Comparison

Observer (note)	ROC Area under Curve with Error Correction			Localized ROC Area under Curve with Error Correction		
	ROI=6mm	ROI=10mm	ROI=16mm	ROI=6mm	ROI=10mm	ROI=16mm
#1	0.90 ± 0.012	0.99 ± 0.008	0.99 ± 0.003	0.23 ± 0.026	0.71 ± 0.023	0.90 ± 0.013
#2	0.89 ± 0.015	0.98 ± 0.009	0.99 ± 0.004	0.23 ± 0.026	0.69 ± 0.024	0.89 ± 0.014
#3	0.88 ± 0.014	0.99 ± 0.005	1.00 ± 0.001	0.23 ± 0.026	0.70 ± 0.024	0.91 ± 0.013
#4	0.93 ± 0.009	0.99 ± 0.003	0.96 ± 0.008	0.25 ± 0.028	0.68 ± 0.025	0.89 ± 0.014

Note: #1 & #2: Experienced Observer, #3 & #4: Inexperienced Observer

B. Human Observer versus Computed CNR Performance Comparison

a) Heterogeneity Size Study

This study is designed to find out how response of the NIR tomography imaging reconstruction algorithm response to different sized heterogeneities would affect the human perception of objects. For the sake of simplicity, a circular geometry was chosen for the heterogeneities with a diameter ranging from 4mm to 16mm, compared to an 86mm diameter region for the background homogeneous tissue. Table 4.4 summarizes the relevant

parameters of 4 observer performance studies, of which each study contained 600 images with optical property ranges.

Table 4.4 Parameter Settings in the Heterogeneity Size and Contrast Studies (Iteration number = 6)

Study #	Noise Level	Heterogeneity Properties		Background Properties
		$\mu_a (10^{-3}\text{mm}^{-1})$	Diameter (mm)	$\mu_a (10^{-3}\text{mm}^{-1})$
1	1% amplitude & 1° phase shift	5.5-10.0	4	5.0
2	1% amplitude & 1° phase shift	5.5-10.0	6	5.0
3	1% amplitude & 1° phase shift	5.5-10.0	10	5.0
4	1% amplitude & 1° phase shift	5.5-10.0	16	5.0

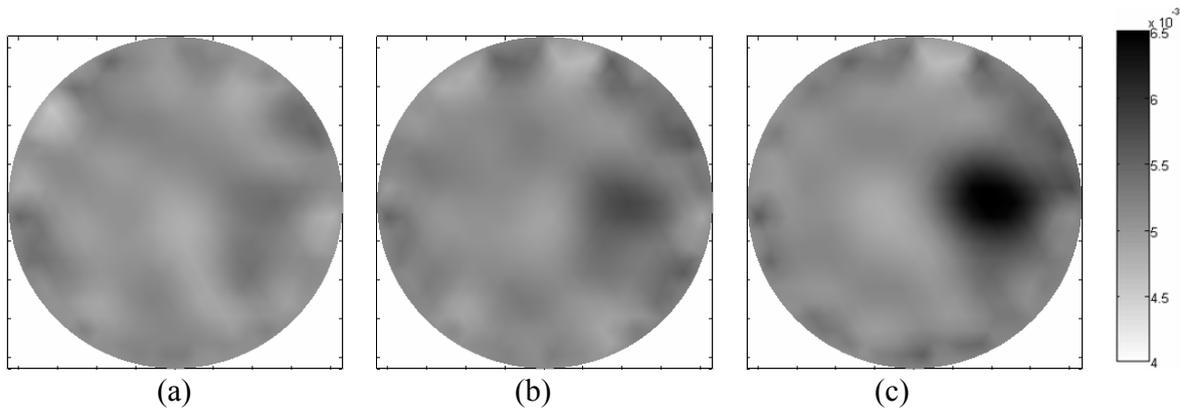


Figure 4.9 Sample images from heterogeneity size studies are shown. The objects all have the same contrast $C = 2.0$, the same reconstruction process iteration number $I = 6$, but with different object sizes. In (a) a small sized object was used, with diameter of 6mm. In (b) a medium sized object was used, with diameter of 10mm. In (c) a large object of diameter 16mm was used.

To maintain experimental consistency, zero-mean Gaussian noise of 1% amplitude and 1 degree in phase shift were added to the calculated boundary data of all studies to simulate a realistic data set from our NIR tomography system. Sample images from heterogeneity size studies, in which objects were sized from 4mm to 16mm and the contrast level defined in the

next section is ranged from 1.1 to 2.0, are shown in Figure 4.9.

Figure 4.10 summarizes the human observer and computed CNR performance in the heterogeneity size studies. From Figure 4.10(a), we can see that when the diameter of the target is less or equal 6mm, the computational observer cannot distinguish the heterogeneous images from the homogeneous images as the AUC (Area under the Curve) of ROC is equal to 0.5. However, human observers were able to sort out the heterogeneous images of 6mm or less from the homogeneous images since the AUC value is well above 0.5 (0.82 for size 4mm and 0.88 for size 6mm). As the size of the targets increase, the computational observer diagnosis accuracy becomes much better because the AUC of ROC increases under the same condition. On the other hand, as shown in Figure 6(b), if we take the localization accuracy factor into account and set “AUC of LROC equal 0.5” as the threshold to determine whether the observers can choose the correct locations of the heterogeneities, neither computational CNR nor human observer can accurately localize the heterogeneity of size 6mm or less in the reconstructed tomography images.

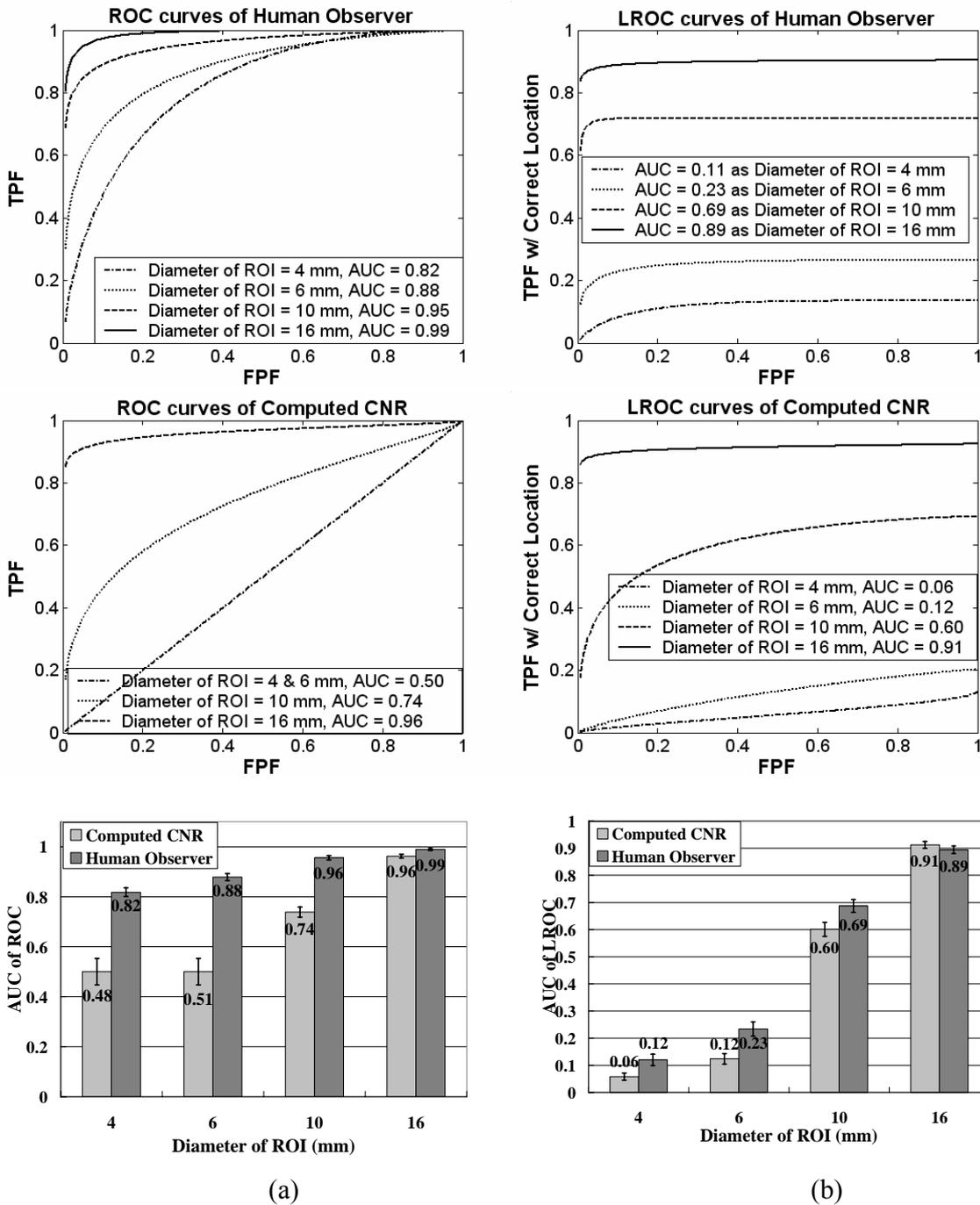


Figure 4.10 The heterogeneity size study for both human and computational observers is shown in (a) with ROC and area under curve data, and in (b) with LROC and area under curve data. Both of these are shown for data sets where the absorption contrast ranged from 1.1 to 2.0, with a fixed 6 iterations.

b) Heterogeneity Contrast Level Study

The optical property contrast (C) between the object and background was provided by the absorption coefficient difference between the heterogeneity and the background. The heterogeneity to background contrast is defined as

$$C = \frac{\mu_{ROI}}{\mu_{Background}} \quad (4-16)$$

where μ_{ROI} is the mean absorption coefficient within the target and $\mu_{background}$ is the mean absorption coefficient value over the variable background. Using fixed object diameter $D = 10\text{mm}$ and reconstruction process iteration number $I = 6$ and varied contrast with increment step of 0.1, we generated 30 images for each contrast value ranged from 1.1 to 2.0 giving 300 images in total. Same human and computational observers are used in these 10 studies. An example image from $C = 1.4, 1.7$ and 2.0 is shown in Figure 4.11.

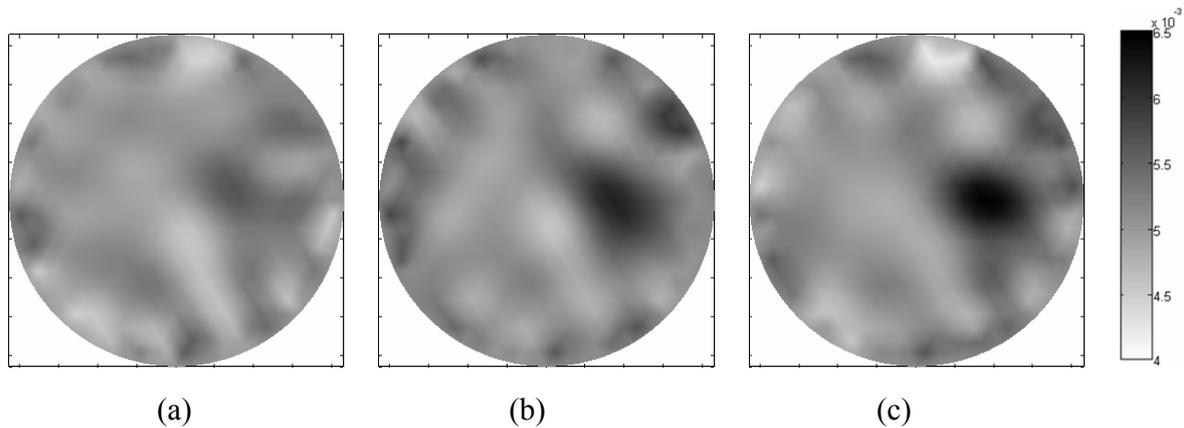


Figure 4.11 Sample images from heterogeneity contrast studies are shown. The objects had same size, with diameter of 10mm, same reconstruction process with iteration number $I = 6$ but with different contrasts. In (a) the contrast was $C = 1.4$; in (b) the contrast was $C = 1.7$, and in (c) the contrast was $C = 2.0$.

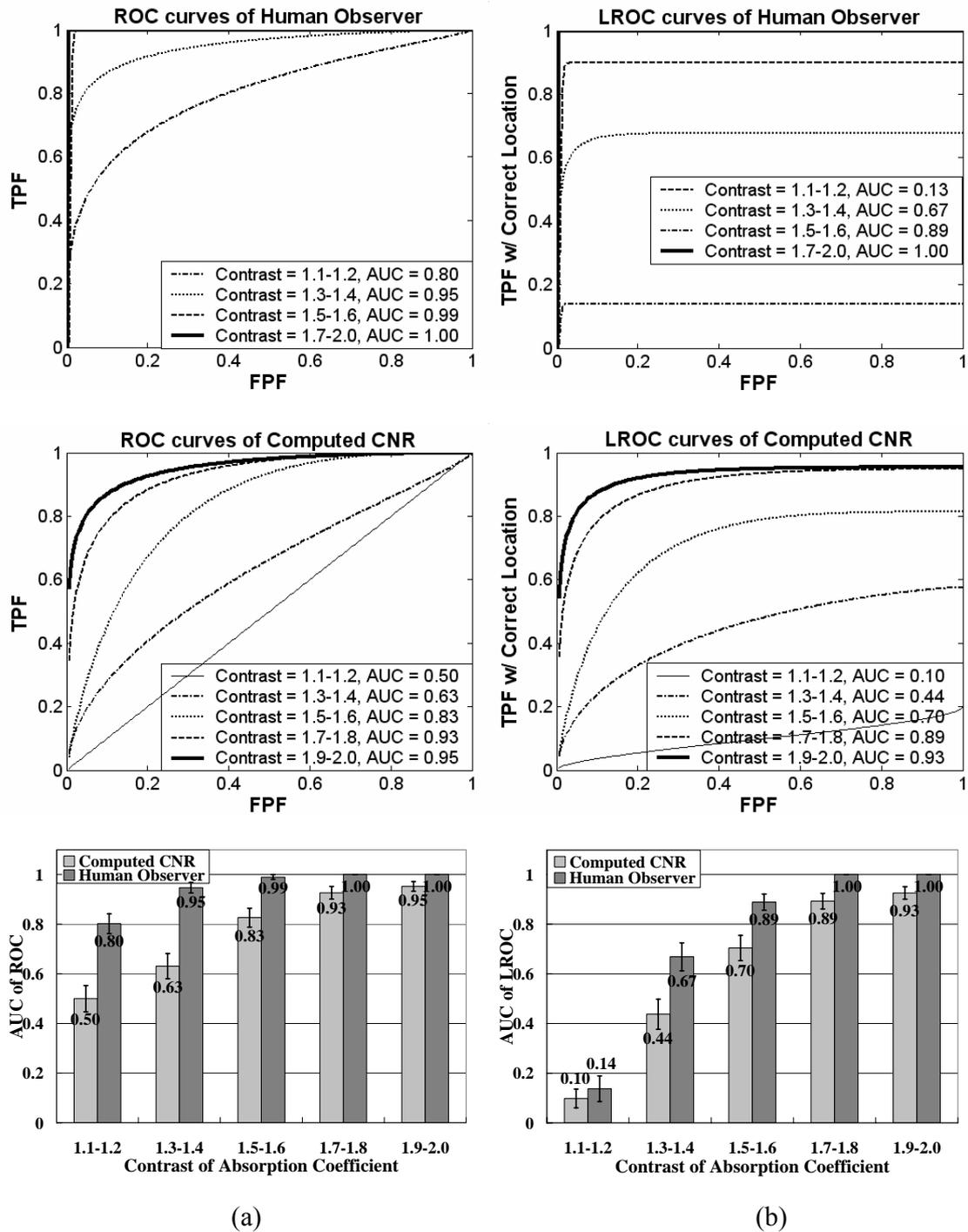


Figure 4.12 The heterogeneity contrast study for both human and computational observers is shown, with (a) showing the ROC and area under curve, and (b) showing the LROC and area under curve. The diameter of ROI was equal 10 mm for all, and 6 iterations were used.

Figure 4.12 summarizes the human observer and computed CNR performance in the heterogeneity contrast level studies for un-localized (Figure 4.12(a)) and localized (Figure 4.12(b)) detection tasks. In Figure 4.12(a), where the ROI size was fixed at 10mm, the human observer performed well even at low contrast levels and achieved perfect detection (AUC = 1.0) performance when the contrast level was above 1.6. The computed CNR on the other hand, was unable to achieve perfect detection at any tested contrast level but was very close to perfect when the contrast level became higher. Human observers have more capability to detect heterogeneities with small contrast levels, yet this difference was negligible if the heterogeneity was at higher contrast levels compared to the homogenous background.

c) Analysis of the Reconstruction Process Iteration

The most dominant factor in the reconstruction process which affects image quality is the regularization parameter and the number of iterations used in formation. In our studies we have always utilized a modified Levenberg-Marquardt scheme in which the regularization parameter starts out a factor of 100 relative to the normalized Hessian diagonal, and then is systematically decreased by a factor of 2 at each successive iteration [42, 49]. Using this approach, the iteration number is the major factor influencing detection of objects, and this has been studied here. The outline of the iteration study is summarized in Table 4.5, where 600 images were used in total. As mentioned in the previous section, a Finite Element Method based reconstruction program written in Matlab is used in the near infrared frequency domain absorption and scattering tomography imaging technique. During a single

reconstruction process step i.e. linear reconstruct, the 2D image was first constructed and used as the input data for the forward problem. The output of the forward problem, also known as the calculated system measurement, was compared to the measured data and thereafter to determine the quality of the reconstruction process.

Table 4.5 Parameter Settings in the Reconstruction Process Iteration Number Study

Study #	Noise Level	Heterogeneity Properties		Iteration Number
		Contrast	Diameter (mm)	
1	1% amplitude & 1° phase shift	1.1-2.0	10	4
2	1% amplitude & 1° phase shift	1.1-2.0	10	6
3	1% amplitude & 1° phase shift	1.1-2.0	10	8
4	1% amplitude & 1° phase shift	1.1-2.0	10	10

An important measure of error in the reconstruction process is the objective function defined as:

$$\chi_k^2 = \frac{\|\Phi_M - \Phi_C(\mu^k)\|^2}{M \sigma^2} \quad (4-17)$$

where Φ_M is the measured set of data, $\Phi_C(\mu^k)$ is the calculated measurements at the k^{th} iteration for optical property μ , M is the number of measurements and σ is the entire image standard deviation. This is also sometimes referred to as the projection error (or Chi squared) function, as it is a direct measurement of the squared error of measured and simulated projection data through the tissue.

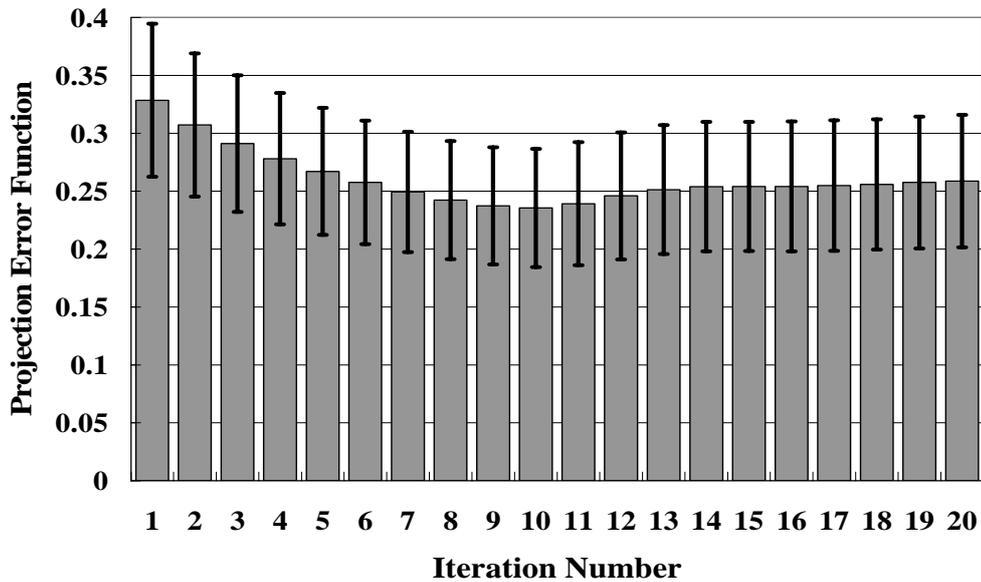


Figure 4.13 The mean of the projection error function of 294 heterogeneous reconstructed images is shown as the function of iteration number, using a diameter ROI of 10 mm and an absorption contrast range from 1.1 to 2.0. The error bars are the standard deviations of the 294 images for all of 20 iterations.

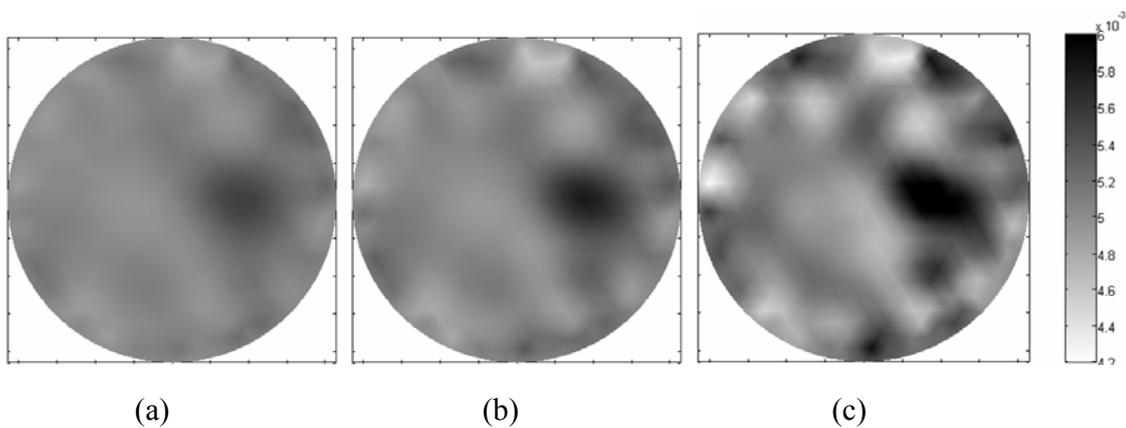


Figure 4.14 Sample images from the reconstruction process are shown for different iteration numbers. The images have same object size, $D = 10\text{mm}$, the same object-to-background contrast of $C = 2.0$, but are from different numbers of iteration in the algorithm. In (a) the result from 4 iterations is shown. In (b) the result from 6 iterations is shown, and in (c) the result from 10 iterations is shown. Higher numbers of iterations tend to recover higher spatial frequency information and higher contrast, albeit with the addition of a non-linear level of noise increased.

In practice, as the iteration number increases, the objective function first decreased and reached a lowest point, and then increased very slowly, as shown in Figure 4.13 where the ROI was 10mm diameter and the contrast ranged from 1.1 to 2.0. The error was estimated from a number of reconstructions, and the mean error is plotted as a function of iteration number. In Figure 4.13 the minimum objective function value occurs at an iteration number of 9. Theoretically the reconstructed image parameter value should be closest to the true image value at the point where the objective function is at its lowest value, but in practice the reconstruction iteration process also inevitably introduces some high frequency noise so that it affects the judgment of both the human observers and the computational observer. A sample image from iteration numbers of 4, 6, 8 and 10 is shown in Figure 4.14.

Figure 4.15 summarizes the human observer and computational CNR value performance in the iteration number studies for un-localized (Figure 4.15(a)) and localized (Figure 4.15(b)) detection tasks. From Figure 4.15 where the ROI size was fixed at 10 mm and the contrast level was varied from 1.1 to 2.0, it can be seen that the human observer always has better performance than a computational observer, in terms of detectability and localization accuracy. In the range of 4 to 10 iterations, the iteration number has either negligible impact (human observer data in Figure 4.15(a)) or negative impact (computed CNR data in Figure 11(a) and all Figure 4.15(b)) on both human observer and CNR detection and localization decisions.

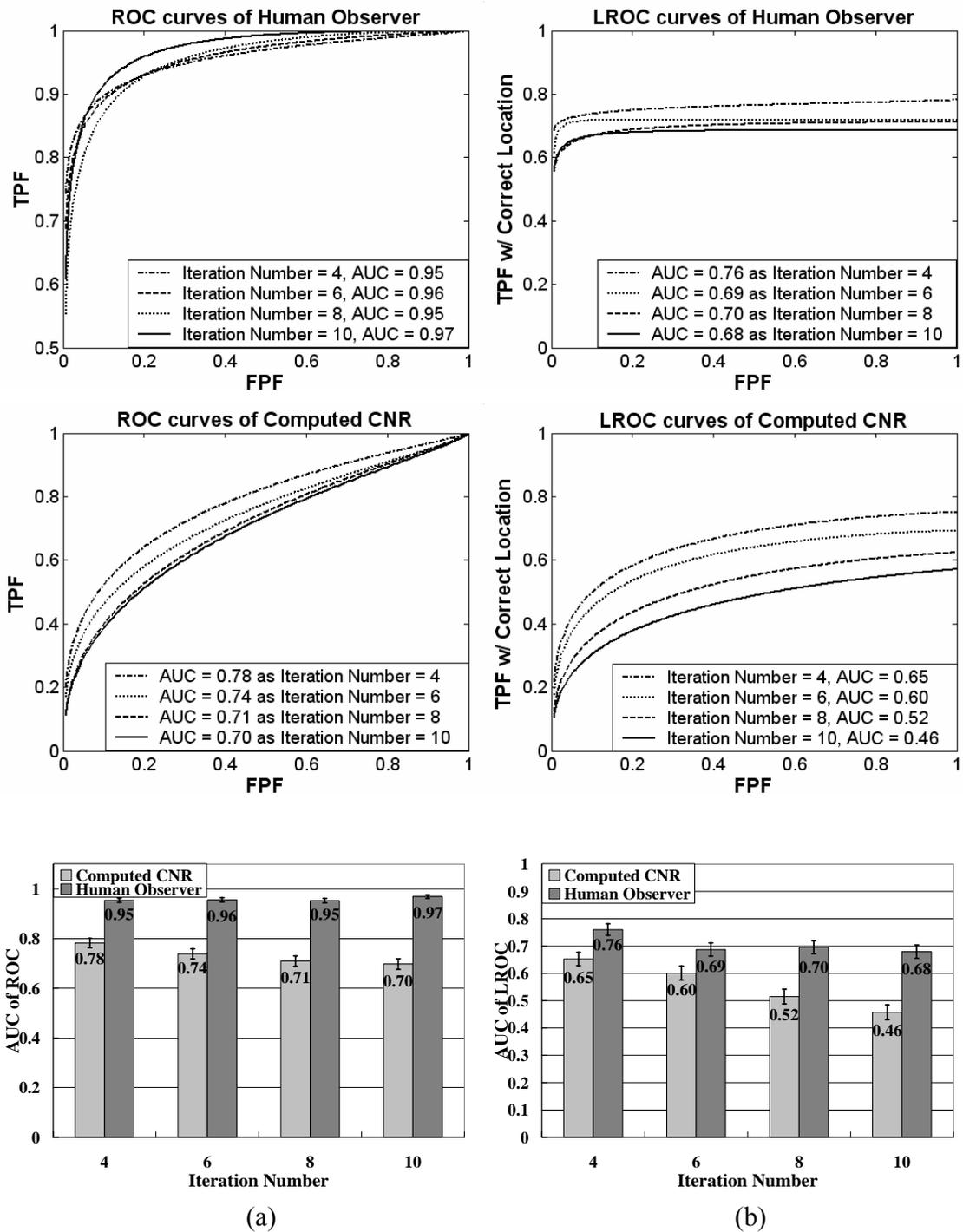


Figure 4.15 The reconstruction process iteration number was studied for both human and computational observers. In (a) the ROC and area under curve are shown, and in (b) the LROC and area under curve are shown, for a fixed diameter equal to 10 mm and with varying absorption contrast from 1.1 to 2.0.

d) Analysis of the Location of ROI

In our prior studies, it shows that NIR tomographic images are spatially dependent [69]. As the NIR light diffuses in the phantoms or tissue, the location close to the edge has higher spatial resolution than that at the center. Meanwhile, the reconstructed contrast becomes larger as the target gets closer to the edge; but the noise from the closest light source and detector have the most impact on the overall image quality. Thus, it is necessary to investigate this phenomenon and find out how the heterogeneity position affects the judgment of human and computational observers. Setting the diameter of ROI at $D = 6\text{mm}$, we generated 43 images for each ROI location ranging from 5mm to 35 mm to the background center and 300 in total, as shown in Table 4.6(a).

Table 4.6a Parameter Settings in the Heterogeneity Location Study (Iteration number = 6)

Study #	Noise Level	Heterogeneity Properties		Distance between the Centers of the ROI & Background (mm)
		Contrast	Diameter (mm)	
1	1% amplitude & 1° phase shift	1.1-2.0	6	5
2	1% amplitude & 1° phase shift	1.1-2.0	6	10
3	1% amplitude & 1° phase shift	1.1-2.0	6	15
4	1% amplitude & 1° phase shift	1.1-2.0	6	20
5	1% amplitude & 1° phase shift	1.1-2.0	6	25
6	1% amplitude & 1° phase shift	1.1-2.0	6	30
7	1% amplitude & 1° phase shift	1.1-2.0	6	35

In the same manner, for ROI diameter $D = 10\text{mm}$, we generate 50 images for each ROI location ranging from 11mm to 31 mm to the background center and 300 in total, as shown in Table 4.6(b).

Table 4.6b Parameter Settings in the Heterogeneity Location Study (Iteration number = 6)

Study #	Noise Level	Heterogeneity Properties		Distance between the Centers of the ROI & Background (mm)
		Contrast	Diameter (mm)	
1	1% amplitude & 1° phase shift	1.1-2.0	10	11
2	1% amplitude & 1° phase shift	1.1-2.0	10	15
3	1% amplitude & 1° phase shift	1.1-2.0	10	19
4	1% amplitude & 1° phase shift	1.1-2.0	10	23
5	1% amplitude & 1° phase shift	1.1-2.0	10	27
6	1% amplitude & 1° phase shift	1.1-2.0	10	31

The reconstruction process iteration number I is set to 6 and the contrast is set to range from 1.1 to 2.0. Same human and computational observers are used in all of the studies. Sample images from Diameter of ROI = 6 and 10mm are shown in Figure 4.16.

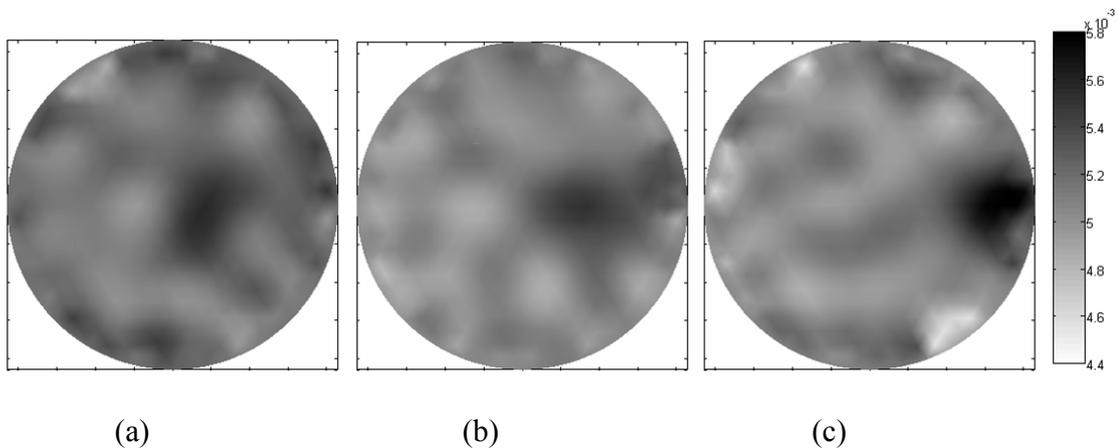


Figure 4.16 Sample images from heterogeneity location studies are shown. The objects had same size, with diameter of 10mm, same reconstruction process with iteration number $I = 6$, and with the same contrast $C = 1.7$, but at different locations. In (a) the distance between the object and phantom centers was $D = 11\text{mm}$; in (b) the distance between the object and phantom centers was $D = 21\text{mm}$, and in (c) the distance between the edges of the object and phantom was $D = 6\text{mm}$.

Figure 4.17 summarizes the human observer and computational CNR value performance in the target position studies for un-localized (Figure 4.17(a)) and localized (Figure 4.17(b)) detection tasks, the error bars are the standard deviation of the four human observers' performance. When the ROI size was fixed at 6 mm and the contrast level was varied from 1.1 to 2.0, the mean AUC of ROC for human observers was around 0.9, and had no significant difference at the different target locations, but all of the AUC for the lower 95% confidence interval band from computed CNR performance were less than 0.5, as shown in Fig 4.17(a).

In the meanwhile, the mean AUC of LROC for human observers is increased significantly as the target moved to the just off-center to the midway of the whole background, but the AUC of LROC for computational observers increased at the edge area of

the background, but all of the values are less than 0.5. As the ROC size increases to 10mm, the mean AUC of ROC for human observers and computational observers do not change much, except the AUC of ROC for computational at the edge area of the background, which is obviously larger than that at other position. Also, the AUC of ROC by the computational observer increases as the target moves toward the edge.

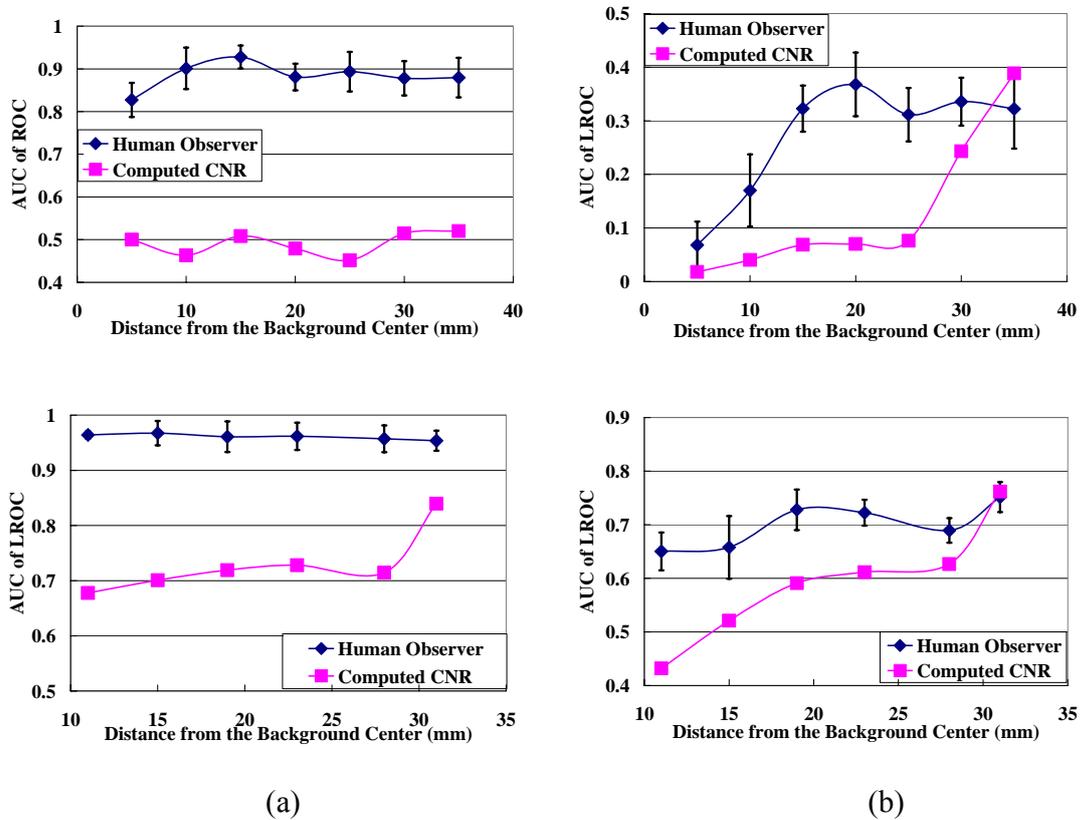


Figure 4.17 The heterogeneity location study for both human and computational observers. In (a) the AUC of ROC are shown, and in (b) the AUC of LROC are shown, for diameter of ROI equal to 6 (top pair of graphs) and 10 mm (bottom pair of graphs) and with varying absorption contrast from 1.1 to 2.0.

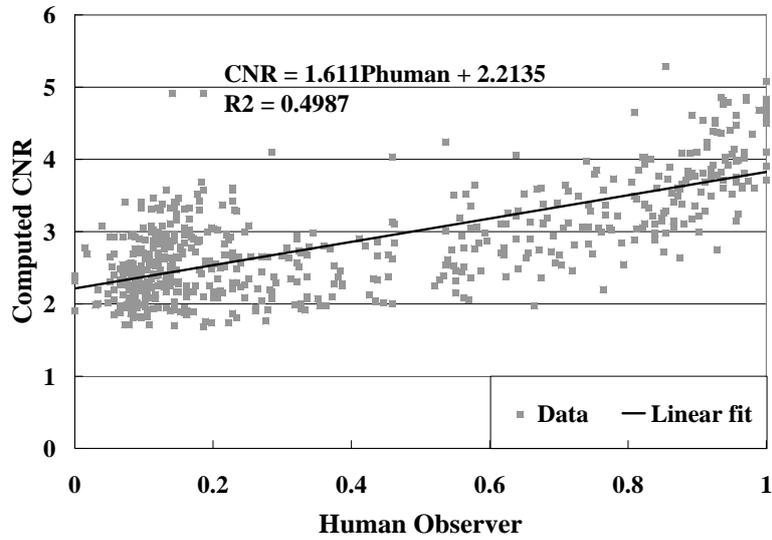
4.4.1.5 Computer Simulation Conclusions

A. Human Factor

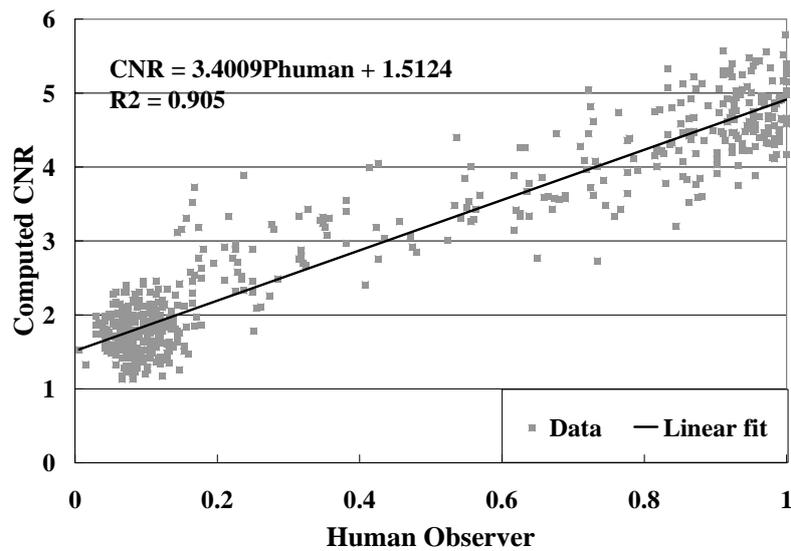
In the Human Observer Performance Comparison study, we plotted ROC and LROC curves of which the area under curve was summarized in Table 4.3. As we know, the area under the curve of an ROC and LROC curve is a reasonable indicator of the quality of the observer performance. Given the fact that observer #1 and #2 had NIR medical image interpretation experience and observers #3 and #4 were inexperienced, the main observation from Table 4.3, was that after the training procedure, readers without medical imaging experience had negligible performance difference compared to readers with medical imaging experience. We saw this trend also when tested with different heterogeneity contrast levels and when the reconstruction process iteration number varied, which makes it reasonable that all the four human readers' experimental data was mixed together in successive studies, and averaged data was used to in the human and computational observer comparison studies.

B. Human Observer Decision to Computed CNR Correlation

In the human observer detection process, readers were required to continuously rate the possibility of heterogeneity being present in the images. In the computational observer detection process, CNR value was calculated at each point within the model and the maximum value was considered in the “decision” or “rating” of the computational observer



(a)



(b)

Figure 4.18 Human observers and computed CNR performance comparisons are shown for paired sets of images analyzed. The human observer decision (confidence level that the image is abnormal) and computational observer decision (contrast-to-noise ratio) for same group of images are plotted in CNR-Human decision 2-D diagram. A linear fitting algorithm was then applied and the R^2 values are present for the two situations of (a) where contrast was set to 2.0, iterations set at 6 and object size set at 10 mm. In (b) the contrast was still 2.0, iterations were the same, but the object size was increased to 16 mm diameter.

for the NIR tomography images. Both the human observer and computational observer decisions data are plotted as scatter pairs for two image groups, including where the ROI size was 10 mm and where it was 16 mm. These are shown in Figure 4.18.

The solid line in Figure 4.18 is a linear fit of human observer and CNR values, of which the fitted equation and R square value are also listed in the diagram. From Figure 4.18, the main observation is that the CNR of the reconstructed images and the human observer responses is more correlated when the diameters of the targets increase.

C. Analysis of Human Observer versus Computed CNR Decisions

In the former section, we presented the results of heterogeneity size and contrast level study; we also discussed the result from the analysis of the reconstruction process iteration. We have seen clear difference between human observer and computational observer. When presented a suspicious NIR tomography image, human and computational observers both tend to search the area with highest contrast to noise ratio and make their decision based on the information acquired. However, there are apparently different processes of decision making between human and computer observers on different NIR tomography images.

When the object size is as small as 6 mm, the existence of object will change the diffused light throughout the entire field but its impact is not significant enough to be comparable to impact of the added noise. Though the object is not big enough to be seen on the reconstructed image at the object location, it makes the image seems noisier than the average homogenous image. This entire image profile change is detectable by human observer whose

strategy and overall ability are more sensible to noise appearance but not detectable by computational observer whose detect capability solely relies on relative contrast to noise ratio across whole image. As the heterogeneity size increased to around 10 mm, it dominates the whole image profile and diffusion process; as a result both computational and human observers get good detect capability and localization veracity.

On the other hand, contrast level of objects play different role in human and computational observers' decision making. When taking account of the location information, as summarized in Figure 4.12(b), the AUC value of the human observer LROC dramatically decreased at very small contrast levels, and since the corresponding ROC AUC values in Figure 4.12(a) were fairly good, it suggests that the human observer is capable of sorting out the abnormal images but is unable to specify the correct location of the heterogeneities. Besides the conceptual and intrinsic difference between ROC and LROC, this phenomenon is not apparent for the computational observers whose position and AUC value didn't dramatically change from ROC to LROC plots. This latter point indicates that the computational decision of tumor presentation in a NIR tomography images, would also provide a reasonably good tumor location at same confidence level.

Moreover, the target location will not affect the decision of human observers to distinguish the homogeneous and heterogeneous images obviously, although the mean AUC values of ROC for both 6mm and 10mm ROI sizes, where the object at the midway between the background center and edge, is a little bit larger than that at the other places in the image. Since the AUC of LROC for both human and computational observers is less than 0.5 at

everywhere in the images, we can consider that neither the human observers nor computational observer could find the correct target location as the ROI is equal than 6mm. Meanwhile, as the object size is increased to 10mm, the computational observer can get better and better decision and location accuracy, when the target moves to the image edge.

4.4.2. Phantom Studies

4.4.2.1. Phantom Studies Design and Image Preparation

Besides computer simulations method described in previous section, phantom studies provide a more practical way to evaluate the human decision with or without computer assistance to the NIR tomographic image. In this Chapter Three resin solid phantoms are used to mimic the physical and optical properties of the human breast. The phantoms are with optical properties similar to breast tissue, having absorption coefficients from 0.0033 to 0.0041 mm^{-1} , and reduced scattering coefficients from 1.16 to 1.31 mm^{-1} , and the cylindrical solid phantom had diameter of 84.5 ± 0.2 mm, as shown in Table 4.7. A hole was cut in each of the phantoms at different relative locations, and filled with Intralipid and ink to produce absorption coefficient contrast in the phantom with a closely matched scattering coefficient to the background. As illustrated in Figure 4.19, the hole in Phantom #1 was just off-center of the phantom, where the distance between the centers of the hole and the phantom was only 3mm; the hole in Phantom #2 is in the midway of the phantom, where the distance between the centers of the hole and the phantom is 18mm; and the hole in Phantom #3 is more closer

to the phantom edge, where the distance between the edges of the hole and the phantom was 6mm. The hole cannot be too close to the phantom edge because of the limitation of the diffuse equation we used in our reconstructed algorithm. Therefore, we could avoid the limitation of the object location when we did the phantom studies, which include the contrast and location analysis.

Table 4.7 Parameters of Resin Solid phantoms for Phantom Studies

Study #	Heterogeneity Properties			The Hole Location in the Phantom
	μ_a (10^{-3}mm^{-1})	μ_s (mm^{-1})	Diameter (mm)	
1	0.0041	1.31	84.6	just off center
2	0.0033	1.27	84.3	in the midway
3	0.0040	1.16	84.4	closed to the edge

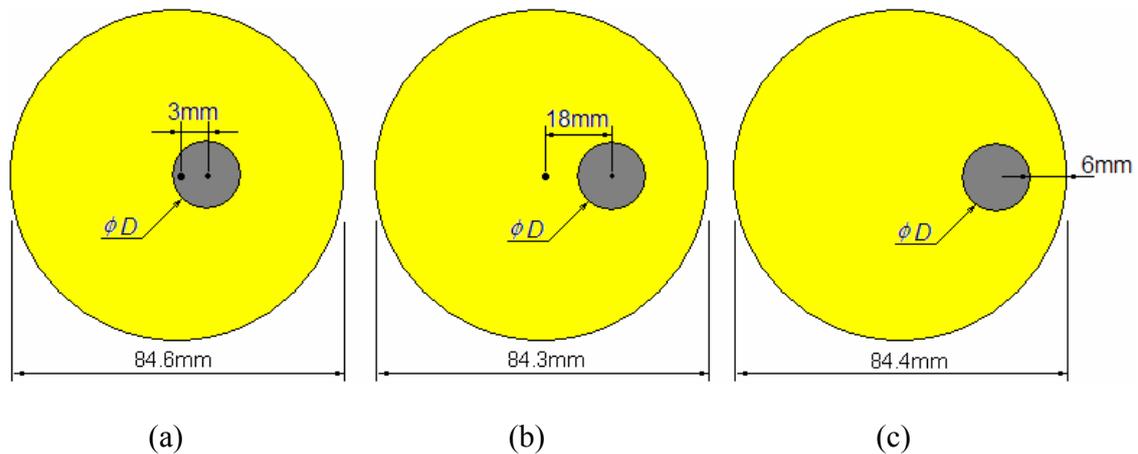


Figure 4.19. The geometries of the phantoms used in the studies for a circular object in solid phantoms whose size slightly varies. The diameters of the heterogeneity in all the 3 phantoms are 6 mm. (a) Phantom #1: $d = 84.6$ mm and the absorption and scattering coefficients of homogeneous area are 0.0041 and 1.31 (b) Phantom #2: $d = 84.3$ mm and the absorption and scattering coefficients of homogeneous area are 0.0033 and 1.16 (c) Phantom #3: $d = 84.4$ mm and the absorption and scattering coefficients of homogeneous area are 0.0040 and 1.03.

We filled the phantom holes with Intralipid and ink at different ingredient ratio to achieve controlled absorption contrast levels. Five contrast levels were used in this study: 1:1, 1.2:1, 1.5:1, and 2.0:1. For the purpose of extracting out hole edge effect, the phantom with 1:1 contrast was used as a homogeneous phantom and the other 4 phantoms were test as heterogeneous phantoms. The experimental data were reconstructed by the same diffuse imaging reconstruction method we used in the simulation study. And the iteration number is 6 for all of the reconstructed images.

In human observer test of this study, we used the same Matlab GUI program used in the simulation section, as shown in Figure 4.7. The four human observers who read images in simulation studies were also asked to read and rate the heterogeneous level of phantom study images. They were also asked to point out the possible ROI location in the Matlab test program, as we introduced in previous part of the thesis. Instead of 600 images they viewed in simulation part, each human observer viewed 228 images in phantom study.

4.4.2.2. Results and Discussion with a 6mm ROI

Table 4.8 indicates the AUC of both ROC and LROC curves for different human observers in the heterogeneity-contrast studies where the ROI is 6mm in diameter. The standard deviation of these four observer performance is larger than that in the simulation studies at each absorption contrast level. Thus, instead of mixing all of the data from the human observers together in the simulation part, we needed to analyze the experiment data

separately, or take the mean value as the average human performance.

Moreover, the experiment noise consists of both random noise and systematic noise, which is higher than the 1% random noise added to the calculated boundary data in our simulation [88, 89]. As the results, the reconstructed image are noisier in phantom study, so the human performance are much worse than that in simulation at the same ROI size and contrast level. When contrast is very low ($C = 1.2$), the mean AUC of ROC is less than 0.60, which means that the human observers could not distinguish the heterogeneity and homogeneity. Then, as the contrast increased to 1.5, the mean AUC is equal 0.75, which is fair to separate the heterogeneous images from homogenous images in typical ROC analysis. When the absorption coefficient increased to 2.0, the AUC of ROC reached the maximum value.

Table 4.8 Human Observers Performance Comparison with Different Contrast (C)

Observer (note)	AUC of ROC			AUC of Localized ROC		
	C = 1.2	C = 1.5	C = 2.0	C = 1.2	C = 1.5	C = 2.0
#1	0.52	0.65	0.74	0.03	0.08	0.08
#2	0.66	0.82	0.82	0.05	0.09	0.11
#3	0.48	0.78	0.79	0.06	0.04	0.09
#4	0.62	0.77	0.85	0.10	0.07	0.08

Note: #1 & #2: Experienced Observer, #3 & #4: Inexperienced Observer

Therefore, if the absorption contrast is greater than 1.5 for 6mm ROI in the phantom study, we can consider that the human observers could distinguish the heterogeneous and

homogenous images. However, the AUC of Location ROC was very low (less than 0.12) at every object contrast level, thus the human observers could not detect the right ROI location even at high contrast ($C = 2.0$).

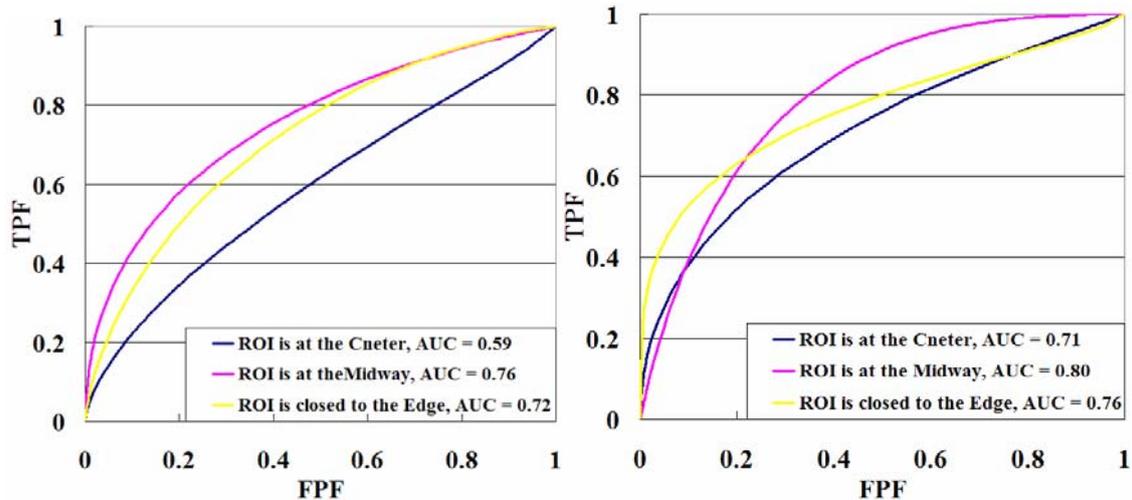


Figure 4.20 ROC curves for the heterogeneity location study from (a) the worst human observer performance and (b) the best human observer performance for 6mm ROI and with varying absorption contrast from 1.2 to 2.0.

We also test the detection abilities of the human observers by their performance on three different ROI locations shown in Figure 4.19. Again, the ROI to background contrast level was ranged from 1.2 to 2.0. Figure 4.20 indicated the worst and best human performance on 6mm ROI where the object was just off-center, in the midway or close to the edge of the phantoms. The human observers had the best detection performance if the object is in the middle of the phantom because its area-under-curve (AUC) of ROC had the largest value. When the object is closed to the edge of the phantom, the AUC of ROC is a little bit less than

it at the midway. The result is similar to that in simulation study where the AUC of ROC also increased as the 6 mm object moved from the just off-center to the midway of the phantom, and then decreased a little as the object moved closer to the edge.

4.4.3. Patient Studies

4.4.3.1. Patients Studies Image Preparation

In this study, we investigated how our NIR tomography imaging system could distinguish breast cancer and normal tissue based on chromophore parameters, such as HbO₂, from 95 cases, including 7 small size cancer lesions (cancer size ≤ 6 mm), 8 large size cancer lesions (cancer size > 6 mm) and 80 normal tissue cases. For each patient, we used 6 different wavelengths (661nm, 761nm, 785nm, 808nm, 826nm and 849nm) to measure the same cross-section of breast, and calculated the optical properties, μ_a and μ_s' for each wavelength. In the previous studies [47, 99], we know that μ_a is representation of three chromophore parameters, oxyhemoglobin (HbO₂), deoxyhemoglobin (Hb) and water (H₂O), which can be expressed as [99]

$$\varepsilon_{[Hb]}^{\lambda}[Hb] + \varepsilon_{[HbO_2]}^{\lambda}[HbO_2] + \varepsilon_{[H_2O]}^{\lambda}[H_2O] = \mu_a^{\lambda} \quad (4-18)$$

where ε (in mm² mol⁻¹) is the extinction coefficient at wavelength λ . Thus, we can calculate the absorbing chromophores, HbO₂, Hb and water with a constrained least-squares estimation based on the Equation 4-18. Generally, we used total hemoglobin (HbT), which

represents as $HbT = HbO_2 + Hb$ in μM , and oxygen saturation (SO_2) as the percentage of HbO_2 in HbT instead of HbO_2 and Hb , to be the two most important chromophores parameters in our patient analysis. [47,91] Meanwhile, based on an empirical equation of scattering theory [90]

$$\mu'_s = a\lambda^{-b} \quad (4-18)$$

we can fit the scattering amplitude (a) and scattering power (b) parameters from the reduced scattering coefficients μ'_s in these 6 wavelengths [90,91].

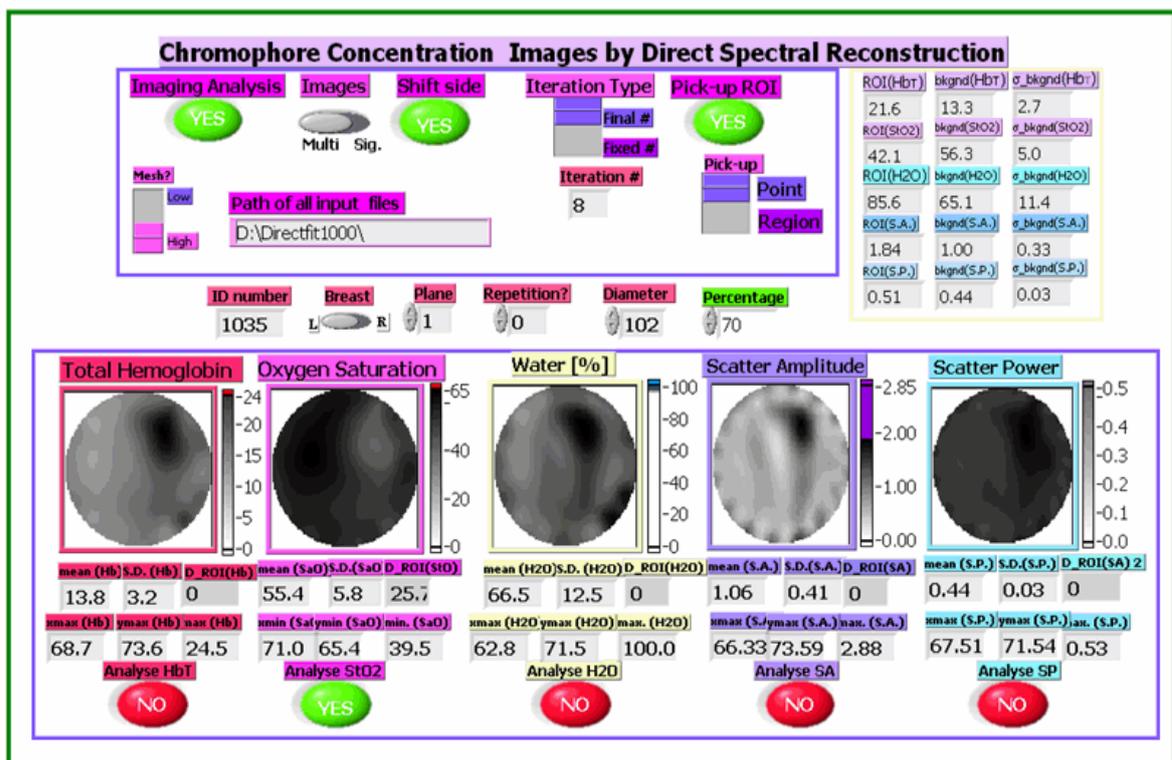


Figure 4.21. A screen shot of the Labview software interface of computer-aid program in our patient studies. the mean and standard deviation values, the maximum value and its location in x, y axis for each image, along with the display controls at the top of the field

Figure 4.12 shows the interface of our Labview program, which calculated the three chromophores parameters, HbT, SO₂ and H₂O, and two scattering parameter, a and b from the absorption and scattering coefficients of the given 6 wavelengths, and display the mean and standard deviation values, the maximum value and its location in x, y axis for each image.

4.4.3.2. Detection Task

Prior to each test, possible tumor information, such as tumor size and deepness from patient chest, were revealed to the observer. However, the observer was not informed whether the tumor information was real from a cancer patient or fake for a normal case. Then, based on the pre-given information, the observer chose the probable tumor location from the reconstructed images of HbT, StO₂, water, scattering amplitude a and scattering power b values as shown in Figure 4.22. Afterwards, the program calculated the normalized parameter for each of chromophores and scattering images, which would be the decision criteria during setting cut-off points to draw the ROC curves.

4.4.3.3. Results

Figure 4.22 shows the ROC curves based on the normalized HbT values of both tumors and normal tissues. If the tumor size is greater than 10mm, our NIR tomography system has very high diagnostic accuracy, as the AUC is equal 0.88; otherwise, the system will fail to distinguish normal and abnormal cases as AUC of the lower 95% confidence interval band is much less than 0.5.

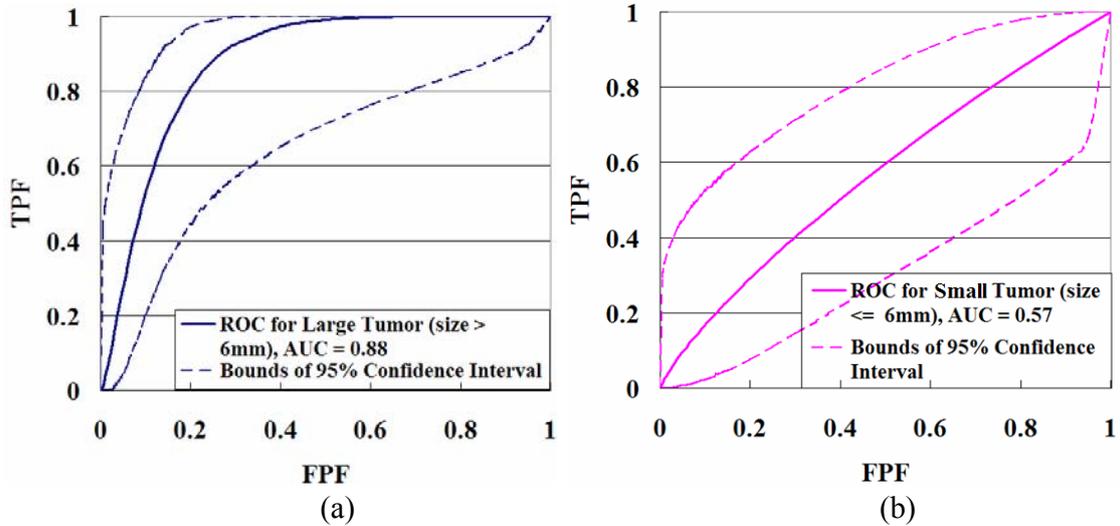


Figure 4.22 The ROC curves and the 95% confidence interval of HbT for both (a) large (size > 6mm), AUC = 0.88 and (b) small (size <= 6mm) breast cancer tumors vs. the normal tissue.

Also, the ROC and 95% confidence interval curves were made for the other normalized chromophores (i.e. StO₂ and water) shown in Figure 4.23 and Figure 4.24, but these could not distinguish the cancer and normal case based upon their ROC curves for both small and large sizes breast cancer tumors to the normal tissue as the lower 95% confidence interval band is below to the 45° diagnosis line with AUC of 0.5. The same curves of normalized scattering amplitude and scattering power parameters also generated (not shown), which could not use for diagnose the breast cancer by the same reason of StO₂ and water. Therefore, we consider the normalized HbT as the main computed feature to distinguish the cancer and normal tissue in our clinical optical tomography imaging at this time.

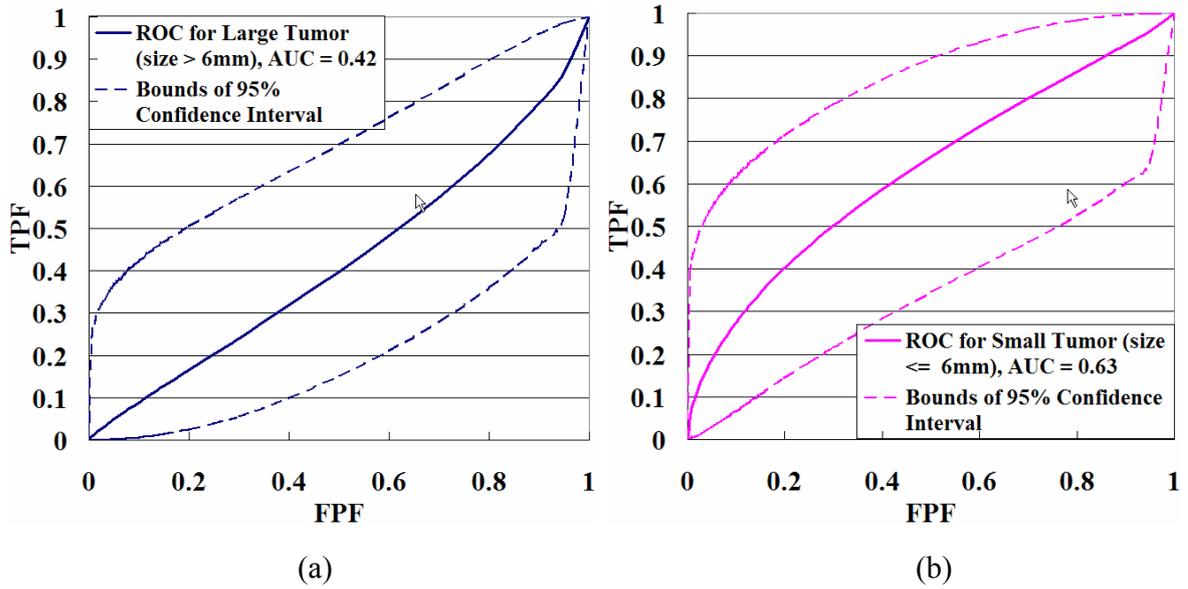


Figure 4.23 The ROC curves and the 95% confidence interval of StO₂ for both (a) large (size > 6mm) and (b) small (size <= 6mm) breast cancer tumors vs. the normal tissue.

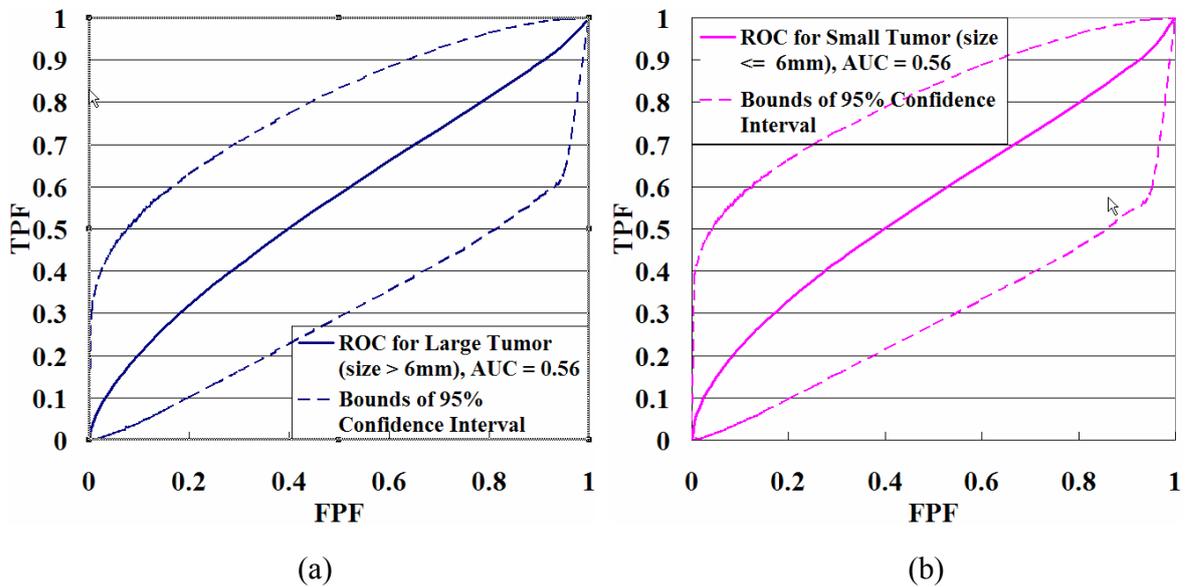


Figure 4.24 The ROC curves and the 95% confidence interval of water for both (a) large (size > 6mm) and (b) small (size <= 6mm) breast cancer tumors vs. the normal tissue.

4.5 Conclusions

In the simulation study, NIR tomography image analysis was completed with observer performance assessment. Based on the results of experienced and inexperienced human observers and CNR characterized computational calculations, NIR tomography images were assessed in terms of sensitivity to human observer discrepancy, detectable object size and object-background contrast level, reconstruction process efficacy. While computed CNR and human observer have similar capability to accurately localize the heterogeneity, human observers are much more capable to just sort out the heterogeneous images of 6mm or less from the homogeneous images. Human observers are also more capable of detecting heterogeneities with contrast levels as small as 1.1, yet this difference was negligible if the heterogeneity was at higher contrast levels compared to the homogenous background. The results also indicate that effects of iteration and algorithm performance significantly alter detectability of objects in NIR tomography images for both human observers and in the computed CNR value. Four iterations of the reconstruction process is sufficient to achieve best image quality, further iterations thereafter have negligible impact upon human observers and a negative impact upon the computed CNR detection and localization decisions. Compared to the computed observe, who has higher capability as the heterogeneity closed to the edge, the human observers have stable response as the target location changes.

In the phantom study of 6mm ROI, we could consider that the human observes could distinguish the heterogeneous images from the homogenous images when the objected

contrast is equal or greater than 1.5 as the mean AUC of ROC is equal to 0.75 at the 1.5 absorption contrast. However, all of the human observers could not detect the right ROI location as the contrast is from 1.2 to 2.0 because the AUC of LROC for the entire contrast level under 2.0 is less than 0.12. Also, the human observer could easiest to distinguish the heterogeneous images from the homogenous if the object is at the midway of the phantom when the object is small. In the future, the larger size object (ROI = 10mm) need to be tested.

In the patient study, we employed three normalized chromophores and two normalized scattering parameters as the computed features to determine the clinically ability for distinguishing breast cancer tumors from benign tumors in system. AUC of 0.88 was seen for the normalized HbT for large breast cancers showing that our clinical system has good discernment from the breast cancer cases to normal tissue if the tumor size is larger than 6mm. However, we did not achieve useful get satisfied results from normalized SO₂, water, scattering amplitude and scattering power measure to date, or for the smaller size cancer tumors, as the lower 95% confidence interval band for all of these parameters are below the 45-degree diagonal of the ROC space since the AUC was 0.5. This study should be verified with larger numbers of patient cases.

CHAPTER 5

Conclusion Remarks

In this thesis, assessment of statistical accuracy of NIR tomographic imaging and computer-assisted diagnosis systems was present. Mean-squared error (MSE) was used to evaluate methods for regularizing the inverse image reconstruction problem. In computer simulation study, it was observed that the bias error dominates at high regularization parameter values while variance dominates as the algorithm is allowed to approach the optimal solution. This optimum does not necessarily correspond to the minimum projection error solution, but typically requires further iteration with a decreasing regularization parameter to reach the lowest image error. Increasing measurement noise causes a need to constrain the minimum regularization parameter to higher values in order to achieve a minimum in the overall image MSE. In phantom studies, the results show that imaging performance is not limited by random measurement error, but rather by calibration issues. The image error over the entire field of view is generally not minimized when an accurate homogeneous estimate of the phantom properties is available; however, local image error over a target region of interest (ROI) is reduced. The image reconstruction process which

includes a Levenberg–Marquardt style regularization provides good minimization of the objective function, yet its reduction is not always correlated with an overall image error decrease. Minimization of the bias in an ROI which contains localized changes in the optical properties can be achieved through five to nine iterations of the algorithm. Pre-calibration of the algorithm through statistical evaluation of phantom studies may provide a better measure of the image accuracy than that implied by minimization of the standard objective function.

The contrast-to-noise ratio (CNR) was then used to determine the detectability of objects within reconstructed images from diffuse near-infrared tomography. Three methods of calculate the CNR were present and compared. This study was performed on both computer simulated images and phantom study images. It was concluded that there was a maximal value of CNR near the location of an object within the image and that the size of the true region could be estimated from the CNR. Experimental and simulation studies led to the conclusion that objects can be automatically detected with CNR analysis and that our current system has a spatial resolution limit near 4 mm and a contrast resolution limit near 1.4. Two approaches were tested to find the location and size of the object based on maximal CNR, including a random node-by-node matching process and a new 2-D linear convolution method. Although the results from these two methods are similar, the convolution method is faster than the node-by-node matching process

At last, Receiver operating characteristic (ROC) analysis based observer performance studies on NIR tomography imaging system, a widely used methodology to address the clinical efficacy of medical imaging systems, were performed and interpreted. Parametric and

nonparametric approaches have been applied for estimating the ROC curve. Two MATLAB GUI based programs were developed for this study. One was used as the observer image reading test program and the other one was used for the ROC/LROC curve analysis tasks performed in this thesis. Binormal model based categorical and continuous data ROC/LROC fitting methods were embedded in this software platform. Parametric and nonparametric approaches have been applied for estimating the ROC curves. Computational CNR and human observers are then compared in the capability to accurately localize the heterogeneity under high heterogeneity contrast level and the results are discussed thereafter. The performance study results indicate that effects of iteration and algorithm performance significantly alter detectability of objects in NIR tomography images.

APPENDIX

Statistics and Probability Theory and Terms

"It has become accepted in today's world that in order to learn about something, you must first collect data..." [1]. Probability theory and statistics are the practice (science) of data analysis summarizing data and drawing inferences about the larger population from which it was drawn. They involve not only collecting, presenting and characterizing data, but also estimation and hypothesis testing. This chapter is dedicated to form the probability and statistics foundation of the thesis. The statistics terms, formulas and theories used in the thesis are reviewed in the following sections.

A.1 Fundamental Concepts of Probability

A.1.1 Random Experiments

Probability theory is based on the paradigm of a random experiment; that is, an experiment whose outcome cannot be predicted with certainty, before the experiment is run. We usually assume that the experiment can be repeated indefinitely under essentially the same conditions. Probability theory is concerned with the long-term behavior as the

experiment is replicated. Naturally, a complete definition of a random experiment requires a careful definition of precisely what information about the experiment is being recorded, that is, a careful definition of what constitutes an outcome.

A.1.2 Sets, Sample Space and Events

A set is simply a collection of objects; the objects are referred to as elements of the set. The statement that s is an element of set S is written $s \in S$. If A and B are sets then A is a subset of B if every element of A is also an element of B :

$$A \subseteq B \text{ if and only if } s \in A \text{ implies } s \in B.$$

By definition, a set is completely determined by its elements. Thus sets A and B are equal if they have the same elements:

$$A = B \text{ if and only if } A \subseteq B \text{ and } B \subseteq A.$$

By contrast, the empty set, denoted as \emptyset , is the set with no elements.

The sample space of a random experiment is a set S that includes all possible outcomes of the experiment. For simple experiments, the sample space may be precisely the set of possible outcomes. For complex experiments, the sample space is a mathematically convenient set that includes the possible outcomes and perhaps other elements as well.

Certain subsets of the sample space of an experiment are referred to as events. Thus, an event is a set of outcomes of the experiment. Each time the experiment is run, a given event A either occurs, if the outcome of the experiment is an element of A , or does not occur, if the outcome of the experiment is not an element of A . Intuitively, statisticians usually think of an event as a meaningful statement about the experiment. The sample space S itself is an event

and by definition it always occurs. On the other hand, the empty set \emptyset is also an event and by definition it never occurs. More generally, if A and B are events in the experiment and A is a subset of B, then the occurrence of A implies the occurrence of B.

A.1.3 Axioms of Probability

The probability of an event A in the sample space is by definition a number $P(A)$ that satisfies the following three axioms [2]:

1. $P(A)$ is a nonnegative number; that is,

$$P(A) \geq 0 \quad (\text{A-1})$$

2. The probability of the event S (i.e., the certain event) is unity:

$$P(S) = 1 \quad (\text{A-2})$$

3. If two events A and B have no common outcomes (such events are said to be disjoint or mutually exclusive), the probability of the event $\{A \text{ or } B\}$ is the sum of their probabilities:

$$P(A \text{ or } B) = P(A) + P(B) \quad (\text{A-3})$$

A.1.4 Conditional Probability

Suppose that we have a random experiment within sample space S, and probability measure P. Suppose also that we know that an event B has occurred. In particular, if A is another event then A occurs if and only if A and B occur; effectively, the sample space has been reduced to B. Thus, the probability of A, given that we know B has occurred, should be proportional to $P(A \cap B)$. However, conditional probability, given that B has occurred,

should still be a probability measure, that is, it must satisfy the axioms of probability. This forces the proportionality constant to be $1/P(B)$. Thus, we are led to the following definition of conditional probability:

Let A and B be events in a random experiment with $P(A) > 0$ and $P(B) > 0$. The **conditional probability** of A given B is defined to be

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (\text{A-4})$$

and the conditional probability of B given A is defined to be

$$P(B | A) = \frac{P(A \cap B)}{P(A)} \quad (\text{A-5})$$

Apparently, by solving both (A-4) and (A-5) we also have:

$$P(A \cap B) = P(A | B)P(B) = P(B | A)P(A) \quad (\text{A-6})$$

A.2 Probability Distributions

A.2.1 Random Variable

A random variable is a function that assigns a real number to each outcome of a chance experiment which defined in A.1.1. Random variables can be discrete, continuous, or mixed. A random variable is a function whose definition is the sample space of a chance experiment and whose range of values is the real line. If X is a random variable, any function of X is also random variable. A discrete random variable can be described by its probability mass distribution as: $P(X = x_i)$, $i = 1, 2, 3, \dots$, is plotted versus x_i , where X is the random variable and the x_i is the value it takes on with probabilities $P(X = x_i)$. Cumulative probability

distribution function (cdf) and its probability density function (pdf) are two other possible descriptions of a random variable, which are usually used to describe continuous random variables.

A.2.2 Cumulative Probability Distribution Function (cdf)

For a random variable X , the cumulative probability distribution function (cdf) $F_X(x)$ is defined as

$$F_X(x) = P(X \leq x) \quad (\text{A-7})$$

Which means, the cdf is the probability that random variable X is less than or equal to the running value x . This description works for discrete, continuous, or mixed random variable.

Properties of cdf:

- a) $0 \leq F_X(x_0) \leq 1$, for all x_0
- b) Monotonically non-decreasing function
- c) The probability that the random variable lies in a given range is

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$$

A.2.3 Probability Density Function (pdf)

The pdf $f_X(x)$ of a continuous random variable is defined as

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (\text{A-8})$$

Properties of pdf:

- a) Nonnegative everywhere
- b) The area under the pdf curve is 1
- c) Its integral is cdf, i.e. $F_X(x) = \int_{-\infty}^x f_X(u) du$

REFERENCE

1. Arridge, S. *Forward and inverse problems in time-resolved infrared imaging*. in *Medical Optical Tomography: Functional Imaging and Monitoring*. 1993. Bellingham, WA: SPIE.
2. Arridge, S. and M. Schweiger, *Inverse methods for optical tomography*, in *Information Processing in Medical Imaging*, A.F.G. H.H. Barrett, Editor. 1993, Springer-Verlag: Flagstaff, AZ. p. 259-277.
3. Arridge, S. and M. Schweiger, *Image reconstruction in optical tomography*. *Philosophical Transactions: Biological Sciences Lond. B*, 1997. **352**: p. 717-726.
4. Baierl, P., et al., *The contrast to noise ratio as a measure of tissue contrast in nuclear magnetic resonance tomography*. *Digitale Bilddiagn.*, 1986. **6**(3): p. 101-110.
5. Barrett, H., C. Abbey, and E. Clarkson, *Objective assessment of image quality: III. ROC metrics, ideal observers and likelihood-generating functions*. *Journal of the Optical Society of America A*, 1998. **15**(6): p. 1520-1535.
6. Barrett, H. and K. Myers, *Foundations of Image Science*. 1st edition ed. 2003: Wiley-Interscience.
7. Berkson, J., *Maximum-likelihood and minimum chi-square estimates of the logistic function*. *Journal of the American Statistical Association*, 1955. **50**: p. 130-162.
8. Birdwell, R., et al., *Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection*. *Radiology*, 2001. **291**(1): p. 192-202.
9. Boas, D., et al., *Detection and characterization of optical inhomogeneities with diffuse photon density waves: a signal-to-noise analysis*. *Applied Optics*, 1997. **36**(1): p. 75-92.
10. Bushberg, J., et al., *Chapter 10: Image Quality*, in *The Essential Physics of Medical Imaging*. 2001, Lippincott Williams & Wilkins. p. 255-292.

11. Cerussi, A., et al., *Sources of absorption and scattering contrast for near-infrared optical mammography*. Academic Radiology, 2001. **8**(3): p. 211-218.
12. Chakraborty, D., *Statistical power in observer-performance studies: comparison of the receiver operating characteristic and free-response methods in tasks involving localization*. Academic radiology, 2002. **9**(2): p. 147-156.
13. Chakraborty, D. and L. Winter, *Free-response Methodology: Alternate Analysis and a New Observer-performance Experiment*. Radiology, 1990. **174**(3 Pt 1): p. 873-81.
14. Chan, H., et al., *Improvement of Radiologists' Characterization of Mammographic Masses by Using Computer-aided Diagnosis: An ROC Study*. Radiology, 1999. **212**(3): p. 817-827.
15. Colak, S., et al., *Tomographic image reconstruction from optical projections in light-diffusing media*. Applied Optics, 1997. **36**(1): p. 180-213.
16. Cutler, M., *Transillumination as an aid in the diagnosis of breast lesions*. Surgery, Gynecology & Obstetrics, 1929. **48**: p. 721-729.
17. Dehghani, H., et al., *Effects of refractive index on near-infrared tomography of the breast*. Applied Optics, 2005. **44**(10): p. 1870-1878.
18. Dehghani, H. and D. Delpy, *Linear single-step image reconstruction in the presence of nonscattering regions*. Journal of the Optical Society of America A, 2002. **19**(6): p. 1162-1171.
19. Dehghani, H., et al., *Multiwavelength Three-Dimensional Near-Infrared Tomography of the Breast: Initial Simulation, Phantom, and Clinical Results*. Applied Optics, 2003. **42**(1): p. 135-145.
20. Dorfman, D. and E.J. Alf, *Maximum likelihood estimation of parameters of signal detection theory--a direct solution*. Psychometrika, 1968. **33**(1): p. 117-124.
21. Dorfman, D. and E.J. Alf, *Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—Rating-method data*. Journal of Mathematical Psychology, 1969. **6**(3): p. 487-496.
22. Dorfman, D., K. Berbaum, and C. Metz, *Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method*. Investigative Radiology, 1992. **27**(9): p. 723-731.
23. Dorn, O., *A transport-backtransport method for optical tomography*. Inverse Problems, 1998. **14**(5): p. 1107-1130.

24. Doyleya, M., et al., *Thresholds for detecting and characterizing focal lesions using steady-state MR elastography*. Medical Physics, 2003. **30**(4): p. 495-504.
25. Engl, H. and W. Grever, *Using the L-curve for determining optimal regularization parameters*. Numerische Mathematik, 1994. **69**: p. 25-31.
26. Fantini, S., et al., *Assessment of the size, position, and optical properties of breast tumors in vivo by noninvasive optical methods*. Applied Optics, 1998. **37**(10): p. 1982-1989.
27. Franceschini, M., et al., *Frequency-domain techniques enhance optical mammography: initial clinical results*. Processing of The National Academy of Sciences of the USA, 1997. **94**(12): p. 6468-6473.
28. Goodenough, D. and K. Weaver, *Factors related to low contrast resolution in CT scanners*. Computerized Radiology, 1984. **8**(5): p. 297-308.
29. Graber, H., Y. Pei, and R. Barbour, *Imaging of spatiotemporal coincident states by DC optical tomography*. IEEE Transactions on Medical Imaging, 2002. **21**(8): p. 852- 866.
30. Grey, D. and B. Morgan, *Some aspects of ROC curve-fitting: Normal and logistic models*. Journal of Mathematical Psychology, 1972. **9**(1): p. 128-139.
31. Grosenick, D., H. Wabnitz, and H. Rinneberg, *Time-resolved imaging of solid phantoms for optical mammography*. Applied Optics, 1997. **36**(1): p. 221-231.
32. Grosenick, D., et al., *Development of a Time-Domain Optical Mammograph and First in vivo Applications*. Applied Optics, 1999. **38**(13): p. 2927-2943.
33. Hanley, J. and B. McNeil, *A method of comparing the areas under receiver operating characteristic curves derived from the same cases*. Radiology, 1983. **148**(3): p. 839-843.
34. Hebden, J., et al., *Three-dimensional optical tomography of the premature infant brain*. Physics in Medicine and Biology, 2002. **47**(23): p. 4155-4166.
35. Hebden, J., et al., *Simultaneous reconstruction of absorption and scattering images by multichannel measurement of purely temporal data*. Optics Letters, 1999. **24**(8): p. 534-536.
36. Hielscher, A. and S. Bartel, *Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography*. Journal of Biomedical Optics, 2001. **6**(2): p. 183-192.
37. Hielscher, A., A. Klose, and K. Hanson, *Gradient-based iterative image reconstruction scheme for time- resolved optical tomography*. IEEE Transactions on Medical Imaging, 1999. **18**(3): p. 262-271.

38. Hill, C., J. Bamber, and D. Cosgrove, *Performance criteria for quantitative ultrasonography and image parameterisation*. Clinical Physics and Physiological measurement, 1990. **11**(4A): p. 57-73.
39. Jiang, H., et al., *Frequency-domain optical image reconstruction in turbid media: an experimental study of single-target detectability*. Applied Optics, 1997. **36**(1): p. 52-63.
40. Jiang, H., et al., *Simultaneous reconstruction of optical absorption and scattering maps in turbid media from near-infrared frequency-domain data*. Optics Letters, 1995. **20**(20): p. 2128-2130.
41. Jiang, H., et al., *Optical image reconstruction using frequency-domain data: simulations and experiments*. Journal of the Optical Society of America A, 1996. **13**(2): p. 253-266.
42. Jiang, S., et al., *Quantitative analysis of near-infrared tomography: sensitivity to the tissue-simulating precalibration phantom*. Journal of Biomedical Optics, 2003. **8**(2): p. 308-315.
43. Li, A., et al., *Reconstructing chromosphere concentration images directly by continuous-wave diffuse optical tomography*. Optics Letters, 2004. **29**(3): p. 256-258.
44. Lloyd, C., *Using Smoothed Receiver Operating Characteristic Curves to Summarize and Compare Diagnostic Systems*. Journal of the American Statistical Association, 1998. **93**(444): p. 1356-1364.
45. Lopez, H., M. Loew, and D. Goodenough, *Objective analysis of ultrasound images by use of a computational observer*. IEEE Transactions on Medical Imaging, 1992. **11**(4): p. 496-506.
46. McBride, T., *Spectroscopic Reconstructed Near Infrared Tomographic Imaging for Breast Cancer Diagnosis*, in *Thayer School of Engineering*. 2001, Dartmouth College: Hanover. p. 1-195.
47. McBride, T., et al., *Spectroscopic diffuse optical tomography for quantitatively assessing hemoglobin concentration and oxygenation in tissue*. Applied Optics, 1999. **38**(25): p. 5480-5490.
48. McBride, T., et al., *A parallel-detection frequency-domain near-infrared tomography system for hemoglobin imaging of the breast in vivo*. Review of Scientific Instruments, 2001. **72**(3): p. 1817-1824.
49. McBride, T.O., et al., *Near-infrared tomographic imaging of heterogeneous media -- a preliminary study in excised breast tissue*. Proc. SPIE, 2001. **4250**.
50. McBride, T., et al., *Strategies for Absolute Calibration of Near Infrared Tomographic*

Tissue Imaging, in *Oxygen Transport to Tissue XXI*, J.F. Dunn, Swartz, H. M., Editor. 2001, Lengerich: Pabst.

51. McBride, T.O., et al., *Multispectral near-infrared tomography: a case study in compensating for water and lipid content in hemoglobin imaging of the breast*. Journal of Biomedical Optics, 2002. **7**(1): p. 72-79.

52. McBride, T.O., et al., *Initial studies of in vivo absorbing and scattering heterogeneity in near-infrared tomographic breast imaging*. Optics Letters, 2001. **26**(11): p. 822-824.

53. McClish, D., *Analyzing a portion of the ROC curve*. Medical Decision Making, 1989. **9**(3): p. 190-195.

54. Metz, C., *ROC Methodology in Radiologic Imaging*. Investigative radiology, 1986. **21**(9): p. 720-733.

55. Metz, C., B. Herman, and J. Shen, *Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data*. Statistics in Medicine, 1998. **17**(9): p. 1033 - 1053.

56. Metz, C., S. Starr, and L. Lusted, *Observer performance in detecting multiple radiographic signals. Prediction and analysis using a generalized ROC approach*. Radiology, 1976. **121**(2): p. 337-347.

57. Mitic, G., et al., *Time-gated transillumination of biological tissues and tissue-like phantoms*. Applied Optics, 1994. **33**(28): p. 6699-6710.

58. Model, R., et al., *Optical imaging: three-dimensional approximation and perturbation approaches for time-domain data*. Applied Optics, 1998. **37**(34): p. 7968-7976.

59. Mourant, J., et al., *Predictions and measurements of scattering and absorption over broad wavelength ranges in tissue phantoms*. Applied Optics, 1997. **36**(4): p. 949-957.

60. Neumaier, A., *Solving ill-conditioned and singular linear systems: a tutorial on regularization*. SIAM Review, 1998. **40**(3): p. 636-666.

61. Nishikawa, R., et al., *Large Scale Observer Study to Measure the Benefits of Computer-aided Detection to Screening Mammography*. Radiology, 1999. **213**(P): p. 150.

62. Ntziachristos, V. and B. Chance, *Probing physiology and molecular function using optical imaging: applications to breast cancer*. Breast Cancer Research, 2001. **3**(1): p. 41-46.

63. Patterson, M., et al., *Diffusion equation representation of photon migration in tissue*, in *Microwave Theory and Techniques Symposium Digest*. 1991, IEEE: New York. p. 905-908.

64. Patterson, M., B. Wilson, and D. Wyman, *The propagation of optical radiation in tissue I. models of radiation transport and their application*. Lasers in Medical Science, 1991. **6**: p. 155-168.
65. Paulsen, K. and H. Jiang, *Spatially varying optical property reconstruction using a finite element diffusion equation approximation*. Medical Physics, 1995. **22**(6): p. 691-701.
66. Paulsen, K.D., Jiang, H., *Enhanced frequency-domain optical image reconstruction in tissues through total-variation minimization*. Applied Optics, 1996. **35**(19): p. 3447-3458.
67. Peters, V.G., Wyman, D. R., Patterson, M. S., Frank G. L., *Optical properties of normal and diseased human breast tissues in the visible and near infrared*. Physics in Medicine and Biology, 1990. **35**(9): p. 1317-1334.
68. Pogue, B., et al., *Three-Dimensional Simulation of Near-Infrared Diffusion in Tissue: Boundary Condition and Geometry Analysis for Finite-Element Image Reconstruction*. Applied Optics, 2001. **40**(4): p. 588-600.
69. Pogue, B., et al., *Comparison of imaging geometries for diffuse optical tomography of tissue*. Optics Express, 1999. **4**(8): p. 270-286.
70. Pogue, B., et al., *Spatially Variant Regularization Improves Diffuse Optical Tomography*. Applied Optics, 1999. **38**(13): p. 2950-2961.
71. Pogue, B., J. Mourant, and J. George, *Topics in biomedical optics - part I: introduction*. Applied Optics, 2005. **44**(10): p. 1784-1784.
72. Pogue, B., et al., *Initial assessment of a simple system for frequency domain diffuse optical tomography*. Physics in Medicine and Biology, 1995. **40**(10): p. 1709-1729.
73. Pogue, B., et al., *Quantitative Hemoglobin Tomography with Diffuse Near-Infrared Spectroscopy: Pilot Results in the Breast*. Radiology, 2001. **218**(1): p. 261-266.
74. Pogue, B., et al., *Statistical analysis of nonlinearly reconstructed near-infrared tomographic images. I. Theory and simulations*. IEEE Transactions on Medical Imaging, 2002. **21**(7): p. 755- 763.
75. Pogue, B., et al., *Instrumentation and design of a frequency-domain diffuse optical tomography imager for breast cancer detection*. Optics Express, 1997. **1**(13): p. 391-403.
76. Pogue, B., et al., *Contrast-detail analysis for detection and characterization with near-infrared diffuse tomography*. Medical Physics, 2000. **27**(12): p. 2693-2700.
77. Press, W., et al., *Numerical Recipes in Fortran*. 2nd ed. 1992, New York: Cambridge University Press.

78. Qi, J., G. Klein, and R. Huesman, *Image properties of list-mode likelihood reconstruction for arectangular positron emission mammograph with DOI measurements*. IEEE Transactions on Nuclear Science, 2001. **48**(4): p. 1343-1349.
79. Robson, K., C. Kotre, and K. Faulkner, *The use of contrast-detail test objects in the optimization of optical density in mammography*. The British Journal of Radiology, 1995. **68**(807): p. 277-282.
80. Ross, S., *Introduction to Probability and Statistics for Engineers and Scientists*. Hardcover ed. Vol. 2nd Edition. 2000: Harcourt/Academic Press.
81. Rownd, J., et al., *Phantoms and automated system for testing the resolution of ultrasound scanners*. Ultrasound in Medicine & Biology, 1997. **23**(2): p. 245-260.
82. Schmidt, F., et al., *A 32-channel time-resolved instrument for medical optical tomography*. Review of Scientific Instruments, 2000. **71**(1): p. 256-265.
83. Schmitz, C., et al., *Instrumentation for fast functional optical tomography*. Review of Scientific Instruments, 2002. **73**(2): p. 429-439.
84. Seltzer, S., et al., *Flattening of the contrast-detail curve for large lesions on liver CT images*. Medical Physics, 1993. **21**(10): p. 1547-1555.
85. Shah, N., et al., *Noninvasive functional optical spectroscopy of human breast tissue*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(8): p. 4420-4425.
86. Shapiro, D.E., *The interpretation of diagnostic tests*. Statistical Methods in Medical Research, 1999. **8**: p. 113-134.
87. Siegel, A., J. Marota, and D. Boas, *Design and evaluation of a continuous-wave diffuse optical tomography system*. Optics Express, 1999. **4**(8): p. 287-298.
88. Song, X., et al., *Automated region detection based on the contrast-to-noise ratio in near-infrared tomography*. Applied Optics, 2004. **43**(5): p. 1053-1062.
89. Song, X., et al., *Statistical analysis of nonlinearly reconstructed near-infrared tomographic images. II. Experimental interpretation*. IEEE Transactions on Medical Imaging, 2002. **21**(7): p. 764- 772.
90. Srinivasan, S., et al., *Interpreting Hemoglobin and Water Concentration, Oxygen Saturation, and Scattering Measured by Near-Infrared Tomography of Normal Breast In Vivo*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(21): p. 12349-12354.

91. Srinivasan, S., et al., *Spectrally constrained chromophore and scattering near-infrared tomography provides quantitative and robust reconstruction*. Applied Optics, 2005. **44**(10): p. 1858-1869.
92. Starr, S., et al., *Visual Detection and Localization of Radiographic Images*. Radiology, 1975. **116**(3): p. 533-538.
93. Suzuki, K., et al., *Quantitative measurement of optical parameters in normal breasts using time-resolved spectroscopy: In vivo results of 30 Japanese women*. Journal of Biomedical Optics, 1996. **1**(3): p. 330-334.
94. Swensson, R., *Unified measurement of observer performance in detecting and localizing target objects on images*. Medical physics, 1996. **23**(10): p. 1709-25.
95. Swets, J., *ROC analysis applied to the evaluation of medical imaging techniques*. Investigative Radiology, 1979. **14**(2): p. 109-121.
96. Swets, J., *Evaluation of Diagnostic Systems: methods from signal detection theory*. 1982, New York, NY: Academic Press.
97. Swets, J., *Indices of Discrimination or Diagnostic Accuracy: Their ROCs and Implied Models*. Psychonomic Bulletin & Review, 1986. **99**(1): p. 100-117.
98. Tosteson, T., et al., *Confidence maps and confidence intervals for near infrared images in breast cancer*. IEEE Transactions on Medical Imaging, 1999. **18**(12): p. 1188-1193.
99. Tromberg, B., et al., *Non-invasive measurements of breast tissue optical properties using frequency-domain photon migration*. Philosophical Transactions: Biological Sciences, 1997. **352**(1354): p. 661-668.
100. van Staveren, H., et al., *Light scattering in Intralipid-10% in the wavelength range of 400-1100 nm*. Applied Optics, 1991. **30**(31): p. 4507-4514.
101. WAGNER, R., et al., *Statistics of Speckle in Ultrasound B-Scans*. IEEE TRANSACTIONS ON SONICS AND ULTRASONICS, 1983. **30**(3): p. 156-163.
102. Wickens, T., *Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data*. Journal of Mathematical Psychology, 1992. **36**(2): p. 213-234.
103. Wieand, S., et al., *A Family of Nonparametric Statistics for Comparing Diagnostic Markers with Paired or Unpaired Data*. Biometrika, 1989. **76**(3): p. 585-592.
104. Wu, C., *Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis*. The Annals of Statistics, 1986. **14**(4): p. 1261-1295.

105. Zheng, B., et al., *Soft-Copy Mammographic Readings with Different Computer-assisted Detection Cuing Environments: Preliminary Findings*. Radiology, 2001. **221**(3): p. 633-640.
106. Zhou, X. and J. Harezlak, *Comparison of bandwidth selection methods for kernel smoothing of ROC curves*. Statistics in Medicine, 2002. **21**(14): p. 2045-2055.
107. Ziemer, R.E., *Elements of Engineering Probability and Statistics*. 1997, New Jersey: Pearson Education. 236pp.
108. Zou, K., W. Hall , and D. Shapiro, *Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests*. Statistics in Medicine, 1997. **16**(9): p. 2143-2156.