ABSTRACT

Title of dissertation: TASK SPECIFIC EVALUATION METHODOLOGY FOR CLINICAL FULL FIELD DIGITAL MAMMOGRAPHY

Haimo Liu, Doctor of Philosophy, 2012

Dissertation directed by: Professor Iacovos Kyprianou
Department of Bioengineering

Purpose: The purpose of this dissertation is to evaluate the image quality of clinical Full Field Digital Mammography (FFDM) systems. This is done by evaluating image acquisition performance of clinical FFDM in a comprehensive way that accounts for scatter, focal spot un-sharpness, detector blur and anti-scatter grid performance using an anthropomorphic phantom. Additionally we intend to provide a limited evaluation of the effects that image processing in clinical FFDM has in signal detectability.

Methodology: We explored different strategies and a variety of mathematical model observers in order to evaluate the performance of clinical FFDM systems under different conditions. To evaluate image acquisition performance, we tested a system-model-based Hotelling observer (SMHO) model on a bench-top system using a uniform anthropomorphic phantom for an signal known exactly background known exactly (SKE/BKE) task. We then applied this concept on two clinical FFDM systems to compare their performance. In a limited study to evaluate the effects of image processing in the detectability

of FFDM, we implemented the channelized Hotelling observer (CHO) model on clinically realistic images of an anatomical phantom for an SKE/BKE task.

Results: Even though the two systems use different detection technologies, there was no significant difference between their image acquisition performances quantified by the Contrast-Detail (CD) curves. We applied the CHO model to investigate the image processing algorithms used in GE Senographe DS FFDM system. For the particular SKE/BKE task with rotationally symmetric signals, the image processing tends to contribute to a non-significant reduction of system detectability.

Conclusion: We provided a complete description of FFDM system performance including the image acquisition chain and post-acquisition image processing. We demonstrated the simplicity and effectiveness of both the MFHO and CHO methods in a clinical setting.

TASK SPECIFIC EVALUATION METHODOLOGY FOR CLINICAL FULL FIELD DIGITAL MAMMOGRAPHY

by

Haimo Liu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Commmittee:

Iacovos Kyprianou, Advisor/Co-Chair
Yu Chen, Chair
Aldo Badano
Luis Benevides
Yang Tao
Nam Sun Wang, Dean's Representative

ACKNOWLEDGMENTS

I am so grateful to those people who have made this research and dissertation possible and to those who have made the four years that I spent at U.S. FDA a remarkable memory. I want to express my thanks to Bruce Fleharty for helping me make products and parts in the machine shop. Without his advice and experience, it would have been impossible to bring so many complex designs into reality. I want to thank Dr. Robert Jennings, Dr. Berkman Sahiner and Dr. Rongping Zeng who helped me review my super long *Medical Physics* paper. I want to thank Dr. Songxiang Gu, who always seemed to have the answers to my questions before I knocked on his door. I would like to thank Dr. Rui Peng for the discussion with him about my research as well as about life experience. I wish to thank the members of my committee at UMD, Dr. Yang Tao, Dr. Yu Chen and Dr. Nam Sun Wang for their comments and suggestions. They provided many valuable ideas to guide the direction of my research. I wish to thank Dr. Luis Benevides for his continuous guidance at the National Naval Medical Center. His advice and knowledge in radiation dosimetry were extremely helpful in building my experiments. I would like to especially thank Kyle Myers, Aldo Badano, Kish Chakrabarti and Richard Kaczmarek at FDA who have worked with me closely on my dissertation project to get it going. More so than anyone else, Iacovos Kyprianou has guided my bioengineering education and my research

projects for the past five years. Whenever and wherever I am in life, I will be thankful for everything I have learned from him. Thanks to him for being such an instructive advisor and a good friend of mine. In the end, I would like to thank my family and girlfriend for being supportive since the very beginning.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1 Background

In 1973, breast cancer ranked first as a cause of cancer death among North American women, and it accounted for 20% of all cancer mortalities[1]. In the next 30 years, the total breast cancer mortality was substantially reduced due to improved treatments including the use of hormonal and cytotoxic treatments[2, 3]. However, breast cancer is still the second leading cause of cancer death (after lung cancer) in women. In 2008, 182,460 cases of invasive breast cancer and 67,770 cases of situ breast cancer were reported. 40,930 breast cancer deaths were estimated in the same year[4], and 39,520 women were estimated to die from breast cancer in 2011[5]. These statistics highlight the clinical significance of reducing breast cancer mortality.

Many scientists have contributed to investigating the prevalence of breast cancer by attempting to better understand the etiology of breast cancer. Even though no confirmed cause of breast cancer has been reported yet[5–10], the analysis of the etiology leads to the development of breast cancer therapies. In most therapies, early cancer detection plays a key role in increasing patient survival rates. The rates are significantly higher when tumors are detected and treated at an early stage, i. e., smaller than 2cm[11].

In the past 15 years, the breast cancer mortality rate has been decreased by 30% to 50% due to the use of screening mammography in early cancer detection[12, 13]. The mortality rate can be remarkably reduced for women who are 40 years of age or older[14, 15]. Lower benefits to younger women are due to their increased risk of rapidly growing cancerous tissues and higher breast densities which affect the sensitivity of the screening mammography[16, 17].

During the past ten years, Full Field Digital Mammography (FFDM) has been gaining popularity and has widely replaced screen-film systems in clinical practice in the USA. A typical FFDM system consists of an x-ray tube that emits x-ray photons; a compression paddle that provides consistent pressure to breasts; a digital detector that interacts with the x-rays transmitted from the breast, absorbs energy carried by the photons, converts this energy to a measurable signal and collects this signal; a computer console that stores the raw images obtained at detector; and a display workstation where the processed images could be reviewed by human observers such as radiologists. One of the biggest image quality disadvantages of film systems is the sigmoid shaped film response to radiation exposure. The contrast of the film is determined by the gradient, or the slope, of its response curve. The non-linear regions of the curve (the "toe" and the "shoulder") have very low contrast, and only the linear region with the higher slope has contrast suitable for imaging. Therefore the linear region of the response curve limits the dynamic range[18]. The effectiveness of film systems is also limited by film processing artifacts that contribute to the false-negative rate[19]. With the development of digital detectors, the response curve in FFDM systems is linearly proportional to incident x-ray intensity, resulting in a wider dynamic range. Unlike the film, which has the roles of acquisition, storage and display, in FFDM, image acquisition, storage and display functions are performed by different subsystems[20]. The film contrast is determined by how the film is processed; whereas in FFDM, contrast can be adjusted and optimized based on the dynamic range of the system. In clinical practice, FFDM systems have significantly lower recall rates[21] and better diagnostic accuracy for young women (under age 50) with dense breasts[22]. In addition, the FFDM decreases the image acquisition time and reduces film library maintenance costs[23].

1.2 Image Quality and Model Observers

A medical image is a representation of human organs. It reveals tissue structures and it is relevant to clinical diagnosis. Various evaluation methods have been proposed in the

past to quantify x-ray image quality. In the early 1970s, evaluation methods such as the mean square difference between objects and their images did not take into account the purpose of medical images. Without linking the phrase "Image Quality" to diagnostic tasks, at that time, the assessment of system performance entirely relies on the subjective impression of humans. Running clinical trials links image quality with how well the image can fulfill its medical tasks or purposes. Previous research has shown that even though humans are skilled at signal detection and discrimination tasks[24], they have a number of inefficiencies including inter- and intra-observer variabilities and a limited range of sampling and integration[25]. In addition, running clinical trials is very expensive and time consuming. As an alternative, in clinical practice, image quality is often characterized by pixel SNR that describes the relative strength of signal with respect to that of noise (mean pixel value divided by the standard deviation). It is commonly used because of its simplicity. However, pixel SNR ignores correlations between pixels and does not account for the shape of signals. The pixel SNR could provide inaccurate or misleading information if it is used as a figure of merit[26]. It is therefore necessary to develop clinically practical objective assessment methods that can provide meaningful measurements of image quality.

Classification is a common clinical task in breast cancer diagnosis. Signal detection belongs to a classification task: to detect whether breast cancer is present or absent based on x-ray images. In this dissertation, we consider a binary detection task with two hypotheses: signals are present in the images, or signals are absent in the images.

In order to provide objective assessment of image quality and to avoid the uncertainty due to inter- and intra-human observer variability, model observers are often used to quantify image quality. A model observer is a decision making function that extracts information from images. Given an input image, the model observer calculates and returns a test statistic, which is a scalar value, according to its template. For the aforementioned detection task, given images under the two hypotheses, the test statistic values will generate two distributions: one from signal present images, and the other one from signal absent images. These two distributions are called probability density functions (PDFs).

By setting a threshold value, the model observer will be able compare test statistic values to the threshold and decide which of the two hypotheses (signal present or signal absent) is true for each image. Each threshold value determines a detection sensitivity, i. e., the True Positive Fraction (TPF) of decisions, and a detection specificity, one minus the False Positive Fraction (FPF) of decisions.

In detection tasks, the detection probability is often quantified by the Receiver Operating Characteristic (ROC) curve[27, 28]. This curve plots TPF as a function of FPF with moving the decision threshold value from positive infinity to negative infinity. The Area Under ROC Curve (AUC) determines the separability of the two distributions under different hypotheses, which represents detection probability of signals[29]. Fig.3.1 shows three ROC curves and the corresponding PDFs. $pr(t|H_i)$ represents the PDF for hypothesis i.

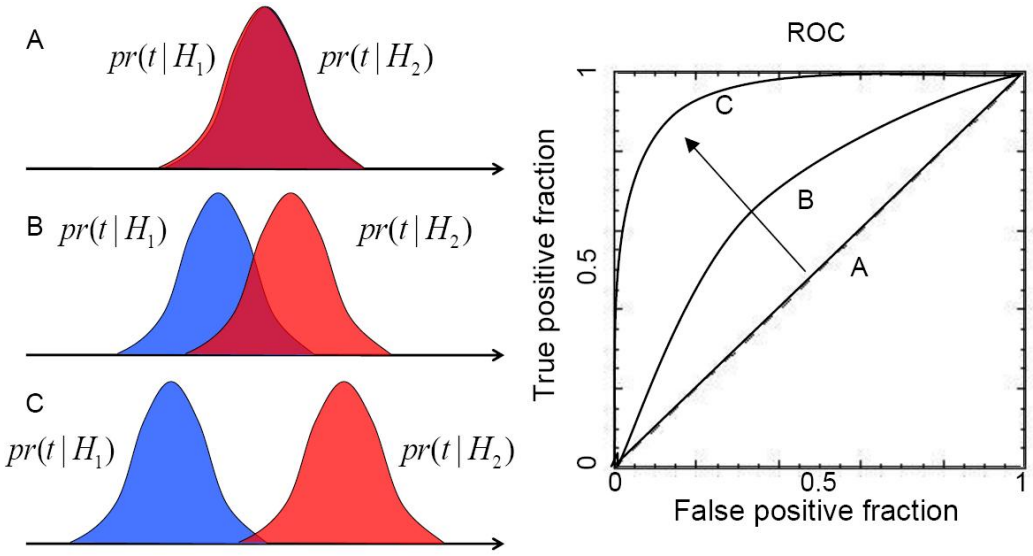


Fig. 1.1: The ROC curves and the corresponding PDFs. The separability of the two PDFs under different hypotheses determines the AUC, which represents detection probability.

System detectability is quantified by signal to noise ratio (SNR). Burgess *et al.* linked SNR to detection probability for multiple-alternative-forced-choice (MAFC) detection tasks[30]. For a 2AFC binary detection task, the link between SNR and AUC is thus given as:

$$\mathrm{AUC} = \frac{1}{2} + \frac{1}{2}\mathrm{Erf}\left(\frac{\mathrm{SNR}}{2}\right). \tag{1.2.1}$$

In general, higher SNR equivalents to better detection probability, and thus an infinitely large SNR corresponds to 100% detection probability. The SNR, which is directly linked to detection probability, can be used as a figure of merit to quantify image quality in detection tasks.

Consider the signal detection task with hypothesis H_1 corresponding to signal present images and hypothesis H_2 corresponding to signal absent images, the system detectability, or the SNR, for a model observer can be calculated as[29]:

$$\mathrm{SNR} = \frac{\bar{t}_2 - \bar{t}_1}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2}}, \tag{1.2.2}$$

where t_i is the test statistic of the observer when hypothesis H_i is true, $\bar{t}_i$ is the mean test statistic under hypothesis H_i, and σ_i^2 is the variance of test statistic under hypothesis H_i.

SNR can be calculated differently according to different observer models. Commonly used model observers for detection tasks include: the ideal observer, the non-prewhitening observer, and the Hotelling observer. Among all model observers, including both mathematical observers and human observers, the ideal Bayesian observer is a decision maker that yields the best possible performance of the imaging system[31]. It uses all statistical information to optimally perform the imaging task. The strategy of the ideal observer is to calculate a test statistic called the likelihood ratio, which is defined as the ratio of the PDFs under the two hypotheses[29]. In realistic imaging conditions, since the PDFs are usually unknown, it might be impossible to estimate its performance without sufficient statistics of the data[32]. Therefore, the Hotelling observer, which is an ideal linear observer that achieves the best performance of all observers constrained to linear operations on the data, is a more desirable alternative[32–34]. This observer is equivalent to the ideal Bayesian observer in detection tasks that only involve Gaussian data. The decision

Observer Model	**Test Statistic**	**System Detectability (SNR)**
Ideal Observer	$t(\mathbf{g}) = \frac{\mathrm{pr}(\mathbf{g}\|H_2)}{\mathrm{pr}(\mathbf{g}\|H_1)}$	No general mathematical expression
Non-prewhitening Observer (Spatial)	$t(\mathbf{g}) = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{g}$	$\mathrm{SNR}^2 = \frac{\left(\Delta\mathbf{S}^{\mathrm{t}}\Delta\mathbf{S}\right)^2}{\Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}\Delta\mathbf{S}}$
Non-prewhitening Observer (Frequency)	$t(\mathbf{g}) = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{g}$	$\mathrm{SNR}^2 = \frac{\left[\int \mathrm{d}\mathbf{f}\Delta\mathbf{S}(\mathbf{f})^2\mathrm{MTF}(\mathbf{f})^2\right]^2}{\int \mathrm{d}\mathbf{f}\Delta\mathbf{S}(\mathbf{f})^2\mathrm{MTF}(\mathbf{f})^2\mathrm{NPS}(\mathbf{f})}$
Hotelling Observer (Spatial)	$t(\mathbf{g}) = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\mathbf{g}$	$\mathrm{SNR}^2 = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\Delta\mathbf{S}$
Hotelling Observer (Frequency)	$t(\mathbf{g}) = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\mathbf{g}$	$\mathrm{SNR}^2 = \int \mathrm{d}\mathbf{f}\frac{\Delta\mathbf{S}(\mathbf{f})^2\mathrm{MTF}(\mathbf{f})^2}{\mathrm{NPS}(\mathbf{f})}$
Channelized Hotelling Observer	$t(\mathbf{V}) = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{U}\mathbf{K}_{\mathrm{c}}^{-1}\mathbf{V}$	$\mathrm{SNR}^2 = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{U}\mathbf{K}_{\mathrm{c}}^{-1}\Delta\mathbf{U}^{\mathrm{t}}\mathbf{S}$

Tab. 1.1: Commonly used model observers and their templates to calculate test statistics.

function of the Hotelling observer can be practically calculated. The upper bond to the system detectability is thus tractable and can be quantified by the Hotelling observer SNR. The channelized Hotelling observer applies a channel mechanism to the Hotelling observer to reduce the dimensionality of the calculation. The non-prewhitening observer is not designed to represent the best possible performance of the system, but to simulate human performance without prewhiten the noise correlations[32].

The model observer SNR definitions are summarized in Table.1.1 with notations defined in Table.1.2. The Hotelling observer uses the known signal as well as the noise covariance matrix $\mathbf{K}$ as a template to prewhiten noise correlations. The non-prewhitening observer uses only the known signal as the template without knowing any information about noise correlations. If the system noise is uncorrelated, the performance of the two observers is identical. In detection tasks with correlated system noise, the performance of the non-prewhitening observer is necessarily worse than the Hotelling observer. The channelized Hotelling observer uses the same observer template as the Hotelling observer, but partially captures the information within the images. Details of calculating different model observer SNR are discussed in later Chapters.

Notation	Definition
x	Discrete spatial dimension, parallel to chest wall
y	Discrete spatial dimension, perpendicular to chest wall
f_x	Discrete frequency along the x direction
f_y	Discrete frequency along the y direction
Δx	Pixel size along the x direction
Δy	Pixel size along the y direction
$\mathbf{g}$	Input image
$\mathbf{w}$	Ideal projection of an object
$\mathbf{n}_{\mathrm{q}}$	Quantum noise
$\mathbf{n}_{\mathrm{e}}$	Electronic noise
$\mathrm{pr}\,(\mathbf{g}\|H_{\mathrm{i}})$	Probability when hypothesis H_{i} is true
$\Delta\mathbf{S}$	Difference signal in spatial domain
$\Delta\mathbf{S}(\mathbf{f})$	Difference signal in frequency domain
$\mathbf{K}$	Noise covariance matrix in spatial domain
$\mathbf{H}$	System transfer function in spatial domain
R	Disk signal radius
h	Disk signal radius
μ	Linear attenuation coefficient
m	Magnification factor
NPS($\mathbf{f}$)	Noise power spectrum in frequency domain
GNPS($\mathbf{f}$)	Generalized noise power spectrum in frequency domain
MTF($\mathbf{f}$)	Modulation transfer function in frequency domain
GMTF($\mathbf{f}$)	Generalized modulation transfer function in frequency domain
GNEQ($\mathbf{f}$)	Generalized noise equivalent quanta in frequency domain
SNR	Signal to noise ratio
$\mathbf{U}$	Matrix of channels
$\mathbf{V}$	A channelized image
$\mathbf{K}_{\mathrm{c}}$	Channelized covariance matrix
$\mathrm{SNR}_{\mathrm{c}}$	Channelized Hotelling observer SNR

Tab. 1.2: Definitions of notations in calculating model observer SNR.

In this dissertation, we choose to use the Hotelling observer that provides the best possible system performance for the signal known exactly/background known exactly (SKE/BKE) detection task we considered. It maximizes the SNR of all observers that are constrained to linear operations including model observers and humans. Appendix B provides a mathematical proof of this statement.

1.3 Motivation

Evaluating the performance of FFDM systems involves the assessment of FFDM image acquisition, as well as the evaluation of image processing designed to help human observers extract information from FFDM images. Our motivation in this dissertation is to develop clinically practical methodology to evaluate both the image acquisition performance of clinical FFDM systems and the effects of image processing algorithms used in clinical FFDM on image quality.

In the evaluation of FFDM image acquisition, our goal is to develop a system-model-based Hotelling observer method that accounts for not only detector blur, but also scatter blur, focal spot unsharpness, magnification factor and noise correlations of the system. A complete description of the performance of the image acquisition part of clinical FFDM systems can be provided by this method with making one set of measurements. The image acquisition performance should be evaluated by this method with taking a reasonable amount of images. In a clinical setting, there are limits to the number of images one can obtain for testing purposes because of considerations of the x-ray tube and detector lifespan. Additionally, data collection restrictions are imposed by patient scheduling. This method should provide a means to build an empirical model of the FFDM system, which can be used to predict system performance at different doses without the needs of collecting additional images. A validation of this method needs to be provided by comparing its evaluation results to other published literatures.

In the evaluation of FFDM image processing, our goal is to develop a methodology that evaluates the effects of image processing on the best possible system performance.

Non-linear image processing/adaptive image processing algorithms are commonly used in clinical FFDM, this method should be able to take into account the effects of non-linear image processing on images. Because most of the non-linear image processing used in clinical FFDM adapts to different imaging regions, we need to design appropriate phantoms in the development of this methodology to trigger image processing (that will be otherwise turned off) including but not limited to edge enhancing and noise smoothing. This method needs to be practical in clinic as well, with the collection of limited number of images.

2. EVALUATION OF FFDM IMAGE ACQUISITION IN SPATIAL DOMAIN

2.1 Introduction

In Nov. 2010, the U.S. Food and Drug Administration (FDA) published a Class II Special Control Guidance for FFDM systems[35]. Any firm intending to market a new FFDM system needs to either show that its device meets the recommendations of this guidance or provide equivalent assurances of safety and effectiveness with respect to previously cleared FFDM systems. This guidance provides a recommended procedure for evaluating FFDM image acquisition performance and identifies the classification regulation. FFDM image acquisition is recommended to be evaluated through characterization of system properties including: the contrast between different tissues, the system noise properties, and the system spatial resolution, which refers to the ability of the system to resolve small details[36]. However, this evaluation of system properties has its limitations. Detector based bench tests are recommended in the guidance, such as Noise Power Spectrum (NPS), Modulation Transfer Function (MTF) and Detective Quantum Efficiency (DQE) measurements. These metrics do not take into account the effects of scatter from patients, focal spot unsharpness, and magnification factors because any phantoms that model x-ray attenuation of patient breasts are not involved in these measurements. As a complement, the guidance suggests a human observer-based phantom study using the Contrast-Detail Phantom for Mammography (CDMAM). The development of phantom is an outcome of the project: "Quality Assurance in Mammography", done at the Department of Radiology, University Medical Centre Nijmegen, (St. Radboud), the Netherlands. The CDMAM phantom consists of gold disks of various thicknesses and diameters, attached to an aluminum base

with up to four PMMA covers (1 cm thick each). The disks are distributed within a matrix of 16 rows and 16 columns. Each matrix cell contains two identical disks: one in the center and another one randomly located in one of the corners. Human observers are required to determine the locations of disks by reading the CDMAM images. The threshold thickness (for the least detectable disks) as a function of disk diameter, i. e., the Contrast-Detail (CD) Curve, can be obtained in the results to represent FFDM system performance. However, such a study might be affected by large uncertainties associated with inter- and intra-observer variability between humans. Furthermore, in this guidance, the link between the measurements of the physical characteristics of the system and the human observer studies is missing. Some expensive and time consuming clinical trials could be potentially avoided through such a link.

Many scientists have contributed to implementing the Hotelling observer model in the evaluation of FFDM image acquisition performance in an effort to provide objective assessment. Sandrik and Wagner[37] developed a Fourier based expression of the Hotelling observer SNR that they applied to a specific imaging task in the assessment of the performance of film systems. The Fourier approach makes the simplifying assumptions that signal transfer is shift-invariant and noise is cyclo-stationary. The Hotelling SNR can be expressed in terms of a difference signal (the difference between signal present and signal absent images) and the noise equivalent quanta (NEQ). The NEQ can be determined from the detector Modulation Transfer Function (MTF) and the Noise Power Spectrum (NPS)[37–39], which characterize the deterministic signal transfer of the detector and the system noise respectively. Furthermore, the detective quantum efficiency (DQE) obtained by normalizing NEQ with the photon fluence that reaches the detector, can be used as a figure of merit to quantify detector performance. The DQE quantifies the efficiency of a detector to detect x-ray photons[40]. Jennings *et al.* developed a database of measured mammographic x-ray spectra that can be used to determine the number of incident photons[41, 42] for the DQE calculation, and Dobbins *et al.* presented examples for calculating the DQE[42, 43] for different imaging systems. Many authors have used this Fourier

approach to evaluate the performance of mammographic systems[41, 44]. This method is based on measurements of the detector without the presence of a phantom to simulate realistic scatter and focal spot blurring conditions.

In an attempt to improve the Hotelling observer based evaluation method, a number of authors contributed to developing practical approaches for calculating the Hotelling observer SNR in spatial domain (defining the system properties in spatial domain) with phantoms placed in the Field of View (FOV). Gagne *et al.* calculated the Hotelling observer SNR with an image based methodology for a given SKE/BKE task to evaluate clinical FFDM systems[45]. By including a phantom in the evaluation, this method accounts for scatter from the phantom and focal spot unsharpness. However, the results are limited to the specific test objects and system settings. Since this approach does not create a model of the system, one must retake all the images in order to evaluate the system performance for other types of signals or at different technique parameters. Kyprianou described a method for building an empirical model of a bench-top imaging system, which models projection radiography, by analyzing the system response function and the system noise in spatial domain[46].

Inspired by Kyprianou's work, in this Chapter, we present an experimental methodology that calculates the Hotelling observer SNR in spatial domain as a figure of merit to evaluate the image acquisition performance of FFDM systems for a SKE/BKE detection task. We apply this evaluation method to a bench-top system that simulates projection mammography in our lab. This allows us to be able to make adjustments and improvements to the method prior to the application on clinical FFDM systems, so that we can take advantages of our limited time spent in the clinic. With placing a phantom in the FOV, it accounts for the scatter from the phantom, focal spot unsharpness and magnification factor in the estimation of image quality. Moreover, this method provides an empirical model of the system and thus allows for the prediction of system performance as well as for the optimization of technique parameters.

2.2 Materials and Methods

2.2.1 Theory

Based on the linearity assumption, system noise is considered an additive to the blurred image. The image $\mathbf{g}$ is thus defined as:

$$\mathbf{g} = \mathbf{H}\mathbf{w} + \mathbf{n}_{\mathrm{q}} + \mathbf{n}_{\mathrm{e}} \tag{2.2.1}$$

where the $\mathbf{H}$ matrix is an imaging operator, or system response function, that maps a discrete object phase-space to a discrete image. It transfers the projection through the imaging system and returns a blurred image. $\mathbf{w}$ is the ideal projection of the object at detector surface considering the geometry of the system and the magnification factor. $\mathbf{n}_{\mathrm{q}}$ is the quantum noise that is proportional to exposure, and $\mathbf{n}_{\mathrm{e}}$ represents the electronic noise. The notations used in this dissertation are all summarized in Table.1.2.

The test statistic $t(\mathbf{g})$ of the Hotelling observer is defined by the following equation[31]:

$$t(\mathbf{g}) = (\Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1})\mathbf{g} = [\mathbf{H}(\mathbf{w}_1 - \mathbf{w}_2)]^{\mathrm{t}}\mathbf{K}^{-1}\mathbf{g}, \tag{2.2.2}$$

where $\mathbf{K}$ is the covariance matrix that characterizes both the quantum noise and electronic noise of the system. It accounts for not only the variance of each pixel, but also the covariance between any two pixels. It is generally accepted that higher order of noise correlation can be ignored[31]. It is defined as: $\mathbf{K} = \langle(\mathbf{g} - \overline{\mathbf{g}})(\mathbf{g} - \overline{\mathbf{g}})^{\mathrm{t}}\rangle$, where $\overline{\mathbf{g}}$ is the noise-free average image. $\mathbf{w}_1$ is the projection of signal and background, and $\mathbf{w}_2$ is the projection of background only. $\Delta\mathbf{S}$ is therefore the expected difference signal between signal present and absent images. For instance, if using the CDMAM phantom, signal-present images are those containing gold disks while signal absent images are those with phantom background only. For each given image $\mathbf{g}$, the Hotelling observer test statistic $t(\mathbf{g})$ returns a scalar value. The returned test statistic values are normally distributed for both the signal present and absent input images.

The Hotelling observer SNR measures the overlap between the two distributions under different hypotheses, and it quantifies the system detectability. This SNR can be calculated following its definition in Eq.1.2.2[29].

The mean test statistic of t_{i} under hypothesis i for the Hotelling observer is:

$$\langle t(\mathbf{g})|H_{\mathrm{i}}\rangle = \langle \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\mathbf{g}|H_{\mathrm{i}}\rangle = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\mathbf{S}_{\mathrm{i}}. \tag{2.2.3}$$

The variance of t_{i} under the i^{th} hypothesis is:

$$\begin{aligned}\sigma_{\mathrm{i}}^2 &= \langle [t(\mathbf{g}) - \langle t(\mathbf{g})|H_{\mathrm{i}}\rangle]^2|H_{\mathrm{i}}\rangle \\ &= \langle \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^{\mathrm{t}}\mathbf{K}^{-1}\Delta\mathbf{S}|H_{\mathrm{i}}\rangle - \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\mathbf{S}_{\mathrm{i}}\mathbf{S}_{\mathrm{i}}^{\mathrm{t}}\mathbf{K}^{-1}\Delta\mathbf{S}.\end{aligned} \tag{2.2.4}$$

The variance is thus equal to:

$$\sigma_{\mathrm{i}}^2 = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\Delta\mathbf{S}. \tag{2.2.5}$$

If we put all the pieces together, the Hotelling observer SNR can be expressed as:

$$\mathrm{SNR}^2 = \frac{\Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\mathbf{S}_2 - \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\mathbf{S}_1}{\Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\Delta\mathbf{S}} = \frac{[\Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\Delta\mathbf{S}]^2}{\Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\Delta\mathbf{S}} = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\Delta\mathbf{S}. \tag{2.2.6}$$

In this Chapter, we consider disk-shaped gold signals when estimating the Hotelling observer SNR. The continuous phase-space of the disk signal is defined as:

$$\begin{aligned}&\Delta w(x, y; kVp, R, h, \mu_g, m) \\ &= X\sqrt{(\frac{x}{m})^2 + (\frac{y}{m})^2 < R^2}\Big(1 - \int_0^{kVp} dEpr(E; kVp)e^{-\mu_g(E)h}\Big),\end{aligned} \tag{2.2.7}$$

where pr(E; kVp) is the normalized x-ray spectrum for a specific kVp, h is the thickness

of the disk with radius R, μ_g is the energy-dependent linear attenuation coefficient of gold and m is the magnification factor. The continuous phase-space can be discretized and then used for the calculation of $\Delta\mathbf{S}$.

2.2.2 Methodology Development

Experimental Setup

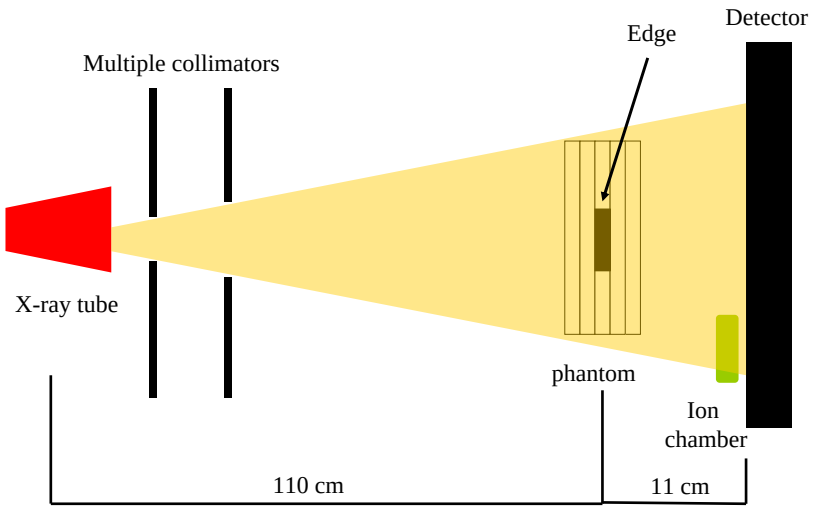


Fig. 2.1: The system setup with the uniform phantom assembly consisting of four PMMA plates (1 cm thick each) and a 0.5 mm thick Al base. Note that the edge test object for the determination of edge response function was placed between the second and the third PMMA plates, where the disk objects are supposed to be positioned.

In the attempt to develop the evaluation methodology for FFDM using the definition of the Hotelling observer in spatial domain, we consider a bench-top imaging system which models projection mammography in this Chapter. This allows us to test the method and optimize technique parameters so that when the method is applied to clinical FFDM

systems, their time and resources will be used efficiently. We use a Varian 4030CB (Varian Corp. Salt Lake City UT) detector with a 2048 $\times$ 1596 detection area, 195 $\times$ 195 μm pixels and 600 μm thick columnar CsI(Tl) scintillator. The detector is set at 14 bit mode without binning. The x-ray tube is a Varian B180 with 0.6 mm focal spot, Tungsten target and 1 mm Al inherent tube filtration. Tube voltage is set to the lowest available voltage 40 kVp in order to simulate clinical FFDM settings. Two collimators are used to collimate the incident x-ray cone beam to fit the shape of the phantom. The phantom is positioned 112 cm away from the x-ray tube (distance measured from the center of the phantom) and 11cm away from the detector. The corresponding magnification factor for the objects, which are placed in the center of the phantom, is 1.1. The phantom holder is made of plastic material to avoid generating additional scatter x-rays. Exposures are measured with a Rad Cal ion chamber (Radcal MOD 10X5-6) on the detector surface. The experimental setup is schematically shown in Fig.2.1.

We choose to evaluate image quality using the CDMAM phantom, since it is built according to the published data in the EUREF Type Testing Protocol relating PMMA thickness to equivalent breast thickness[47], and it is recommended in the FDA Class II Special Control Guidance for FFDM system. Automated software scoring and human observer reading are the available means for assessing the quality of the CDMAM images[48, 49]. However, the automated scoring method suffers from

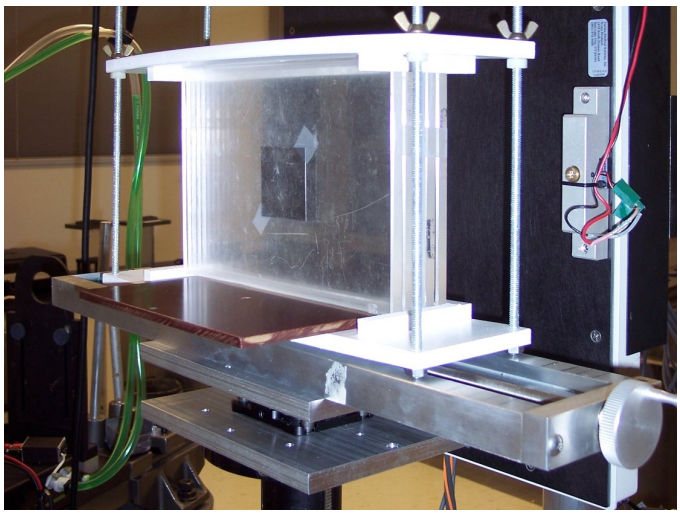


Fig. 2.2: The uniform phantom that models the background of the CDMAM phantom. In the center, a 5 cm$\times$5 cm, 0.25 mm thick Dysprosium (Dy) edge test object is taped and sandwiched between the second and the third PMMA plates.

limited statistics while the human observer reading is significantly affected by reader variability and reduced precision. These methods yield biased assessments of the system performance. It is therefore necessary to develop a task specific evaluation method that takes advantage of using the CDMAM phantom to provide objective and reliable evaluation results.

We designed a uniform phantom assembly, which simulates the background of the CDMAM phantom of version 3.2. Therefore, the two phantoms have the same attenuation, same noise properties and generate the same amount of scatter. Using this phantom allows us to separately analyze the system noise and the transfer function simulating a situation in which the CDMAM phantom is used. The CDMAM phantom consists of a 3 mm thick PMMA cover and a 0.5 mm thick Al base (99.5% aluminum) with gold disks (99.99% gold) of varying thicknesses (from 0.05 to 1.60 μm) and diameters (from 0.10 to 3.20 mm). The phantom itself has a PMMA equivalent thickness of 1 cm under standard mammography exposures, and it comes with four 16.25 $\times$ 24 cm PMMA plates, 1 cm thick each. To simulate the uniform background of the CDMAM phantom, we use four 1 cm thick PMMA plates and a 0.5 mm thick Aluminum base placed in the mid-distance between the four PMMA plates. The phantom assembly matches the size of the CDMAM phantom (162.5×240 mm) and its attenuation in the mammography energy spectrum. The CDMAM phantom has an Al Half Value Layer (HVL) of 1.108 mm, and the uniform phantom

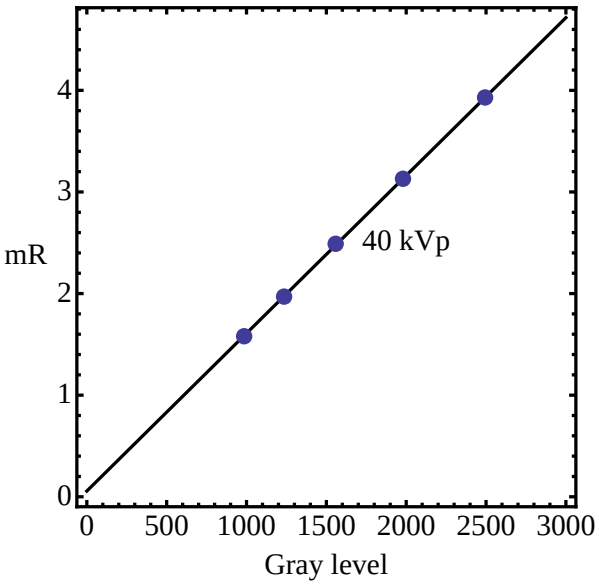


Fig. 2.3: Detector linearity at 40 kVp for exposures up to 5 mR.

assembly has an Al HVL of 1.113 mm at the lab settings.

Image Corrections

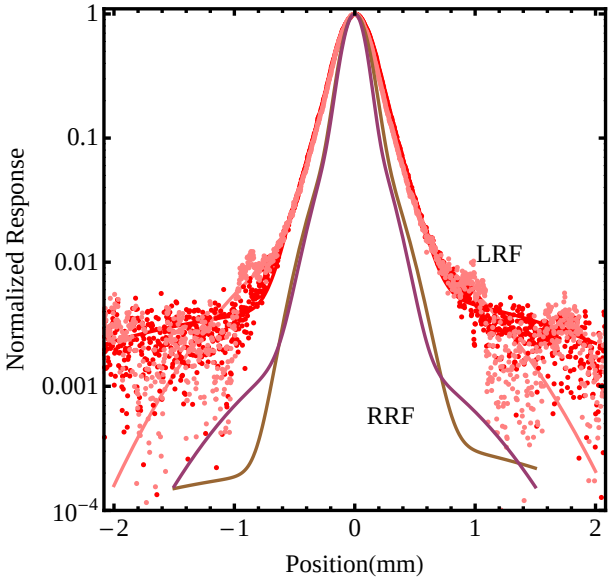
We performed the flat-field correction to raw images in order to remove the variations in the pixel-to-pixel sensitivity of the detector. The flat-field corrections consist of the correction for the nonuniform pixel gain and the subtraction of the dark-field (a mean of 100 dark-filed images was subtracted from each image).

We measured exposures and the corresponding grey-level pixel values for six different tube currents. Based on the measurements, we fit a linear relation between the exposure and grey-level pixel value. The linearity curve is shown in Fig.2.3. It thus allowed us to convert the image pixel values into meaningful exposures.

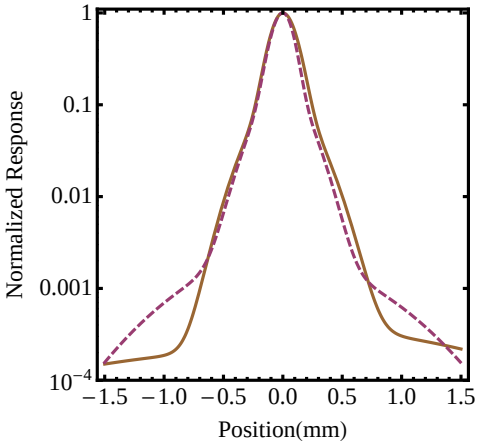
System Deterministic Properties

The detector transfer function is determined by the detector Ray Response Function (RRF), which is the detector response to a simulated infinitesimal x-ray beam. Kyprianou *et al.* developed an experimental method to estimate the detector RRF from a known detector Edge Response Function(ERF)[50]. We implemented this method with the improvement of placing a phantom in the FOV to take into account the scatter from phantom, the focal spot unsharpness and magnification factor, rather than only accounting for detector blur in the RRF. Therefore, in order to differentiate this RRF from the detector RRF, we name it the system RRF.

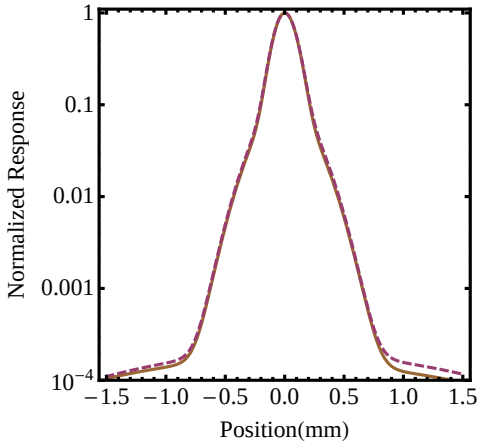
In an attempt to estimate the system RRF using the uniform phantom assembly, we placed a 5×5 cm square-shaped, 0.25 mm thick Dysprosium edge object between the second and the third PMMA plates counting from the detector, at the center area of the uniform phantom assembly. The edge was tilted 1.5 degree with respect to the horizontal axis as shown in Fig.2.2. We measured the ERF of the four edges and then obtained the average ERF. The system RRF was acquired from the ERF following the procedure described in Ref.50.



(a) The horizontal and vertical LRFs and the corresponding system RRFs at 40 kVp. The Dy edge was sandwiched between the second and the third PMMA plates.



(b) System RRFs at 40 kVp. The Dy edge was sandwiched between the second and the third PMMA plates. The dashed line represents the horizontal RRF, the solid line represents the vertical RRF.



(c) Detector RRFs at 40 kVp. The Dy edge was positioned directly at the detector surface. The dashed line represents the horizontal RRF, the solid line represents the vertical RRF.

Fig. 2.4: LRF and RRF measurements at 40 kVP. Test Dy edge was either taped directly at the detector surface or sandwiched between the second and the third PMMA plates.

We obtained the system response function, **H** matrix, by sub-pixel-simulated-scanning a 32×32 detector area with the system RRF and simultaneously recording the accumulated detector responses[50]. The system RRF was effectively sampled at four different locations within one pixel. The dimensions of the resulting **H** matrix are therefore 1024×4096.

System Noise Covariance Matrix

We estimated the system noise covariance matrix at 2.49 mR (detector entrance exposure) by taking 90 flat-field corrected images of the uniform phantom assembly. We selected a 384×384 central region within each image assuming that the system noise is stationary within this region. This central region was separated into 32×32 sub-regions and tiled in a 12×12 array. If $\mathbf{K}_{\mathrm{ROI}}$ represents the covariance matrix calculated for each Region of Interest(ROI), the average covariance matrix over all ROIs within each image, i. e., $\mathbf{K}_{\mathrm{g}}$, is thus defined as: $\mathbf{K}_{\mathrm{g}} = \langle \mathbf{K}_{\mathrm{ROI}} \rangle$. The overall average covariance matrix **K** is acquired by averaging $\mathbf{K}_{\mathrm{g}}$ over all 90 images. The dimensions of the overall average covariance matrix are 1024×1024.

Hotelling observer SNR

We simulated a series of gold disks of thickness from 0.05 μm to 1.6 μm and diameter from 0.1 mm to 3.2 mm as input signals inspired by the CDMAM phantom of version 3.2. The ideal projection of disk signals $\Delta\mathbf{w}$ was specified by the disk radius and height taking into account the energy spectrum dependent linear attenuation coefficient following the procedure of Ref.51. The Hotelling observer SNR was calculated following its definition for each disk signal.

2.3 Results

Fig.2.3 shows the linearity of the detector. It is used to convert the image pixel values into detector entrance exposures.

In Fig.2.4a, discrete points and the corresponding fits for the points represent the horizontal and vertical system Line Response Functions (LRFs). The system RRFs generated from the LRFs along these two directions are also shown in this figure. Fig.2.4b shows the system RRF measured along the two axes. The edge test object was placed in between the second and the third PMMA plates of the uniform phantom assembly. Fig.2.4c shows the detector RRF along the two axes when the edge test object was directly placed at the detector surface without placing the phantom in the FOV.

Fig.2.5 shows the $\mathbf{HH}^{\mathrm{t}}$ representing the system transfer function.

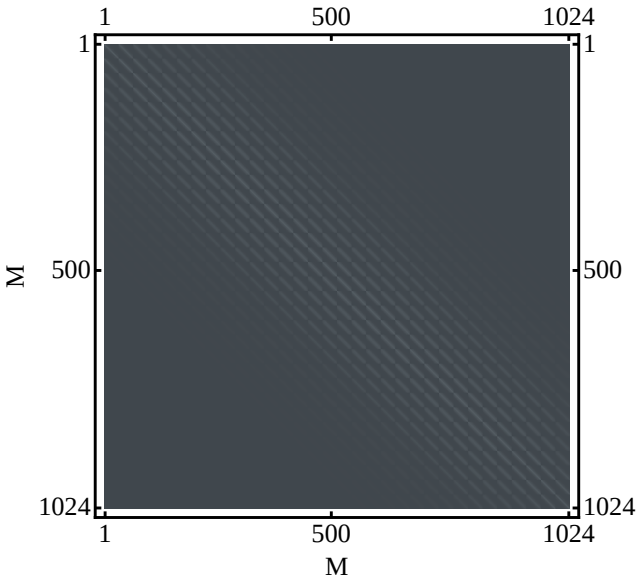


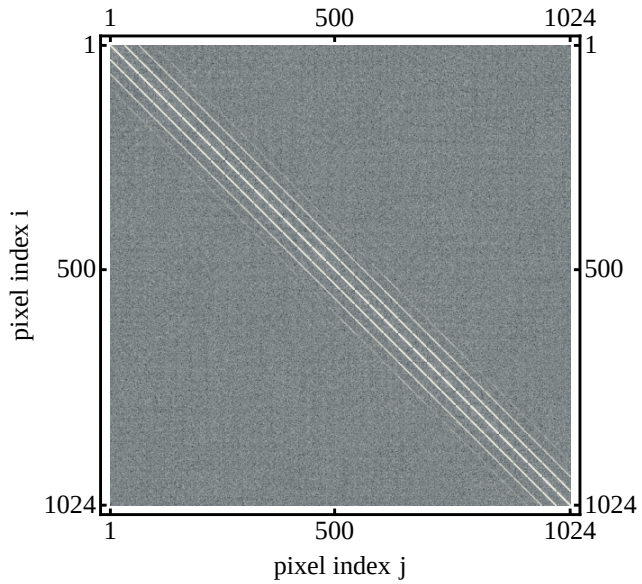
Fig. 2.5: The $\mathbf{HH}^{\mathrm{t}}$ of the system at 40 kVp, representing the system response function.

Fig.2.6a shows the overall average covariance matrix at a detector entrance exposure of 2.49 mR. Fig.2.6c shows the eigenvectors of the aforementioned covariance matrix. Fig.2.6b shows the first 10×10 partial-covariance matrix. The matrix element (i,j) shows the noise correlation between the i^{th} pixel and the j^{th} pixel. The correlations between the first pixel and its adjacent 6×6 neighbors are shown in Fig.2.6d.

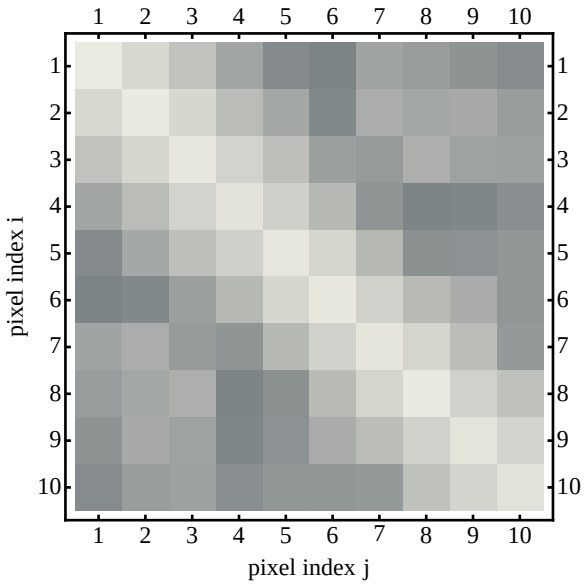
Fig.2.7 shows a sample simulation of the CDMAM disk signal with a thickness of 1 μm and a diameter of 2 mm. The two images on the left are the ideal projections (2D and 3D version) of the simulated disk, while the two images on the right are the corresponding images transferred through the system, blurred by the system response function.

Fig.2.8 shows the 2D Hotelling observer SNR map for different disk diameters and thicknesses at 2.49 mR. Fig.2.9 shows the Hotelling observer SNR as a function of: detector

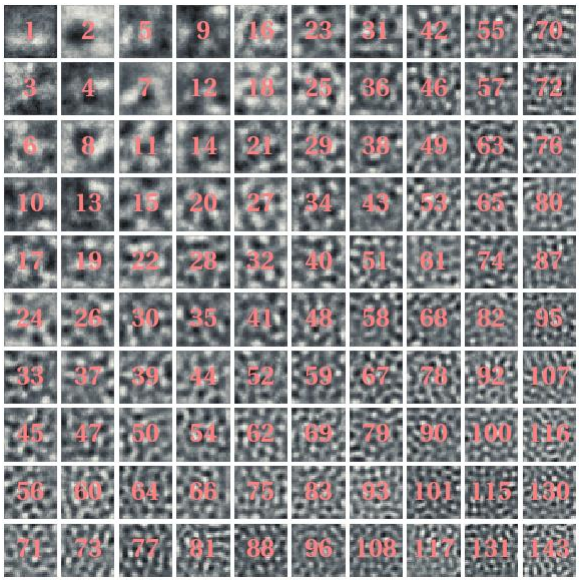
entrance exposure(2.9a), disk area (2.9b), disk thickness (2.9c) and disk radius (2.9d) respectively.



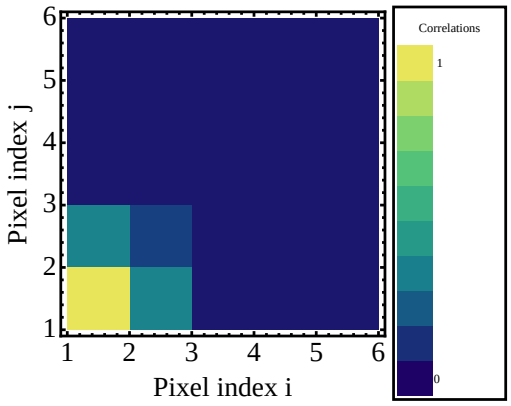
(a) The overall average covariance matrix **K**.



(b) The first 10×10 sub-matrix of the overall covariance matrix.



(c) Eigenvectors of the overall average covariance matrix.



(d) noise correlations of the first pixel with its 6×6 adjacent neighbors.

Fig. 2.6: Covariance matrix measured using the uniform phantom assembly, at 40 kVp and at the detector entrance exposure of 2.49 mR.

The covariance matrix consists of the quantum noise, which is proportional to exposure, and the electronic noise, which can be considered as an additive constant. Since the difference signal $\mathbf{\Delta S}$ is also proportional to exposure, the Hotelling observer SNR satisfies the following relation:

$$\text{SNR} \propto \sqrt{\frac{a\text{X}^2}{\text{X+b}}}, \tag{2.3.1}$$

where a is a constant, X represents the exposure and b represents the electronic noise. Using the estimated SNRs and the corresponding exposures in Fig.2.9a, we solved the constant a to be mR^{-1} and the constant b to be 93 μR.

2.4 Discussion

Placing the uniform phantom assembly in the FOV to take into account the scatter generated by the phantom and the focal spot unsharpness provides a more accurate estimation of the system performance. Without the phantom, the detector RRF shows no difference in system response between the horizontal and vertical directions. However, the system RRF shows that the system response along the horizontal direction has more blur than that along the vertical direction. One possible explanation is that the x-ray scatter generated from the phantom is not identical along the two directions due to the rectangular shape of the phantom. Another possibility is that the focal spot has a rectangular shape,

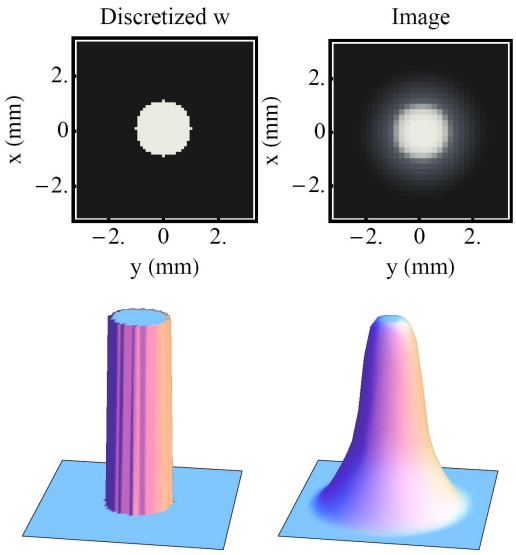


Fig. 2.7: Two images on the left: projection of a gold disk with 1 μm thickness and 2 mm diameter; two images on the right: the corresponding blurred image of the projection.

which also contributes to this difference. After adding a linear anti-scatter grid to the system, which is commonly used in clinical FFDM systems, we expect to see larger difference between the system response functions along the two axes, because the grid removes different amounts of scatter along these two directions. Therefore, this evaluation method provides a potential tool to evaluate the scatter removal method of the FFDM system. Details will be discussed in Chapter 3.

By simulating disk signals instead of taking images of the actual signals, we have the ability to predict system performance for other types of signals embedded in the same background without the collection of additional images. We could acquire as many images as we need using the bench-top system, however, it becomes problematic if we need to eventually apply the method to clinical FFDM systems since our access to clinical FFDM systems is limited by the system lifetime as well as unpredictable patient scheduling. Therefore, the evaluation method becomes more clinically practical when using simulated signals.

In our signal simulation, we have accounted for the x-ray spectrum, the geometry of the system setting, the material, size and thickness of signals in the ideal projection. Using the system response function $\mathbf{H}$ matrix derived from the system RRF, we take into account the system blur and eventually obtain the blurred image of the projection as shown in Fig.2.7. We choose to estimate the system response function within a 32×32 small detector region because our computation is limited by the large dimension of the $\mathbf{H}$ matrix. Since we use the sub-pixel-simulated-scanning method, the dimension of the $\mathbf{H}$ matrix is 1024×4096 for a 32×32 detection region. It also explains why we choose to show the Figure of $\mathbf{H}\mathbf{H}^{t}$ to represent the system response instead of showing the $\mathbf{H}$ matrix.

The noise covariance matrix accounts for the noise correlations between any two pixels. The diagonal of the covariance matrix represents the variance of each pixel, and the off-diagonal elements are the covariances between pixels. The variance of each pixel is higher than the covariance with any other pixels, therefore the diagonal of the covariance matrix is the brightest line in Fig.2.6a. However, the noise correlations between pixels, especially the

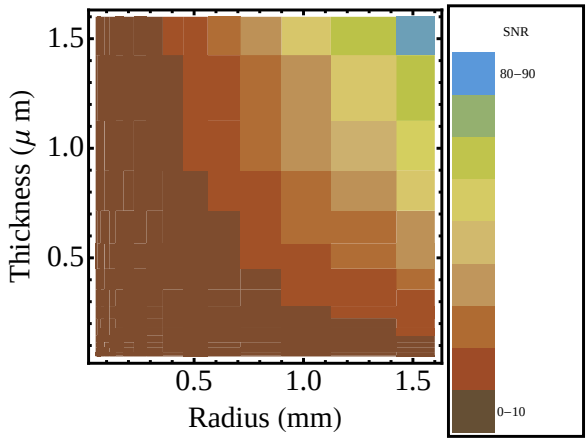
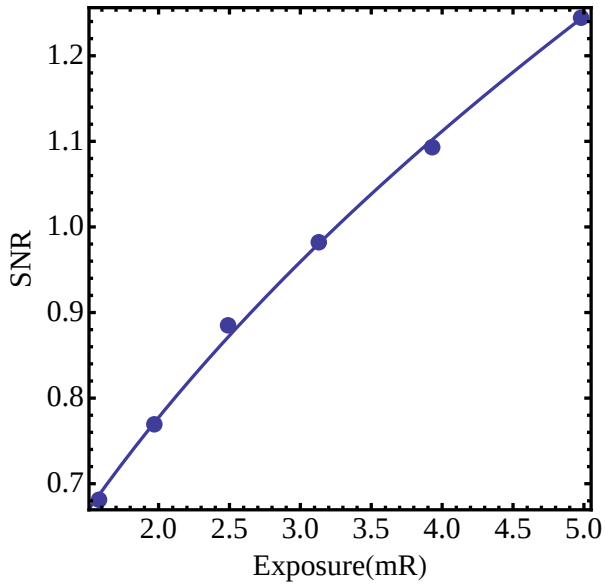


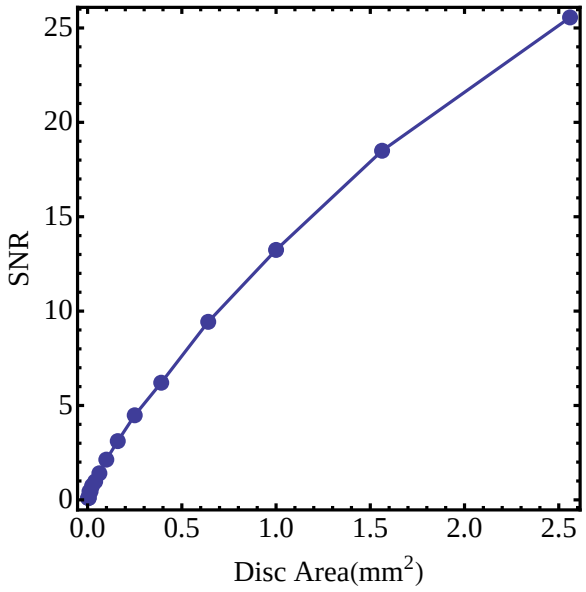
Fig. 2.8: 2D Hotelling observer SNR map for different disk radiuses and thicknesses at 2.49 mR.

neighbor pixels, noticeably contribute to the system noise as shown in Fig.2.6d. Ignoring noise correlations would introduce uncertainties to the estimation of system noise. Ideally, with obtaining an infinite number of images, the eigenvectors of the covariance matrix exactly describe the basis functions of the space. In practice, since limited by the available number of images, assuming sinusoidal-basis-functions is a desirable alternative to reduce the required number of images.

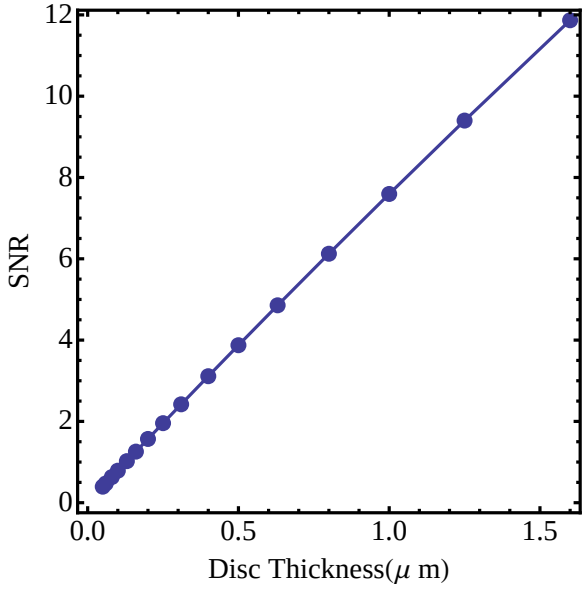
The Hotelling observer SNR as a function of disk diameter/area shown in the results matches the theory described in the literature[25, 52].



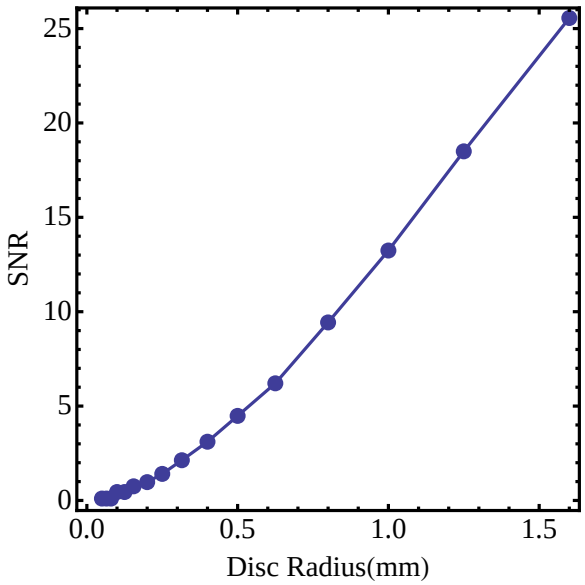
(a) The Hotelling observer SNR as a function of detector entrance exposure for a 0.25 μm thick disk with 0.5 mm diameter.



(b) The Hotelling observer SNR as a function of disk area with fixed disk thickness 0.4 μm.



(c) The Hotelling observer SNR as a function of disk thickness with fixed disk diameter 0.8 mm.



(d) The Hotelling observer SNR as a function of disk diameter with fixed disk thickness 0.4 μm.

Fig. 2.9: The estimated Hotelling observer SNR at 2.49 mR.

2.5 Conclusion

In this Chapter, we developed an experimental method that uses the Hotelling observer SNR as a figure of merit to evaluate the performance of mammographic systems. It calculates the system noise and transfer function independent of signals, and thus provides an empirical model of the system. Therefore, the evaluation results can be used to predict system performance at any exposure with any type of signal embedded in the same background. Although we only applied this method to a bench-top system in our lab, which simulates a projection mammography, the method is general and can be used to evaluate clinical FFDM. However, it requires a huge number of images in order to provide an accurate estimation of noise covariance matrix. A desirable alternative is to extend the method into Fourier frequency domain to reduce the required number of images and to make the methodology practical in clinic. Details will be discussed in the following Chapter.

3. EVALUATION OF FFDM IMAGE ACQUISITION IN FREQUENCY DOMAIN

3.1 Introduction and Background

In Chapter 2, we have developed a task specific system-model-based evaluation methodology that calculates the Hotelling observer SNR in spatial domain to quantify the FFDM image acquisition performance. And we have applied it on a bench-top system which models projection mammography in our lab. However,in order to ensure the estimation accuracy, this method requires a fairly large number of images. In a clinical setting, there are limits to the number of images one can obtain for testing purposes because of considerations of the x-ray tube and detector lifespan. Additionally, data collection restrictions are imposed by patient scheduling. Describing system properties in Fourier frequency domain has the advantage that it requires much fewer images because the shift invariance and stationarity assumptions allow more averaging. The Fourier based approach becomes more attractive for evaluating FFDM systems in a clinical setting as long as the implications of the assumptions made are fully understood. It is thus important to develop a clinically applicable task specific evaluation methodology in frequency domain to perform a complete assessment of the FFDM image acquisition system. Similarly, this method will be able to provide an empirical model of the FFDM system, which allows for the prediction of system performance. And it will take into account the noise correlations of the system, as well as signal size, shape, pixel size, focal spot unsharpness and scatter from patients.

In order to use the Fourier approach for evaluating the image acquisition part of FFDM systems by taking into account the effects of focal spot unsharpness and patient scatter, many scientists have contributed to generalizing the definitions of the MTF, NPS, NEQ

and the Hotelling observer SNR. Wagner investigated the effects of system noise and patient scatter on system detectability[53–55] recognizing the necessity of developing a comprehensive assessment approach that takes into account these effects[37]. Muntz studied focal spot size, magnification factor, patient scatter and described their effects on image quality by combining them in a comprehensive function[56]. Boone *et al.* introduced the scatter MTF that measures the spatial distribution of scatter[57]. Following in the footsteps of this research, Cooper *et al.* proposed an experimental methodology to measure the magnitude and spatial distribution of scattered radiations for different clinically relevant configurations[58]. Similarly, Doi *et al.* initially studied the focal spot unsharpness[59], Shaw *et al.* introduced the focal spot MTF[60], and Ganguly implemented the scatter MTF in a clinically relevant application[61]. In an effort to describe the complete system performance with considering both the focal spot unsharpness and scatter, Kyprianou *et al.* analytically generalized the MTF definition by separately defining the scatter MTF, the focal spot MTF and the detector MTF, generalized the NPS definition by considering the magnification factor and derived the generalized NEQ based on these two definitions[51, 62, 63]. For a microangiographic system, Samei developed an experimental methodology for estimating the effective DQE, which accounts for the scatter magnitude, and evaluated its use in selected digital radiographic imaging systems[64].

The task specific Signal to Noise Ratio (SNR) provides an objective assessment of image quality. A model observer is a decision making function that extracts information from an image, evaluates a test statistic, and compares it with a threshold value, to decide which of two populations (i.e. signal present or signal absent in a Signal Known Exactly/Background Known Exactly (SKE/BKE) detection task) an image belongs to. The ideal Bayesian observer is a decision maker that yields the best possible performance of the imaging system[31]. It uses all statistical information to optimally perform the imaging task. The ideal linear, or Hotelling, observer achieves the best performance of any observer constrained to linear operations on the data. It is a more desirable alternative[32–34], because this observer is equivalent to the ideal Bayesian observer in detection tasks

that involve Gaussian data, and its decision function can be practically calculated.

A number of authors have contributed to the development of practical approaches for calculating the Hotelling observer SNR. Sandrik and Wagner[37] first introduced an expression of the Hotelling observer SNR derived in Fourier domain to estimate the performance of a film system. Gagne *et al.* calculated the Hotelling observer SNR with an image-based methodology for an SKE/BKE task to evaluate clinical FFDM systems[45]. By including a phantom in the evaluation, this method accounts for scatter from the phantom and focal spot unsharpness. Kyprianou and Liu described a method for building an empirical model of a bench-top imaging system, which models projection radiography and mammography respectively, by analyzing the system response function and the system noise[46, 65]. In a more recent study, Monnin *et al.* used a non-prewhitened model observer with an eye filter (NPWE) in an effort to model the human observer and estimated the system noise with a phantom placed in the FOV.[66] Their NPWE observer made use of the the system noise evaluated with a uniform phantom (in order to account for the scatter noise) as well as the detector MTF.

In this Chapter, we built upon previous work to develop an experimental methodology for evaluating the performance of clinical FFDM image acquisition chain. The method uses an empirical model of the system to calculate the task specific System-Model-Based Hotelling Observer SNR in the spatial frequency domain (SMFHO SNR). The resulting SNR was used to generate contrast-detail (CD) curves in order to fascilitate the comparison between systems and methods. We demonstrate the practicality and clinical applicability of the method on two clinical FFDM systems (the GE Senographe DS and the Hologic Selenia) and we compare our results with other published methods.

3.2 Materials and Methods

Fig.3.1 summarizes our approach for developing an empirical system model observer in order to generate the CD curves of the image aquisition chain of FFDM. A model of the

image acquisition chain of the system is generated from the GMTF and GNNPS evaluated using a uniform phantom. The Hotelling observer, defined in terms of the GMTF and GNPS and a set of simulated signals, is used to estimate system detectability. From the Hotelling SNR the detection probabilities of disk signals of the CDMAM phantom can be estimated and CD curves can be generated by setting a threshold probability.

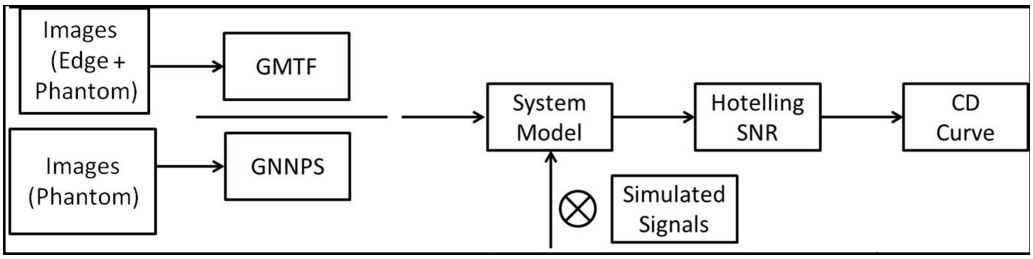
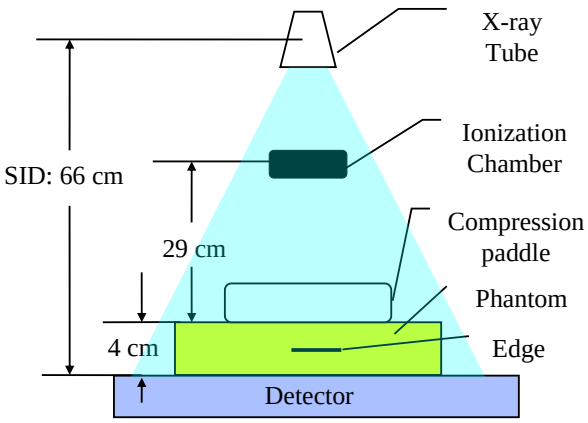


Fig. 3.1: Evaluation procedure of the SMFHO methodology.

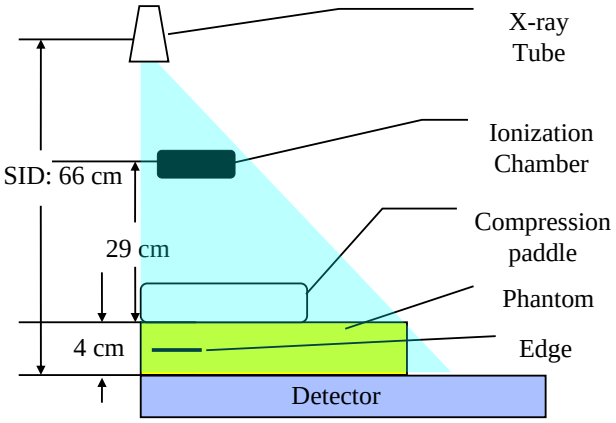
3.2.1 FFDM System Descriptions

In this Chapter we evaluate two clinical FFDM systems with different underlying detector technologies and scatter removal grids: the GE Senographe DS located at National Naval Medical Center, Bethesda, MD and the Hologic Selenia at Sibley Memorial Hospital, Washington DC. The schematic in Fig.3.2 demonstrates the acquisition geometry of both systems.

The GE Senographe DS FFDM system has two focal spots, 0.1 and 0.3 mm nominal, indirect CsI scintillator based digital detector, FOV of 19×23 cm (1920×2304 pixel), pixel size of 100 μm, kVp range 22 to 49, mAs range 4 to 500 and source to image distance (SID) of 66 cm. A 31 line pair/cm moving linear grid is built into the Bucky that covers the detector. The Bucky can be removed when required by our experiments. The system has two imaging modes. The *fine view* mode, uses proprietary sofware to filter the "for-processing" images (images before processing for display). The *standard* mode only performs the basic corrections for flat-field, dark-field bad pixels and gain. For simplicity, the *fine view* mode will be referred to as imaging mode A while the *standard* mode will



(a) The front view of the system schematic. A copper edge was placed in between the phantom slabs for measuring the GMTF.



(b) The side view of the system schematic. A copper edge was placed in between the phantom slabs for measuring the GMTF.

Fig. 3.2: Schematic of the GE Senographe DS and Hologic Selenia FFDM systems showing the X-ray tube, phantom assembly and detector.

be referred to as imaging mode B.

In the GE Senographe DS system, scintillators are used to convert x-ray photons into visible lights.X-rays are detected by a layer of phosphor, typically thallium (Tl)-activated CsI phosphor. Directly behind the phosphor layer, there is a layer of amorphous silicon, on which an array of photodiodes is used to record the emitted light photons. The photodiodes generate electrical signals in proportion to the absorbed light photons. These electrical signals will be amplified and recorded by the electronic system behind the layer of photodiodes[67]. The biggest limitation of this technology is the detector blur due to lateral spread of light photons. GE uses needle-like elements to channel the light photons in an effort to control the lateral spread[68].

The Hologic Selenia FFDM system has two focal spots, 0.1 and 0.3 mm nominal, selenium based direct digital detector, FOV of 24×29 cm (3428×4096 pixel), pixel size of 70 μm, kVp range from 20 to 39, mAs range from 3 to 400, the LORAD HTC anti-scatter

grid[1] and SID of 66 cm. For our experiments we used the "Phantom" mode which is typically used for QA purposes.

In the Hologic Selenia system, the detector directly converts the received x-rays to electronic charges. Amorphous selenium is used to absorb x-ray photons. When this material interacts with x-ray photons, an electric charge is liberated in the form of electron-hole pairs. A electric field is applied to control the lateral spread of electron-hole pairs and electrode pads are used to collect charges. Comparing to the indirect detection, the FFDM system uses this technology has better lateral resolution. However, the selenium based detectors are usually made thicker than indirect detectors in order to increase the detection efficiency. In addition, the cost is higher than indirect detectors[70].

For both systems, we performed all of our experiments with the large focal spot, molybdenum target, molybdenum filter, 30 kVp and three tube outputs: 20 mAs, 100 mAs and 200 mAs.

The uniform phantom we used consists of four PMMA plates (10 mm thick each), a 3 mm thick PMMA cover and an aluminum sheet (0.5 mm thick) placed midway between the four PMMA plates. The phantom assembly matches the size and thickness of the Contrast Detail Mammography (CDMAM) phantom[71], version 3.4 (Capintec, Inc., NJ) (162.5×240 mm) as well as its attenuation for the typical mammography energy range. Both phantoms have 5 cm equivalent thickness of PMMA. According to Dance *et al.*[72], this PMMA thickness is equivalent to a 6 cm thick breast with 20% glandularity. We measured and compared the Al half value layer (HVL) for this phantom assembly and for the CDMAM phantom on the GE Senographe DS system set at 100 mAs. The CDMAM phantom HVL was 0.676 mm, while the uniform phantom HVL was 0.673 mm.

For both systems, we measured detector entrance exposures by placing the phantom as close as possible to the x-ray tube and placing an ionization chamber between the phantom and the detector surface, as shown in Fig.3.2, to avoid back-scatter. We fit a linear relation between tube output (mAs) and the measured exposures (mR), after correcting for the

[1] The HTC grid has a cross-hatch design that reduces scatter in two directions[69].

distance from the detector using the inverse square law. This way, we obtained the scatter free detector entrance exposures for a given tube output. Note that when the phantom is placed near the detector the detector entrance exposure will be higher due to scatter from the phantom. The detector entrance exposures are used to compare our evaluation results to those from the literature.

We estimated the mean glandular breast dose (MGD) by implementing the method described in Refs.73 and 72, assuming a 6 cm thick standard breast with 20 percent glandularity. For the three tube outputs we considered, the MGDs are 0.53 mGy, 2.71 mGy and 5.42 mGy for the GE Senographe DS and 0.49 mGy, 2.87 mGy and 4.97 mGy for the Hologic Selenia system.

3.2.2 Detector MTF and Generalized MTF evaluation

The MTF describes the spatial resolution of the detector. It is defined as the normalized modulus of the Fourier transformation of the Line Response Function (LRF)[31, 74, 75]. In the attempt to fully characterize a clinical FFDM imaging system with regards to the detector blur, focal spot unsharpness due to magnification and the scatter from the phantom, we estimated the Generalized Modulation Transfer Function (GMTF)[62, 63, 76, 77]:

$$\mathrm{GMTF}(f_x, 0) = \mathrm{GMTF}(f) = \frac{\left| \sum_{i=1}^{256} \mathrm{LRF}(x_i) e^{-\frac{2\pi i f x_i}{256}} \right|}{\sum_{i=1}^{256} \mathrm{LRF}(x_i)}, \tag{3.2.1}$$

where f is the discrete frequency (1/mm) and x is the discrete spatial dimension in mm. Eq.3.2.1 shows the 1D GMTF along the x axis as one slice of the 2D GMTF. In our experiments, the x direction is defined as parallel to chest wall. The phrase "Generalized" is used to identify that the GMTF was estimated with the uniform breast phantom in place and that it is different from the traditional detector MTF.

For both systems, an edge test object (2.5 cm × 2.5 cm) was placed in the mid-distance between the four PMMA plates (the location where the disk signals inside the CDMAM phantom would have been positioned) 1.5 cm away from the chest side where the incident X-ray beam is almost perpendicular. Note that clinical systems commonly set very low pixel values to zero in order to reduce electronic noise, therefore high attenuation materials such as Dysprosium or lead, suggested in the Refs.65 and 50, could not be used in our experiments. Instead, a 0.14 mm thick copper plate was chosen as the edge test object because it does not attenuate the beam completely. To reduce noise, an average of five images of the phantom assembly was taken at each exposure, and a second order polynomial fit (avoiding the test object area) was subtracted from the mean image to correct for the heel effect. Following the procedure described in Refs.65 and 50, the Edge Response Function (ERF) and the LRF were estimated using the corrected average image while the pre-sampled GMTF was derived from the LRF[50]. For the GE Senographe DS system, the pixel size at the detector plane was estimated to be 100 μm, the magnification factor from the center of the phantom to the detector was 1.03, hence the pixel size at the object plane was 97 μm, and the Nyquist frequency was 5.15 mm^{-1}. For the Hologic Selenia system, the pixel size was 68 μm at the object plane and the Nyquist frequency was 7.35 mm^{-1}. The zero frequency in the Fourier space corresponds to infinite length in the spatial domain. We estimated the zero frequency GMTF by linear extrapolation of the first few points of the GMTF. To reduce noise in the GMTF data, we performed a piecewise fit to the GMTF. The low frequency drop section of the curve represents the scatter MTF. Since the drop is very steep[63], we were only able to obtain three data points. And a line was thus fit to this section. A twentieth order polynomial was fit to the remaining points. In order to obtain the 2D GMTF, two profiles (along x and y directions) of the 2D asymmetric GMTF were acquired from the LRFs, and the 2D GMTF was obtained by fitting cubic splines between the horizontal and vertical 1D GMTF profiles. For a given GMTF magnitude GMTF_0, four points on the x and y directions that satisfy $\mathrm{GMTF}(f_x, f_y) = \mathrm{GMTF}_0$ were located, and a cubic spline was fit between the

four points. This procedure was repeated for 200 different magnitudes. It should be noted that the results can be trusted only along the axis directions, and the interpolation between measured points provides an estimate of the 2D GMTF. "Boot-strapping" with replacement (the idea was introduced and explained in Refs.45) was used to estimate the mean and standard deviation of the GMTF with finite number of samples. For the same edge, we derived 10 LRFs. We then created a new set of LRFs by randomly picking 10 LRFs from the original set and we calculated their mean. By repeating this procedure 100 times, we were able to collect 100 mean LRFs. From each mean LRF, we can derive a corresponding 1D GMTF. And from the whole set of mean LRFs, we were able to estimate the mean and variance of the GMTF. Two times of the standard deviation was used as error bar.

3.2.3 NPS and Generalized Normalized NPS evaluation

The detector based Normalized Noise Power Spectrum (NNPS) was acquired by normalizing the detector NPS with the squared average pixel value for a given exposure[78]. Since both the primary and scatter X-rays exiting patients contribute to the noise recorded by the detector, we estimated the Generalized NNPS (GNNPS)[62] with the uniform phantom assembly placed on the detector underneath the compression paddle (the typical location of a breast):

$$\mathrm{GNNPS}(f_x, f_y) = \frac{\mathrm{GNPS}(f_x, f_y)}{\mathrm{d}^2}, \tag{3.2.2}$$

where the average detector output signal is given as d and the GNPS can be calculated by:

$$\mathrm{GNPS}(f_x, f_y) = \frac{\Delta x \Delta y}{\mathrm{N} \times 256 \times 256} \sum_{n=1}^{\mathrm{N}} \left| \sum_{i=1}^{256} \sum_{j=1}^{256} \left(\mathrm{I}_n\left(x_i, y_j\right) - \mathrm{S}_n\left(x_i, y_j\right)\right) e^{-\frac{2\pi i f_x, f_y x_i y_j}{256 \times 256}} \right|^2. \tag{3.2.3}$$

$I_n(x_i, y_j)$ is the nth Region Of Interest (ROI) of the flat field corrected images of the uniform background and S is the 2D second-order polynomial fit of this image. N is the total number of ROIs (number of ROIs within one image times number of images). Δx and Δy are the physical dimensions of the detector pixels.

We acquired five images of the uniform phantom assembly. From each image, a 640 × 640 pixel region at the same location where the edge was placed for the MTF calculation was selected and subdivided into 256 × 256 ROIs following refs.[62, 63, 78]. Each ROI overlapped by three quarters with its neighbors. This provides 80 ROIs in total from the five images to be used for the GNPS calculation. "Boot-strapping"[45, 51] with replacement was used to randomly sample ROI regions to obtain error bars for the GNNPS.

3.2.4 Hotelling observer SNR

To make the best possible estimate of the image acquisition performance of the two FFDM systems, we calculated the Hotelling observer SNR. The Hotelling observer integrates the frequency content of a signal, filtered and blurred by the system transfer function and hidden by the system noise. The Hotelling observer SNR is defined as[31, 62]:

$$\mathrm{SNR}^2 = \Delta x \Delta y \sum_{k=1}^{256} \sum_{l=1}^{256} \frac{\mathrm{GMTF}^2(f_x, f_y)}{\mathrm{GNNPS}(f_x, f_y)} \Delta \mathrm{S}_\mathrm{F}^2(f_{x_k}, f_{y_l}), \tag{3.2.4}$$

where $\Delta \mathrm{S}_\mathrm{F}(f_x, f_y)$ is the Fourier transformation of an object of interest, here a set of gold disks with thicknesses from 0.03 μm to 2.0 μm and diameters from 0.06 mm to 2.0 mm as input signals to match the targets in the CDMAM phantom. In the simulation, each disk signal was specified by signal size and thickness, taking into account the energy-spectrum-dependent linear attenuation coefficient following the procedure of Ref.51. The standard deviation of the Hotelling observer SNR was determined from the propagation of errors of the GMTF and the GNNPS.

In order to compare the SMFHO SNR with other published methods, we implemented three additional observer models on the GE Senographe DS system. In their image-based,

image-space method, Gagne *et. al.*[45] obtained both signal present and absent images using the CDMAM phantom. They generated the difference signal ΔS by subtracting sample averages of the two. They then estimated the covariance matrix which represents system noise and calculated the Hotelling observer $\mathrm{SNR}_{\mathrm{ISIB19}}$[1] which is defined as: $\mathrm{SNR}_{\mathrm{ISIB19}} = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\Delta\mathbf{S}$. To implement this method, we acquired five images at each exposure with the CDMAM phantom placed in the middle of the four uniform PMMA plates (between the second and the third PMMA slabs replacing the 0.5 mm Al sheet of the uniform phantom) in order to obtain signal present ROIs. Five images of the uniform background with the aluminum sheet replacing the CDMAM phantom were acquired to obtain the signal absent ROIs (5048 ROIs in total). To generate ΔS for each disk size, we averaged five 19×19 pixel signal present ROIs and then subtracted the average signal absent ROI. All 5048 19 × 19 signal absent ROIs were used to estimate the covariance matrix for each exposure. In order to investigate differences between Gagne's image-space, image-based method with Fourier-space, image-based methods we calculated $\mathrm{SNR}_{\mathrm{FSIB19}}$ and $\mathrm{SNR}_{\mathrm{FSIB256}}$ by dividing the square of the Fourier transformation of 19×19 and 256×256 pixel ΔS, by the appropriate 2D GNNPS. Because of its larger size the 256 × 256 GNNPS was calculated with only 80 ROIs, while the 19 × 19 GNNPS was calculated with 5048 ROIs.

3.2.5 Contrast-Detail Analysis

Contrast-Detail (CD) curves are commonly used to determine the boundary between visible and invisible objects[79, 80]. The CD curves were generated by determining the disk thickness-diameter pairs that result in a fixed detection probability. For our 4AFC experiments, the detection probability of 62.5% was chosen as the threshold probability because it is the mid distance between chance (25%) and 100% detection. In practice, we converted the threshold detection probability to a threshold SNR[30]. For each disk diameter,

[1] Note that the symbol SNR_{ISIB19} identifies the image-space, image-based SNR with 19×19 pixel ROIs using Gagne's method

because of the discrete disk thickness of the CDMAM phantom, we compared the 16 SNRs (corresponding to 16 thicknesses) to this threshold. The thicker one of the two disks that produce SNR closest to the threshold SNR was chosen as the threshold thickness. For example, if the threshold SNR is equal to 0.5, the 1.42 μm disk has a SNR of 0.3 and the the 2 μm disk has a SNR of 0.51, then 2 μm will be chosen as a threshold thickness. A CD curve was obtained as an interpolation between all 16 threshold thicknesses.

To determine the probability of detection when scoring CDMAM images, we considered a four alternative forced choice (4AFC) detection task, based on the visual detection task perfromed by the human readers of the CDMAM: choosing the correct signal location out of four possibilities within each cell of the CDMAM phantom. For the 4AFC imaging task, the detection probability p can be related to the SKE/BKE Hotelling observer SNR as described in Ref.30 by Burgess.

The CD curves were generated by determining the disk thickness-diameter pairs that result in a fixed detection probability. For our 4AFC experiments, the detection probability of 62.5% was chosen as the threshold because it is the midway between chance (25%) and 100% detection. In practice, we converted the threshold detection probability to a threshold SNR[30]. We simulated a series of gold disks with thicknesses from 0.03 μm to 2.0 μm and diameters from 0.06 mm to 2.0 mm as input signals to match the targets in the CDMAM phantom. For each disk diameter, we compared the 16 SNRs (corresponding to 16 thicknesses) to the threshold SNR. The thicker one of the two disks that produced an SNR closest to the threshold SNR was chosen as the threshold thickness. For example, if the threshold SNR is equal to 0.5, the 1.42 μm disk has a SNR of 0.3 and the the 2 μm disk has a SNR of 0.51, then 2 μm will be chosen as a threshold detectable thickness. CD curves were obtained by linear interpolation between all 16 threshold thicknesses.

To link the SMFHO CD curves to human reader performance, we estimated the human efficiency, defined as the human observer $\mathrm{SNR}^2_{\mathrm{Human}}$ divided by the Hotelling observer $\mathrm{SNR}^2_{\mathrm{SMFHO}}$. We obtained the estimated human performance by implementing an empirical method developed by Young *et al.*[81, 82]. Young's empirical model links the CD curves

generated by the CDCOM software[2] and those by human readers from three different imaging centers. We implemented Young's method on the GE Senographe DS system with imaging mode B for the three exposure levels (20 mAs, 100 mAs and 200 mAs). For each exposure level, we acquired eight images of the CDMAM phantom and used CDCOM software output and Young's model to obtain the estimated human CD curves. From the CD curves we obtained the estimated human observer SNR and compared it to the Hotelling observer SNR. The human efficiency was calculated by averaging over the three exposures.

3.3 Results

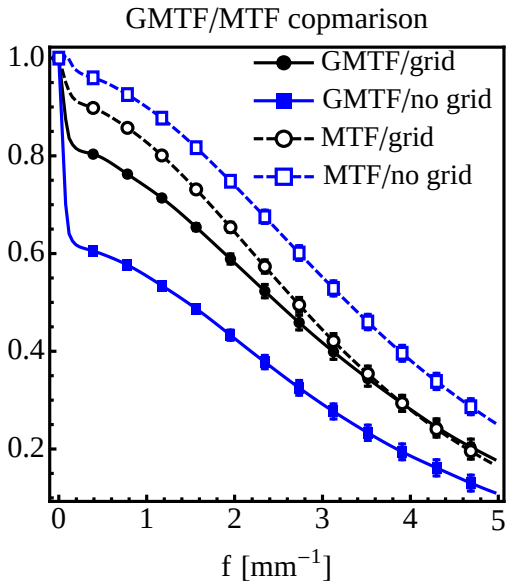
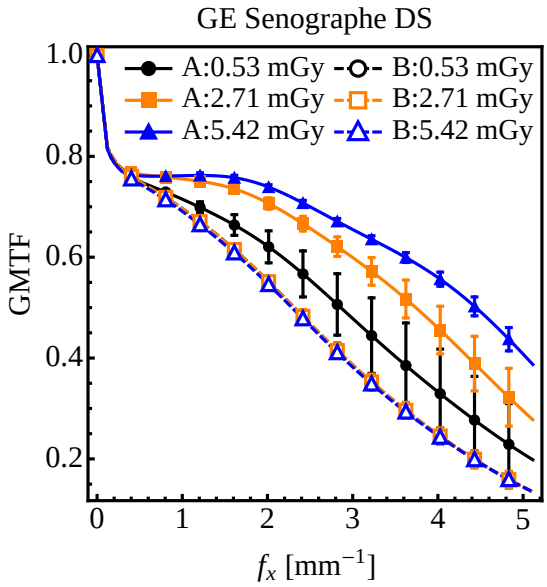


Fig. 3.3: Detector MTF (without phantom) and GMTF (with phantom in place) measurements for the GE Senographe DS system at 100 mAs tube output with imaging mode B. Results are shown with and without the grid installed on the system.

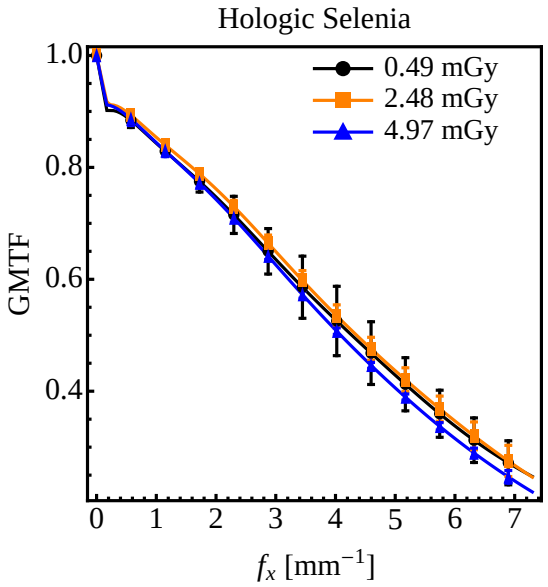
Fig.3.3 shows the comparison between the detector MTF (without phantom) and the GMTF (with the phantom in place) with and without the grid of the GE Senographe DS

[2] CDCOM[71] is a software program for scoring CDMAM images. It works by detecting the two disk signals within each cell (one in the center and the other in one of the four corners). The program compares the detection results to a pre-determined template and returns the detection probability.

FFDM system (mode B) measured at 100 mAs along the x-direction (parallel to the chest wall). The detector MTF was slightly higher when the grid was not installed. The low frequency drop observed when the grid was in place was due to scatter within the grid that degraded the detector MTF. The grid however, significantly improved the GMTF when the phantom was in place because, as evidenced by the low frequency drop, it reduced the scatter fraction from 38.6% to 18.4%.



(a) 1D GMTF for the GE Senographe DS system. Solid lines with filled symbols represent imaging mode A while dashed lines with empty symbols represent mode B.



(b) 1D GMTF for the Hologic Selenia system.

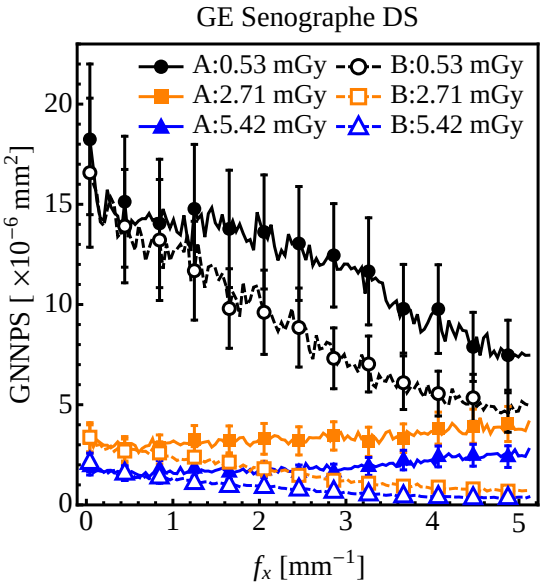
Fig. 3.4: 1D GMTF measured along the x axis (parallel to the chest wall), for the two systems. Three MGDs are denoted by different colors/symbols. The symbols, that correspond to every 10th point, are shown to help identify the curves. Error bars are twice the standard deviation.

Fig.3.4 shows the 1D GMTF derived from the LRF measured along the x-direction, for the GE Senographe DS system (3.4a), and the Hologic Selenia (3.4b) for different MGDs. In Fig.3.4a, solid lines with filled symbols represent imaging mode A while dashed lines with empty symbols represent mode B. The symbols are shown to help identify the curves and they do not correspond to specific points. In mode B, the GMTF did not significantly

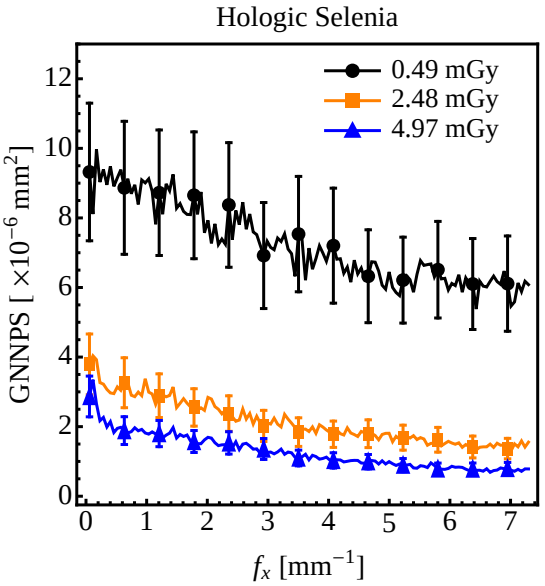
change with MGD. In mode A, the GMTFs were all higher than those in mode B for all MGDs due to possible image sharpening employed by mode A. Furthermore, the GMTF increased with higher MGD, indicating that higher sharpening was employed when the dose was increased. In Fig.3.4b, for the Hologic Selenia system, the GMTFs at three MGDs were identical within error bars. Although the plots were not shown here, we observed the same behaivior in the y-direction GMTF with respect to dose dependence. However, for the GE system the GMTF had a larger low frequency drop in the y-direction. This difference was due to the different amounts of scatter that reach the detector in the two directions: more scatter is removed along the x-axis, the axis perpendicular to the grid lines (scatter fraction: 18.4%±0.4%), than that along the y-axis (scatter fraction: 16.3%±0.2%). The Hologic Selenia grid removes the same amount of scatter along the two axes (scatter fraction: 9.8%±0.4%) because it employes a 2D cross hatched cell pattern. Note that the small increase in scatter we observed for the GE system along the y-direction only occurs for a very small solid angle (defined by the grid septa spacing). The rotationally averaged scatter fraction was estimated using the beam stop method (17.8%±0.2%) and was about halfway between the x and y scatter fraction estimates from the GMTF.

Fig.3.5 shows the 1D GNNPS for the GE Senographe DS system (3.5a), and the Hologic Selenia system (3.5b). For both systems, the GNNPS decreases with increased MGD. For the GE system, mode A has higher GNNPS than mode B, because the image sharpening algorithm designed to increase system resolution also increases system noise. Note that the numerical GNNPS range is similar between the two systems for similar doses.

Fig.3.6 shows the comparison between (A) SMFHO SNR, (B) SNR_{ISIB19}, (C) $SNR_{FSIB256}$ and (D) SNR_{FSIB19} for imaging task with a specific disk signal: 0.63 mm in diameter and 1 μm in thickness, for the GE Senographe DS system with imaging mode B at 2.71 mGy MGD. Within error bars, the SNR results are close to each other. Notice that the SMFHO SNR and its error bar vary noticeably with the number of ROIs. When using 256 $\times$ 256 pixel ROIs in our measurements, which provides us 80 ROIs, the SNR becomes stable (error bars are very small compared to those estimated with 10 ROIs) and changes very



(a) Profiles of GNNPS for the GE Senographe DS system along the x direction. Solid lines with filled symbols represent imaging mode A while dashed lines with empty symbols represent mode B.



(b) Profiles of GNNPS for the Hologic Selenia system along the x direction.

Fig. 3.5: Profiles of GNNPS along the x direction (parallel to the chest wall), for the two systems. Three MGDs are denoted by different colors/symbols. The symbols, that correspond to every 10th point, are shown to help identify the curves. Error bars are twice the standard deviation.

slowly.

In the CD analysis, we first normalized the SMFHO SNR^2 by MGD for all signals at three MGDs, since SNR^2 is proportional to MGD. We therefore obtained the SNR at 1 mGy MGD. Next, we averaged the SNR at 1 mGy MGD for the three data sets and created an SNR map for all 16 × 16 signals. Fig.3.7 shows the SMFHO CD curve comparison at 1 mGy MGD for the two systems. Notice that because the disk thicknesses are discrete, small fluctuations in the threshold SNR can cause the CD curve to jump to the next available disk thickness. This is a possible explanation for the difference between the CD curves of the two systems at smaller diameters.

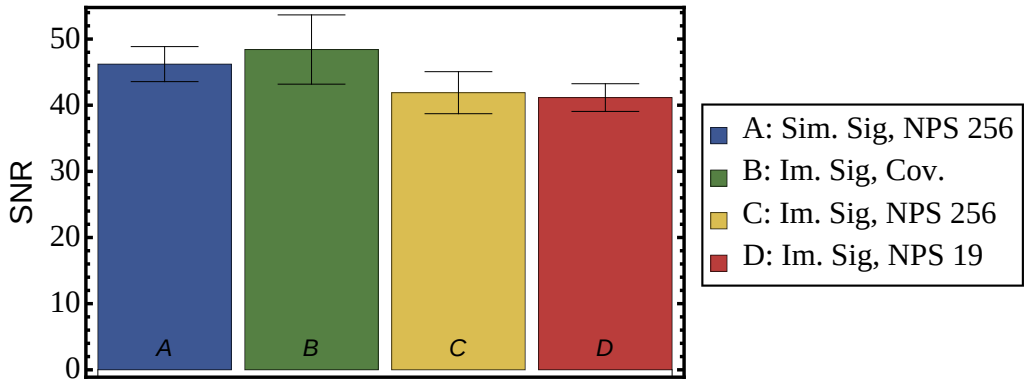


Fig. 3.6: GE Senographe DS system at 2.71 mGy MGD with imaging mode B. For one specific signal: 0.63 mm in diameter and 1 μm in thickness, the comparison between (A) SMFHO SNR, (B) $\text{SNR}_{\text{ISIB19}}$, (C) $\text{SNR}_{\text{FSIB256}}$ and (D) $\text{SNR}_{\text{FSIB19}}$.

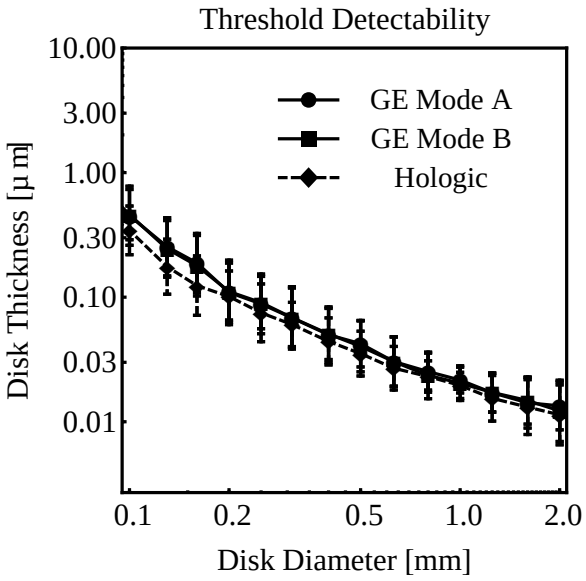
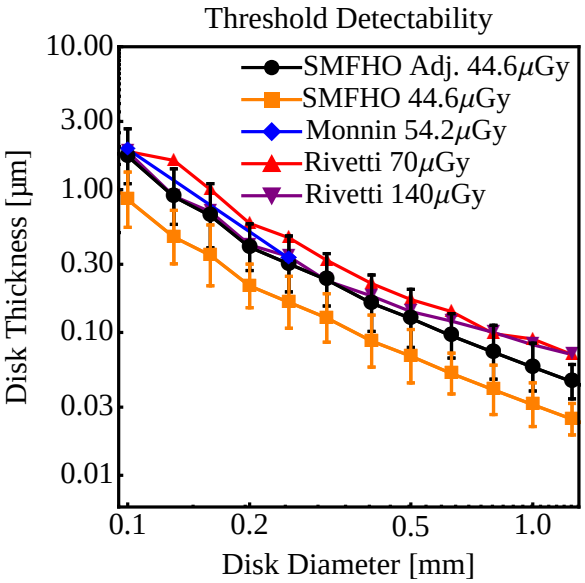


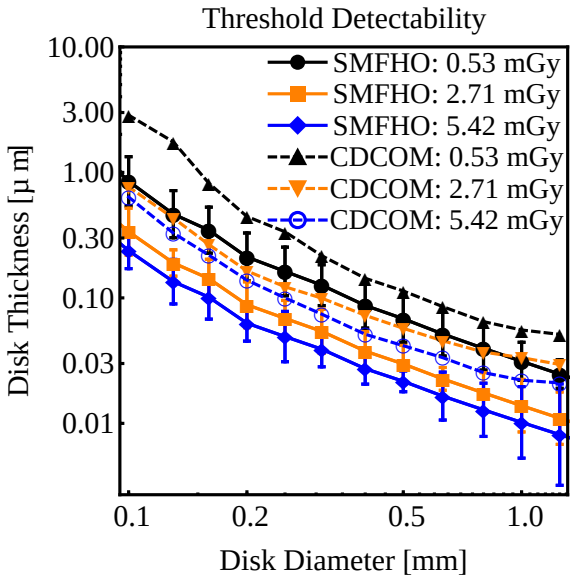
Fig. 3.7: Acquisition of average CD curves (generated from the SMFHO SNR) at 1 mGy MGD for the two systems, including two imaging modes of the GE Senographe DS system.

Fig.3.8 shows the comparison between CD curves acquired with different methods. In Fig.3.8a, we compare the SMFHO CD curves before and after adjusting by a mean human efficiency factor for the GE Senographe DS system, mode A, at 44.6 μGy detector entrance exposure (corresponding to 0.53 mGy MGD). The human efficiency was estimated by dividing the predicted human observer SNR^2, which was derived from the CDCOM

reading outputs as per the procedure described in Refs.81 and 82, to the SMFHO SNR^2. We obtained the efficiency for each disk signal and calculated the mean human efficiency factor, which is 30%±5%, by averaging over all signals. We compare the SMFHO CD curve, adjusted by human efficiency, to that obtained by Monnin at a similar entrance exposure for two specific disk diameters, the comparison shows a fairly good agreement. Notice that in this figure, we also show two human CD curves obtained at 70 and 140 μGy detector entrance exposures for the GE Senographe 2000D FFDM system by Rivetti[83] for comparison purposes. Even though the human CD analysis was performed on an older version of the GE FFDM system, the results are close to the SMFHO CD curve adjusted by human efficiency. Fig.3.8b shows the comparison between the SMFHO CD curves (before adjusting by human efficiency) and the ones obtained from CDCOM readings at three MGDs.



(a) The CD curves of the GE Senographe DS system acquired at 44.6 μGy detector entrance exposure from the SMFHO SNR adjusted by 30% human efficiency and the original SMFHO SNR. Two threshold thicknesses for disk diameter 0.1 mm and 0.25 mm at 54.2 μGy, cited from Ref.66. Another two human CD curves (for the GE Senographe 2000D system at 70 and 140 μGy detector exposures, cited from Ref.83) are shown for comparison purpose.



(b) The CD curves of the GE Senographe DS system at three MGDs: solid lines being the ones obtained from the SMFHO SNR , dashed lines are the ones from the CDCOM readings.

Fig. 3.8: The CD curves of the GE Senographe DS system obtained from using the SMFHO method, the CDCOM software readings and the human observer readings.

3.4 Discussion

The SMFHO method uses a model of system noise and deterministic properties for a uniform breast phantom to calculate the Hotelling observer SNR for clinically relevant FFDM imaging tasks. Using the empirically derived model, the SMFHO SNR can be estimated for different signal shapes and sizes with different system settings. With this method we can practically perform a comprehensive evaluation of clinical FFDM systems, including detector performance, scatter rejection, focal spot unsharpness and patient MGD.

In order to make the SMFHO methodology practical for evaluating FFDM systems in a clinical setting with collecting a limited number of images, we make the following assumptions: the FFDM system under evaluation is linear and cyclostationary. This method, without modification, is not applicable for evaluating systems that use non-linear or adaptive image processing algorithms. Furthermore, since the x-ray incident angles, the amounts of scatter and the focal spot dimensions all vary depending on different detector locations, the system is not shift invariant globally. Therefore, our results are only valid for the relatively small ROI of the experimental measurements, where the system blur and system noise can be considered shift invariant and cyclostationary.

We implemented the "boot-strapping" technology to account for the variabilities in the system noise and system response measurements. The variabilities come from the distribution of x-ray photons as well as system motion. However, since we only considered simulated signals, the variabilities of signals, including the difference between the actual signal diameter/thickness and the nominal diameter/thickness, have not been included in our error analysis.

In the attempt to validate the SMFHO method, we compare it to three other methods in Fig.3.6 and demonstrate that within error bars there is no significant difference between them.

$\mathrm{SNR}_{\mathrm{ISIB19}}$ was estimated in spatial domain using 5048 ROIs to estimate the noise covariance matrix. Since it used the most efficient Hotelling observer, it provided the highest SNR. $\mathrm{SNR}_{\mathrm{FSIB19}}$, on the other hand, was estimated in frequency domain. It used the same amount of ROIs to estimate system noise (GNNPS). The difference between the two SNRs is due to the limiting assumptions made in order to perform Fourier transform. Note that calculating the GNNPS performs more averaging than estimating the covariance matrix, the $\mathrm{SNR}_{\mathrm{FSIB19}}$ therefore has lower error bar than the $\mathrm{SNR}_{\mathrm{ISIB19}}$.

We also investigated how the SNR and its uncertainty vary with the number of ROIs used to calculate the GNNPS. We compared $\mathrm{SNR}_{\mathrm{FSIB256}}$, using 80 256×256 pixel ROIs, to $\mathrm{SNR}_{\mathrm{FSIB19}}$, using 5048 19×19 pixel ROIs. No significant difference was observed between

the two, except for a slightly smaller error bar of SNR_{FSIB19} due to the large amount of ROIs. Therefore, the results had shown that 80 is an adequate number of ROIs in our imaging tasks.

The SMFHO SNR is slightly smaller than the SNR_{ISIB19}, however the difference is not significant. The SMFHO methodology provides equivalent results to the image-space,image-based method, while in the meantime allowing for predicting FFDM system performance with other types of signals at different exposures without the need of collecting additional images.

One of the biggest advantages of the SMFHO method over using the detector MTF for estimating the SNR is that the GMTF accounts for realistic scatter conditions and thus can be used to evaluate the system scatter removal method. To demonstrate the advantages of the GMTF, for the GE Senographe DS system we compared the MTF with and without a grid (without using a phantom) and the GMTF without and without a grid (with the uniform phantom on the detector). Since the MTF only accounts for the detector blur, as well as potential scatter contributed by the grid, but not for the scatter rejected by it, the system with grid has a worse MTF as shown in Fig.3.3. However, the GMTF shows that even though the grid itself contributes to blur, it reduces the scatter fraction by 20%, and therefore significantly improves the system GMTF. The fact that the GMTF is noticeably higher for the system with grid shows that the blur added by the grid is a justified tradeoff. Therefore, the SMFHO evaluation methodology, which uses the GMTF to describe system deterministic properties, can be used to evaluate the performance of the system anti-scatter method.

In this chapter, we presented a practical method to estimate the 2D GMTF from 1D GMTF profiles. The importance of evaluating the 1D GMTF profiles along more than one direction is demonstrated by comparing the GMTF along the two axes. For the GE Senographe DS system the $GMTF_x$ and $GMTF_y$ are significantly different because the linear grid removes more scatter along the x axis than along the y axis. The Hologic Selenia system uses the Lorad High Transmission Cellular (HTC) grid, which rejects scatter with

a cross-hatch shaped grid[69], reduces the same amounts of scatter along the two axes. The resulting GMTF profiles along the two axes are thus identical at low frequencies due to the symmetric scatter fraction. However, there is still a noticeable difference between them at higher frequencies. One possible explanation is the rectangular shape of the focal spot, which has a smaller length along the y axis than that along the x axis, resulting in higher GMTF along the y axis. We obtained the 2D GMTF by fitting cubic splines between the $GMTF_x$ and $GMTF_y$. Even though the results are only accurate along the two axes, it provides additional information of the system blur than using the 1D GMTF.

The SMFHO method can be used to evaluate the effects of linear image processing algorithms on image quality. The *fine view* image processing used in the GE Senographe DS system acquisition mode A improves the GMTF. The GMTF of mode A increases with higher MGD, while the GMTF of mode B is constant with MGD. A possible explanation is that the *fine view* mode applies a MGD depended sharpening filter to images in order to enhance the edges. This processing filter also increases system noise at different MGDs as shown in Fig.3.5. However for the Hotelling observer, the processing effects on the GMTF and the GNNPS cancel out in the CD curves. Using this processing filter does not improve the system detectability of this observer.

It is instructive to examine the behavior of the SMFHO SNR to verify the dependence of the SNR on MGD and signal size. In Ref.84, Burgess provided a simple model to estimate SNR for uniform background and analytically showed that SNR^2 is linearly proportional to MGD while SNR is linearly proportional to signal size. The SMFHO SNR results for the two clinical systems reach a good agreement with Burgess's theory as we expected.

The SMFHO method uses an empirical model of FFDM systems based on the GNNPS and the GMTF. To implement this method, we simulated signals accounting for the linear attenuation coefficient of signals[85], x-ray spectrum, signal size, shape and thickness for a specific uniform background phantom. One advantage of this method is that we can predict FFDM system performance for other types of signals embedded in the same background without the need of collecting additional images.

Notice that we have the ability to obtain the SMFHO CD curves at any MGDs due to the fact that the SMFHO SNR^2 is linearly proportional to MGD. We can also predict the system performance at any given MGDs.

The SMFHO CD curves could be potentially used to predict human performance. Young *et al.* proposed a methodology to predict human observer performance from the CDCOM software readings of CDMAM images[81, 82].Using Young's method, we related the SMFHO CD curves to CDCOM readings, something that allowed us to predict human performance. A relevant research by Rivetti *et al.*[83] studied human performance and provided human observer CD curves for the GE Senographe 2000D. We digitized and displayed the human CD curves at two detector entrance exposures in Fig.3.8a. Although this CD analysis was performed on an older version of the GE FFDM system, the results are close to the SMFHO CD curve adjusted by human efficiency factor at 44.6 μGy detector entrance exposure. The efficiency of the TSFM observer compared to the human observer has not been evaluated experimentally. Future work could include performing human observer studies, comparing with the predicted human CD curves using the SMFHO method and calculating a human efficiency factor that can be directly linked to human performance. Notice that we have the ability to obtain the SMFHO CD curves at any MGDs due to the fact that the SMFHO SNR^2 is linearly proportional to MGD. We can also predict human performance at any given MGDs.

We summarized the advantages of using the SMFHO method to evaluate the performance of clinical FFDM systems. Comparing to the pixel SNR, the SMFHO method provides an objective assessment of the system performance based on a SKE/BKE detection task. It uses an uniform phantom when estimating the system noise and deterministic properties to take into account the scatter from the phantom and focal spot unsharpness. Comparing to the detector based MTF, the SMFHO method can be used to evaluate the performance of the anti-scatter grid in clinical FFDM. It provides equivalent evaluation results with Gagne's image-based method[45], but uses an empirical model of the system to predict system performance without the needs of collecting additional images when varying

exposures or signal types. Comparing to Monnin's method that uses a non-prewhtiening observer to simulate human performance[66], the SMFHO method estimates the best possible performance of FFDM systems quantified by the Hotelling observer SNR. This SNR is calculated in Fourier frequency domain to avoid taking a large number of images on clinical systems in order to calculate an invertible noise covariance matrix in spatial domain. The SMFHO method derives CD curves from the SNR results for the detection task, instead of fitting a function to relate the system detectability to CD curve in Monnin's method. It provides a direct comparison with human studies, and thus can be potentially used to predict human performance.

3.5 Conclusions

A clinically practical SMFHO assessment methodology for evaluating FFDM system performance has been presented in this chapter. This method uses an empirical model of the FFDM imaging system, incorporating a uniform phantom assembly placed in the FOV to account for scatter and focal spot unsharpness. For a SKE/BKE detection task, with taking only five images of the phantom background and another five of the edge test object, SMFHO CD curves can be generated. The evaluation results are not limited to specific system settings, and can be used to predict system performance using other types of signals at various MGDs. In addition, this method can be potentially used to predict human observer performance. We applied the method to evaluate the performance of the GE Senographe DS and the Hologic Selenia systems, and demonstrated that there is no significant difference between them.

4. EVALUATION OF IMAGE PROCESSING USED IN CLINICAL FFDM

4.1 Introduction

Post-acquisition image processing algorithms are typically employed in clinical FFDM systems to generate "for display" FFDM images with the familiar "film-like" apperarance as well as to aid human observers in extracting information. Most of the processing is typically performed by linear filtering, however some more sophisticated algorithms employ non-linear/non-reversable processing that introduce the possibility of affecting system detectability. Therefore, it is necessary to take into account the effects of post-acquisition image processing algorithms in the evaluation of clinical FFDM image quality.

The exact image processing algorithms employed in a clinical FFDM system are usually proprietary. Review of published litterature indicated that typical post-acquisition image processing algorithms include the following steps:

- Pre-processing
- Segmentation
- Normalization
- Post-processing

Pre-processing is the first step in image processing chain[86]. Detail enhancement algorithms can be used to set the grey levels of the direct irradiated detection areas (with no x-ray absorbtion) to zero in order to optimize the dynamic range of the images.

Logarithmic conversion can be used to adjust the grey level range in a digital image, by expanding values of dark pixels and compressing values of bright pixels.

Normalization is a linear processing applied to readjust the brightness and contrast across the whole breast region of a digital image[87]. Typically, without normalization, the dynamic range of monitors does not allow the display of all the variations in both the high density (bright areas near the center of the breast) and low density regions (dark areas near the nipple). Furthermore human readers will not be able to discriminate between small changes in a dark region and at the same time small changes in a bright region.

Region segmentation is the next step. The aim of segmentation is to locate various distinct regions such as compressed tissue areas, skin lines that form strong edges and suspicious masses. Matsubara *et. al.* developed a thresholding technique to categorize mammographic regions based on tissue densities using histogram analysis[88]. They set threshold values in the histogram to detect ROIs that contain suspicious masses. Szekely *et. al.* used a region based technique with a set of decision trees[89] to segment mammographic images[90]. The decision trees are automatically created from a set of training images. Each training image is divided into smaller ROIs, and each ROI has the class label (normal tissue or mass) provided by specialists. Difference in training images will cause the difference in the decision trees, and different decision trees will classify the image segments differently. The image is segmented using vote values provided by all the decision trees. Other authors proposed edge-detection based techniques to segment mammographic images[91, 92]. The edge-detection based technique detects the changing rate in grey levels, measured by the derivative of grey levels, to locate high contrast edges. By detecting the edges, the algorithm will be able to determine the boundaries of objects. After the segmentation, different regions could be separately processed according to different needs.

Edge enhancement is a post-processing algorithm that is commonly applied to enhance edges in order to make the regions/signals easier for humans to detect. Osher *et. al.* demonstrated the necessity of enhancing edges with image processing algorithms and they

described a filter based edge enhancement method and its applications[93]. Nitzberg *et. al.* described a model of nonlinear image filtering for edge enhancement using anisotropic diffusion[94]. This method can be used to enhance not only edges, but also corners as well as T junctions. Pollak *et. al.* proposed an edge enhancement method that uses a family of first-order multidimensional ordinary differential equations: the "Stabilized Inverse Diffusion Equations" (SIDE's)[95]. These equations are defined in a semi-discrete scale space, i.e., continuous in scale (or time) and discrete in space. They demonstrated successful applications of the method in enhancing edges with very high levels of noise, as well as to blurry signals.

Another important post-processing algorithm is noise suppression. Many authors proposed different filter based noise smoothing methods since 1980s[96–98]. Alvarez *et. al.* proposed that diffusion equations can also be used in noise suppression algorithms[99]. Following in the footsteps of their research, Pollak *et. al.* demonstrated the application of using SIDE's to reduce system noise[95].

Well designed image processing algorithms aid human observers to extract more diagnostic information, but inappropriate processing might damage the system detectability. For instance, noise suppression or smoothing could smooth-out low contrast signals while edge enhancement could potentially introduce spurious signals. Therefore, it is necessary to develop a clinically practical methodology that evaluates image processing of FFDM systems.

Image processing for FFDM heavily relies on the content of the images, therefore an image-based model observer would be more appropriate than the system model observer discussed in previous Chapters. Without dimensionality reduction, image-based methods are typically not practical in a clinical environment due to the fact that such methods require large numbers of images to estimate image quality. Myers and Barrett proposed to add a channel mechanism to observer models in order to make image-based model observers practical[100]. Channels can be considered as complete or incomplete basis vectors in a channelized space. A channelized image can be expressed uniquely as a finite

linear combination of a set of channels assuming with enough channels. After channelizing an image, the dimensionality of the image will be reduced to the number of channels used.

Myers and Barrett introduced a channelized ideal observer, constrained to process scenes through frequency selective channels[**?**]. For simplicity, they considered non-overlapping frequency selective channels with constant magnitudes. They found that this channelized ideal observer has equivalent performance with ideal observer for selected detection and discrimination tasks. In addition, they demonstrated that this channelized ideal observer is able to predict human performance in correlated noise. In the late 1990s, Barrett *et. al.* incorporated the channel mechanism in the Hotelling observer[101]. They indicated that the channelized Hotelling observer provides a practical method to quantify image quality.

Barrett *et. al.* indicated that a channelized observer only approximates the performance of that observer when the channels are efficient[101]. The choice of efficient channels is determined by imaging tasks, system properties and objects. Gallas and Barrett validated that the Laguerre - Gauess (LG) channels are efficient channels for the Hotelling observer in a detection task with lumpy backgrounds and rotationally symmetric signals in known locations[101]. Park *et. al.* introduced a LG channelized ideal observer in detection tasks using non-Gaussian distributed lumpy backgrounds and Gaussian signal profiles, and showed that the LG channels are efficient channels for the aforementioned tasks[102, 103].

In this Chapter, we implemented the LG channelized Hotelling observer model to perform a preliminary investigation the post-acquisition image processing algorithms used in the GE Senographe DS system. These non-linear algorithms are intended to improve image quality without affecting the safety and effectiveness of the system.

4.2 Materials and Methods

In this Chapter, we performed a preliminary investigation of the post-acquisition image processing algorithm of the GE Senographe DS FFDM system (for the system specifica-

tions please see Chapter 3) using the LG CHO model. Our goal with this preliminary investigation is develop a simple test to identify whether processing images for display causes information in FFDM images to be lost when compared to unprocessed images. As this is a preliminary investigation we will not make an attempt to evaluate how the human observer detectability is affected by image processing. In this section we describe the specialized signal phantom we developed by making use of the Rachel anthropomorphic phantom and the LG-CHO observer with an SKE/BKE task we implemented.

4.2.1 Theory

Let image $\mathbf{g}$ be an M dimensional vector (M×1). An (M×1) channel $\mathbf{u}_\mathrm{i}$ channelizes an image, returning a scalar value as follows:

$$\mathrm{v_i} = (\mathbf{u}_\mathrm{i})^\mathrm{t}\mathbf{g}. \tag{4.2.1}$$

In order to extract more information from the images, multiple channels are used to channelize an image. Given the total number of channels N, the channel matrix U (M×N) is defined as:

$$\mathrm{U} = [\mathbf{u}_1, \mathbf{u}_2 ... \mathbf{u}_\mathrm{N}] \tag{4.2.2}$$

Notice that in the attempt to reduce the dimensionality of the calculation, N is selected to satisfy N≪M. Therefore, the number of channels used to channelize an image is usually much smaller than the number of pixels of that image. A N-dimensional channelized image $\mathbf{V}$ can be obtained as:

$$\mathbf{V} = \mathrm{U}^\mathrm{t}\mathbf{g} = [\mathrm{v}_1, \mathrm{v}_2 ... \mathrm{v}_\mathrm{N}]. \tag{4.2.3}$$

The N-dimensional channelized image (N×1) reduces the dimension of the noise covariance matrix to N × N, where N is usually smaller than 10. In order to obtain a full

rank invertible covariance matrix, the number of images needed must be equal to or larger than N. For the GE Senographe DS system with a detection area of 1914×2094 pixel, for signals of typical sizes from 10 mm to 40 mm, the size of signal present ROIs would be around 100×100 pixel. Without channelizing the images, a minimum of 10,000 independent images must be taken in order to estimate the inverse of the covariance matrix for this ROI. Even if we had 10,000 images, it is still computationally intense to generate the inverse of such a large covariance matrix[104]. The channelized model observer provides a practical methodology to estimate the noise covariance matrix and its inverse.

The Channelized Hotelling Observer (CHO) uses a test statistic, which is a function of the channelized image $\mathbf{V}$:

$$\mathrm{t}\big(\mathbf{V}\big) = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{U}\mathbf{K}_{\mathrm{c}}^{-1}\mathbf{V}, \tag{4.2.4}$$

where the channelized covariance matrix $\mathbf{K}_{\mathrm{c}}$ is defined as:

$$\mathbf{K}_{\mathrm{c}} = \left\langle \big(\mathbf{V}_{\mathrm{i}} - \overline{\mathbf{V}}\big)\big(\mathbf{V}_{\mathrm{i}} - \overline{\mathbf{V}}\big)^{\mathrm{t}} \right\rangle. \tag{4.2.5}$$

The CHO SNR can be estimated using the following definition:

$$\mathrm{SNR}_{\mathrm{c}}^2 = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{U}\mathbf{K}_{\mathrm{c}}^{-1}\mathbf{U}^{\mathrm{t}}\Delta\mathbf{S}. \tag{4.2.6}$$

It has been previously demonstrated that, with enough LG channels, the LG CHO approximates the performance of the ideal observer in detection tasks of circularly symmetric signals embedded in a lumpy background and positioned at known locations[101].

The LG channels are created from Laguerre polynomials and Gaussian functions defined as:

$$u_i(r|a_u) = \frac{\sqrt{2}}{a_u} exp\left(\frac{-\pi r^2}{a_u^2}\right) L_i\left(\frac{2\pi r^2}{a_u^2}\right), \tag{4.2.7}$$

where a_u is the width of the Gaussian function. a_u determines a Gaussian envelope

($a_u = 2\pi\sigma_u^2$), and L_i determines a series of Laguerre polynomials:

$$L_i(x) = \sum_{j=0}^{i}(-1)^j \binom{i}{j} \frac{x^j}{j!}. \tag{4.2.8}$$

In the LG channel definition, the variable $x = \left(\frac{2\pi r^2}{a_u^2}\right)$ yields an orthonormal family that satisfies:

$$\int_0^\infty r\mathrm{d}r\, u_i(r)u_j(r) = \delta_{ij}. \tag{4.2.9}$$

4.2.2 Experimental Setup

System and Phantom

In the development of our phantom we made use of the Rachel anthropomorphic phantom which when imaged radiographically it generates realistic anthropomorphic projection images. We chose a realistic breast background in order to trigger the type of image processing that will be applied in a clinical situation. Based on our experience, image processing algorithms are disabled when using phantoms with uniform backgrounds. The Rachel phantom consists of an outer case, tissue equivalent material at different thicknesses that model location specific breast attenuation and a high resolution mammogram film. The layer of tissue equivalent materials is used to simulate breast thickness variation. The tissue layer is superimposed to the digitally processed mammogram film to model the complex structure of mammograms.

We designed a holder for the Rachel phantom, which facilitates reproducible positioning of the phantom in the x-ray field. It is very important to maintain the same background with and without signals in order to have a good estimation of the difference signal $\Delta\mathbf{S}$, which is obtained by subtracting the signal absent image (background only) from the signal present image (background with signals).

Measurements

Typically, non-linear image processing algorithms adapt to different imaging conditions. For example, edge enhancing is used to enhance pre-detected high contrast objects, and noise smoothing is typically applied to smooth-out noise in regions with no pre-detected objects. High contrast objects can be enhanced by edge enhancing algorithms; and low contrast objects (contrast lower than a preset threshold when compared to the noise) might be smoothed out by the noise smoothing algorithms.

In an attempt to evaluate the effects of potential non-linear image processing algorithms, we considered both high and low contrast signals. To determine the appropriate contrast of the signals system and exposure we run multiple pilot studies using disk signals of different materials, different thicknesses and diameters. Our goal for the pilot studies was that the low contrast objects should be on the threshold of being undetectable by trained human observers when superimposed over realistic breast background. Such signals could potentially be considered as part of the background by image processing algorithms. On the other hand, our goal for the high contrast objects was that they be slightly over the detectable threshold when superimposed over a realistic mammographic background. Such signals coule be easily identified by image processing algorithms and potentially enhanced. The materials that satisfied the aforementioned conditions were 0.26 mm thick PMMA and 0.16 mm thick Polycarbonate for the high and low contrast signal films respectively. We cut circular holes in the films with a precision LASER cutter. The holes are referred to as negative signals. The shape of the signal films was semicircular in order to match the shape of the breast in the Rachel phantom. Signals with three different diameters: 2 mm (six signals), 3 mm (four signals) and 4 mm (four signals), are positioned on concentric circular arcs at 25, 45 and 65 mm from the center as can be seen in Fig.(4.1).

Since our goal was to determine whether signal information was lost due to image processing without regarding human observer performance, we wanted to maximize the information that our model observer had in its disposal. For this reason we chose an SKE/BKE type of task where the observer has knowledge of the fixed average background

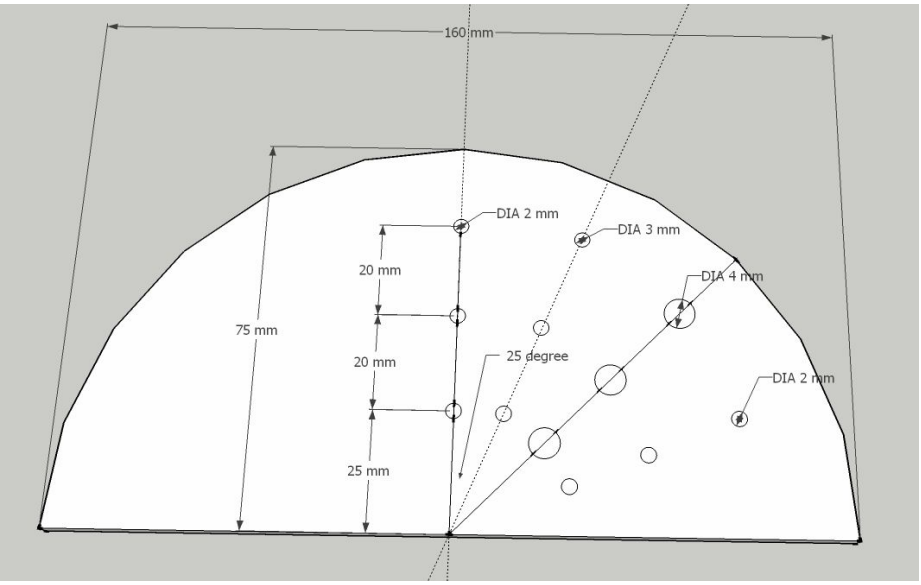


Fig. 4.1: Signal film designed for both high (PMMA) and low (Polycarbonate) contrast signals.

as well as the fixed average signal. Note that typically in mammography images are not quantum limited, but detectability is limited by the background complexity. We therefore expect that even though the low contrast signals are barely visible when superimposed over a mammographic background, when the background is subtracted for the SKE/BKE task the signals will be readily visible, as in the case of digital subtraction angiography.

For the implementation of the methodology we used the GE Senographe DS FFDM system that we had available at the FDA with the following technique parameters: Rh/Rh, 28 kVp and 50 mAs. We placed the high contrast PMMA film on top of the Rachel phantom holder without having the phantom in the FOV to record the signal locations. This system setup facilitates the determination of signal locations in the absence of anatomical background. We took five images of each signal film and determined the signal locations from the average image. We assigned consecutive numbers (starting from 1) to mark the signals with the same diameter. The signal locations (with their assigned numbers) are shown in Fig.(4.3), three signal diameters can be differentiated by colors. We repeated the procedure for the low contrast Polycarbonate film.

After the determination of signal locations, we placed the Rachel phantom in the FOV,

Fig. 4.2: Rachel anthropomorphic phantom, a high contrast signal (PMMA) film and the phantom holder for consistent positioning. This phantom models the breast structure radiographically.

inside the phantom holder. The signal film was kept stationary on top of the holder to facilitate the consistent positioning, as shown in Fig.(4.2). We set the system technique parameters to Rh/Rh, 30 kVp and 110 mAs. We took 100 signal present images (with the signal films) and 100 signal absent images (with the signal films replaced by blank PMMA/Polycarbonate films), and stored the images with and without the post-acquisition processing in order to make the comparison.

We selected 120×120 pixel signal present ROIs from signal present images with the signal located in the center of the ROI. We selected the signal absent ROIs from signal absent images, at the same locations as the signal present ROIs to ensure the same background.

We investigated the number of channels needed to ensure the efficiency of the LG channel set. For three high contrast signals with different sizes, we calculated the LG-CHO SNR by summing over different numbers of channels and plotted the LG CHO SNR as a function of channel number.

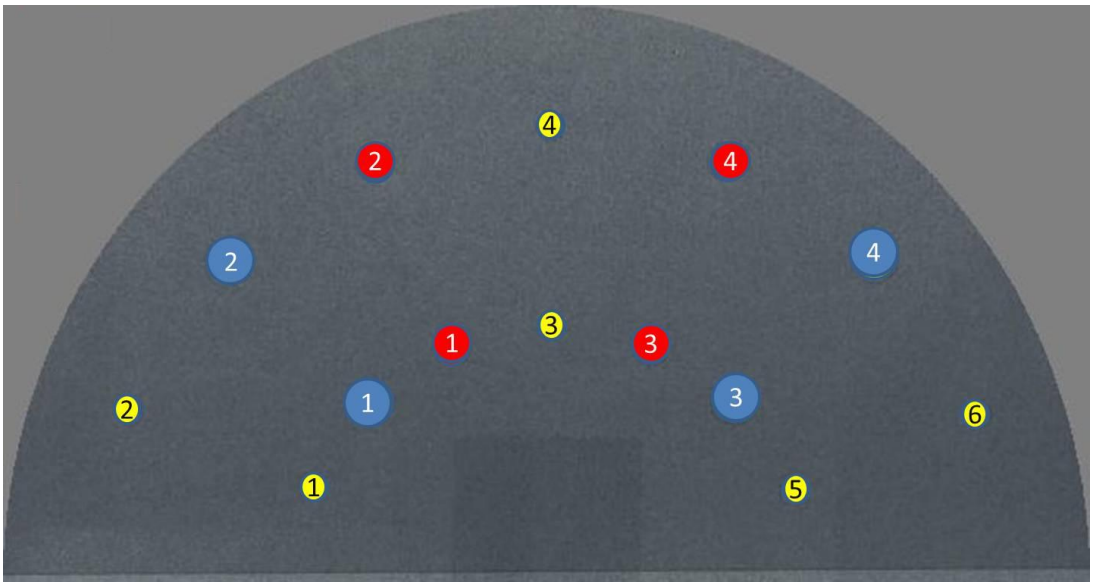


Fig. 4.3: The average image over five individual images of the signal film. For each signal diameter, there are four (3 mm and 4 mm) or six (2 mm) signals. We assigned consecutive numbers (starting from 1) to mark the signals with the same diameter.

Based on the results of our analysis, we decided to use ten channels to build the Channelized Hotelling Observer (CHO) model. For each disk signal, the LG channels satisfy that the width of the first channel is equal to the Full Width at Half Maximum (FWHM) of the signal. Therefore, the width of the Gaussian function a_u was set to 10, 8 and 6 mm corresponding to disk diameter of 4 mm, 3 mm and 2 mm. Note that this LG CHO model reduces the dimension of the data to the number of channels (ten), which is much smaller

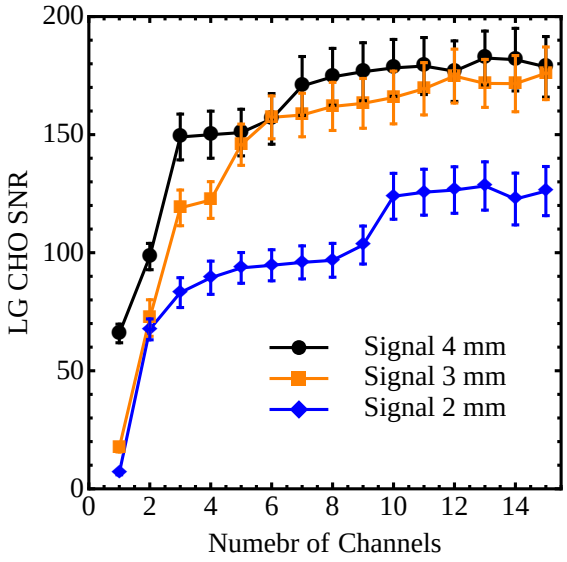


Fig. 4.4: The LG CHO SNR as a function of the number of channels for three high contrast (PMMA) signals with diameters 2 mm, 3 mm and 4 mm. Raw (un-processed) images are used in this analysis.

than the original image.

We applied the 2D LG channels on

both the signal present and absent ROIs following Eq.(4.2.3), and obtained the channelized difference signal $\Delta\mathbf{S}$. We calculated the channelized covariance matrix as defined in Eq.(4.2.5). And we estimated the channelized Hotelling observer SNR using Eq.(4.2.6). "Boot-strapping" with replacement was used to randomly sample ROIs to obtain new data sets of the CHO SNR. The standard deviation of the CHO SNR was estimated from 100 "boot-strapped" ROI samples for each disk signal. Our error analysis accounted for quantum and electronic noise as well as minor x-ray tube and system motion between exposures.

4.3 Results

Fig.(4.4) shows the LG-CHO SNR as a function of the number of channels used in the channelized observer for three high contrast (PMMA) signals with diameters: 4 mm, 3 mm and 2 mm. Error bars were generated from the "boot-strapped" samples of signal present/absent images. The LG-CHO SNR becomes stable for all three signal sizes when using ten or more channels.

Examples of the channels we used for the three signal diameters 4 mm, 3 mm and 2 mm are shown in Fig.(4.5) from top to bottom. The first four channels (out of 10) are shown from left to right.

Fig.(4.6) shows the signal present (left) and absent (right) ROIs as well as the difference signal in the middle for both the high contrast and low contrast signals. Disk diameter is fixed at 2 mm. Fig.(4.6a) and Fig.(4.6b) are the images of the high contrast signals with and without image processing. Fig.(4.6c) and Fig.(4.6d) are the images of the low contrast signals with and without image processing.

Fig.(4.7), Fig.(4.8) and Fig.(4.9) show the resulting CHO SNR for high contrast signals (PMMA) with three diameters: 4 mm, 3 mm and 2 mm respectively. For each signal, the

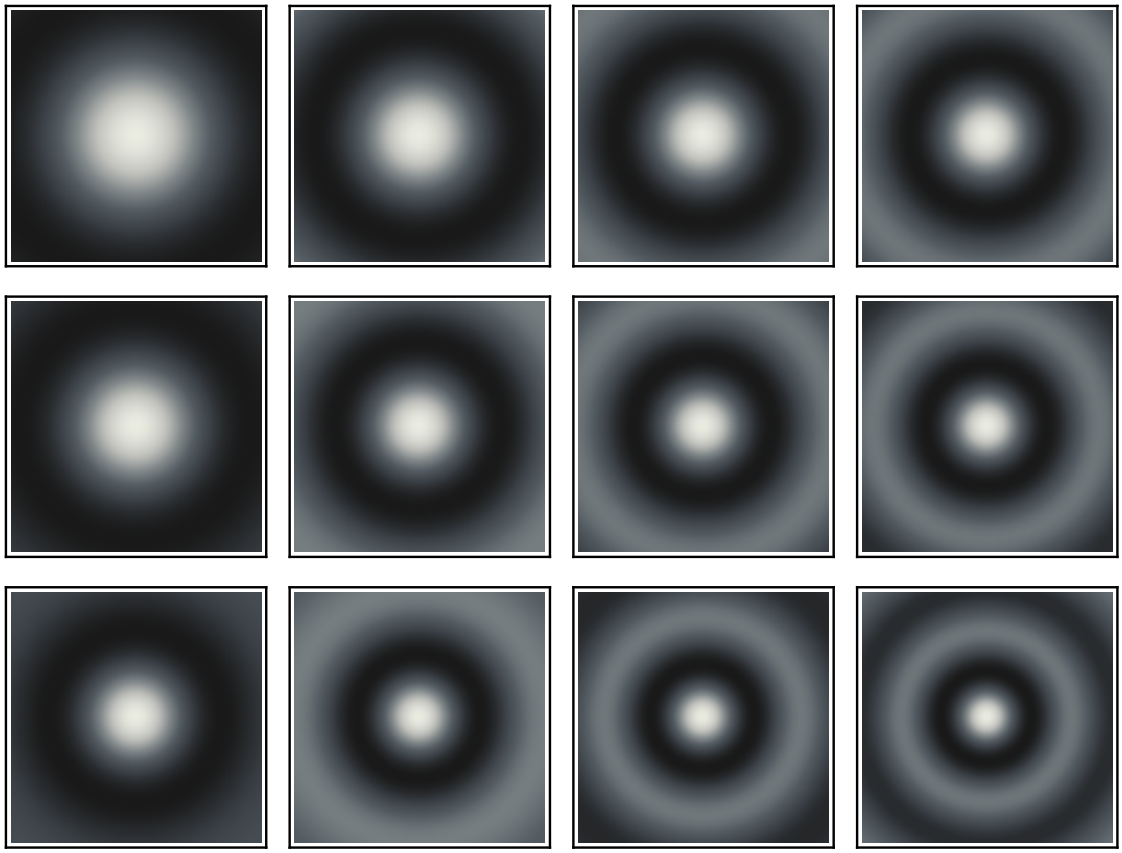
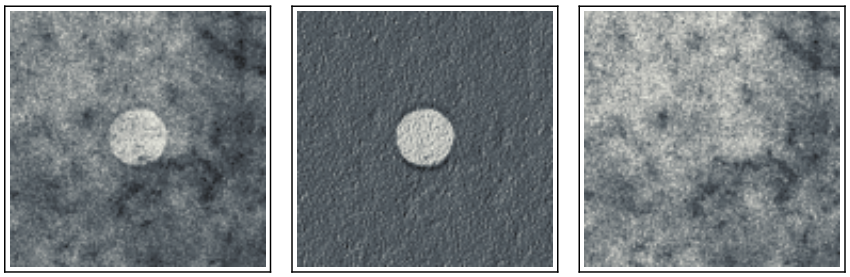


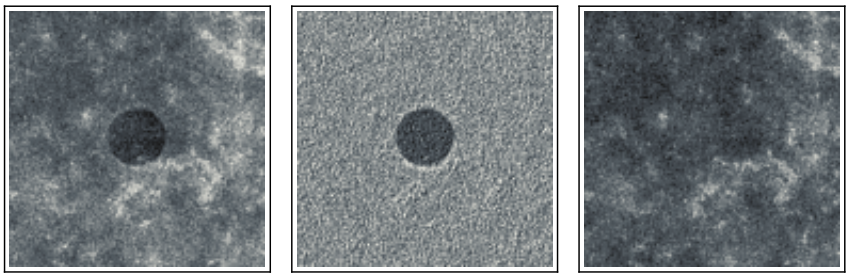
Fig. 4.5: The three rows are the 120×120 LG channels with a_u=10,8 and 6 from top to bottom. The four columns are the first four 120×120 LG channels with i=1,2,3 and 4 from left to right.

CHO SNR of the raw image (denoted by A) is compared to that of the processed image (denoted by B). The signal locations are marked by consecutive numbers from 1 to 6 as shown in Fig.(4.3). Standard deviation of the CHO SNR was estimated from the "bootstrapped" samples. Fig.(4.10) shows the percent difference of the CHO SNR, for high contrast signals, between the raw images and processed images for the different signal sizes and locations. Positive difference indicates higher CHO SNR of the raw images, and negative difference indicates higher SNR of the processed images. Average percentage error of the percent difference was 13%. Fig.(4.11), Fig.(4.12) and Fig.(4.13) show the CHO SNRs and their error bars for low contrast signals. Fig.(4.14) shows the percentage

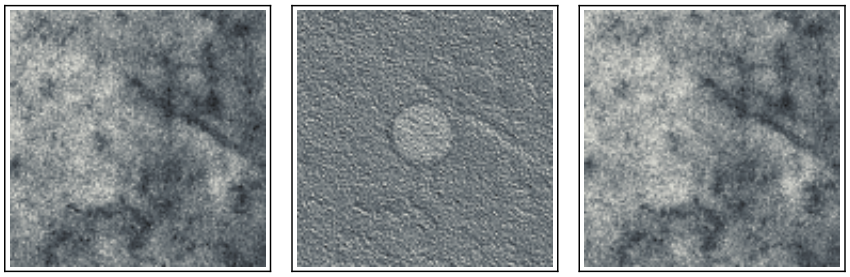
difference of the CHO SNR for low contrast signals. Average percentage error of the percent difference was also 13%.



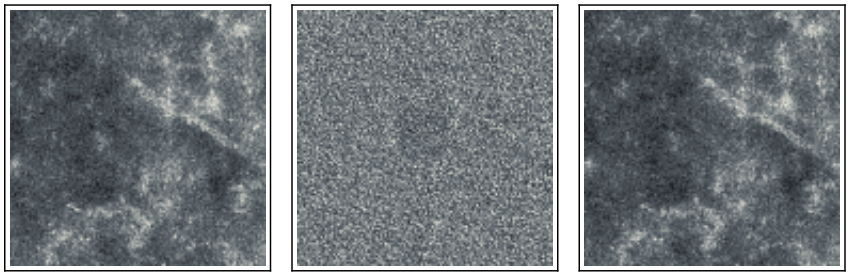
(a) High contrast PMMA signals, without image processing.



(b) High contrast PMMA signals, with image processing.



(c) Low contrast Polycarbonate signals, without image processing.



(d) Low contrast Polycarbonate signals, without image processing.

Fig. 4.6: Signal present (left), absent (right) ROIs as well as the difference signal (middle) for both the high contrast and low contrast signals. Signal diameter is fixed at 2 mm.

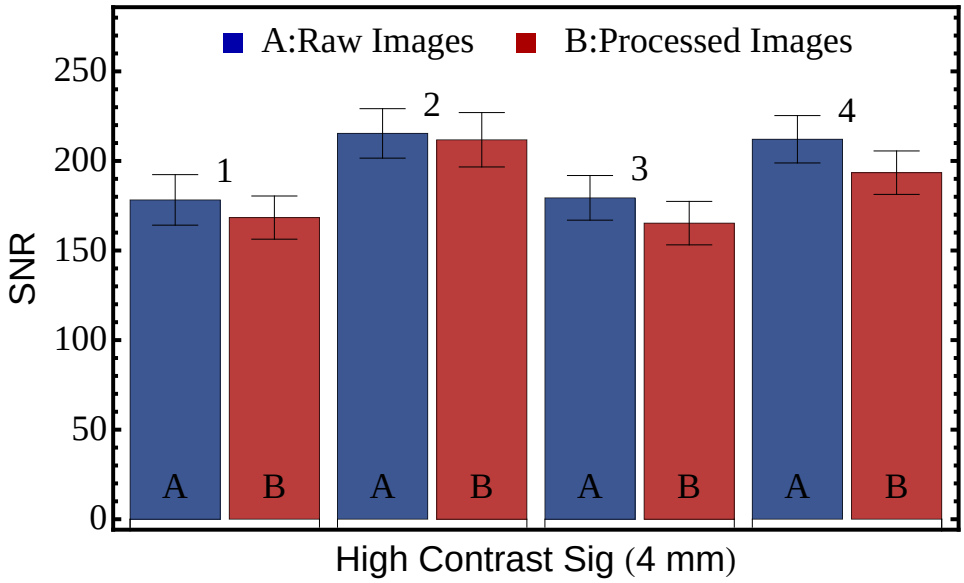


Fig. 4.7: The CHO SNR for high contrast signals with diameter 4 mm at four different locations.

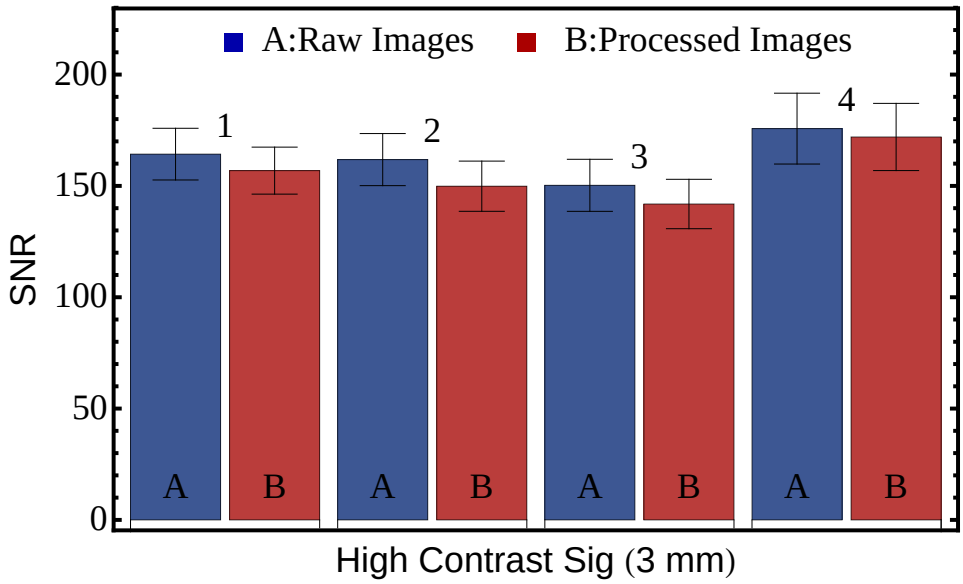


Fig. 4.8: The CHO SNR for high contrast signals with diameter 3 mm at four different locations.

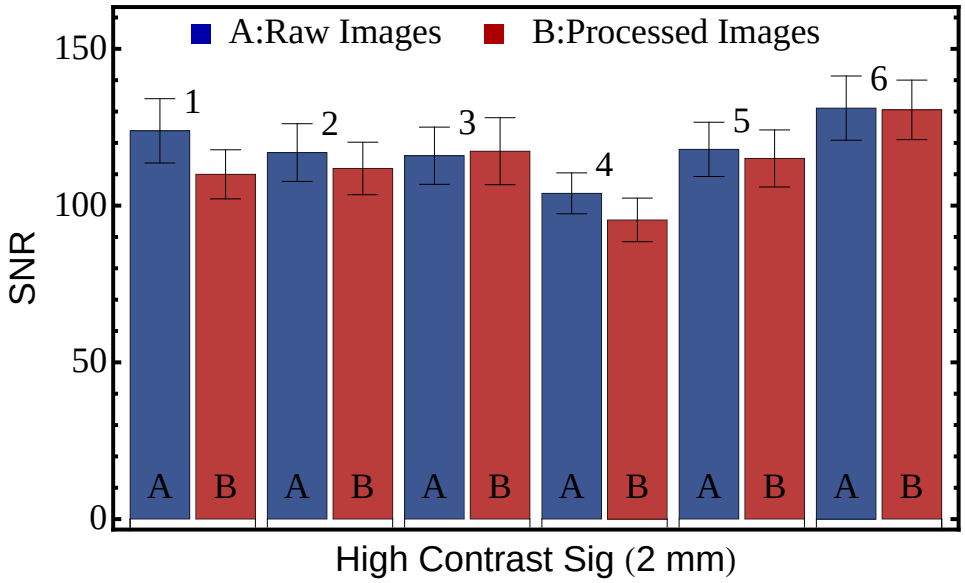


Fig. 4.9: The CHO SNR for high contrast signals with diameter 2 mm at six different locations.

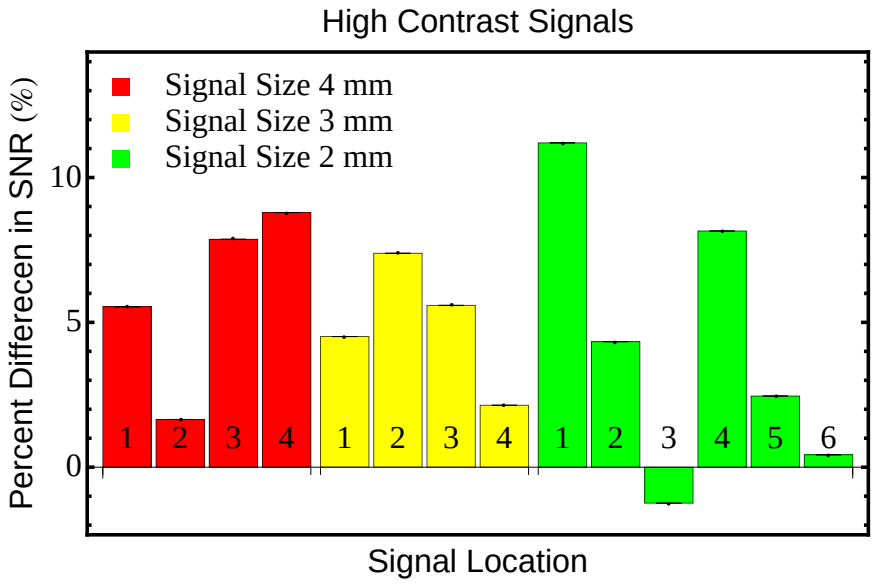


Fig. 4.10: The percent difference of the CHO SNR between raw images and processed images of high contrast signals. Positive difference indicates higher CHO SNR of the raw images, and negative difference indicates higher SNR of the processed images. Three colors represent different signal sizes, numbers represent the corresponding locations. The average percentage error of the percent difference was 13%.

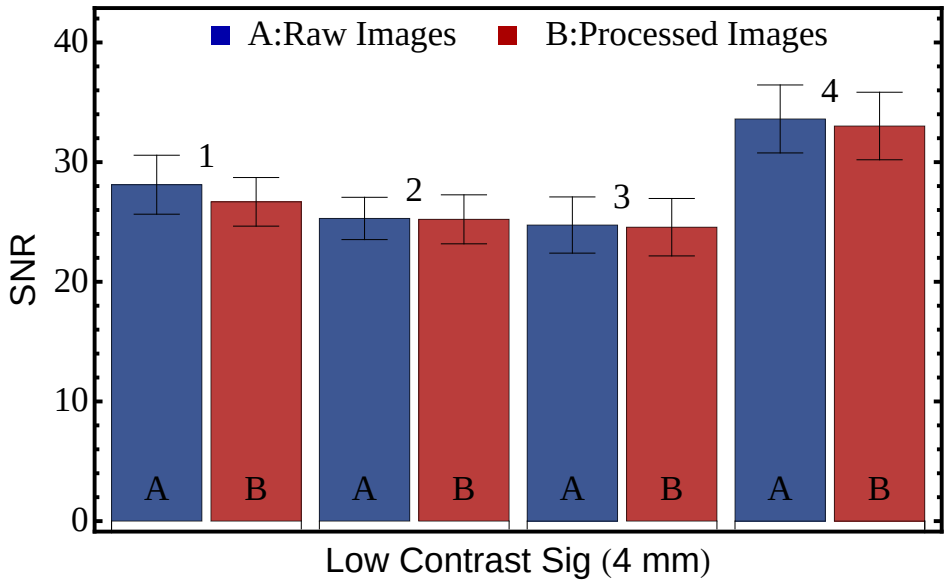


Fig. 4.11: The CHO SNR for low contrast signals with diameter 4 mm at four different locations.

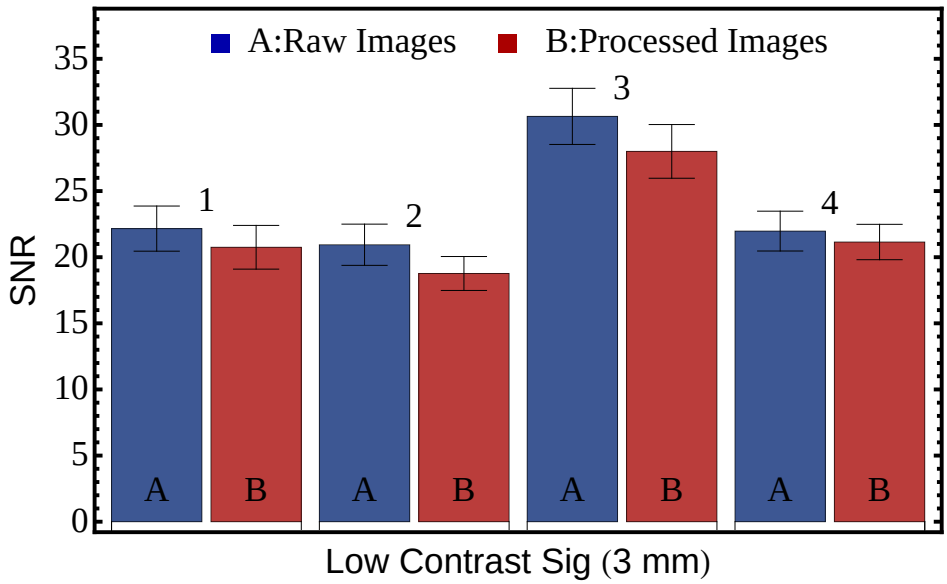


Fig. 4.12: The CHO SNR for low contrast signals with diameter 3 mm at four different locations.

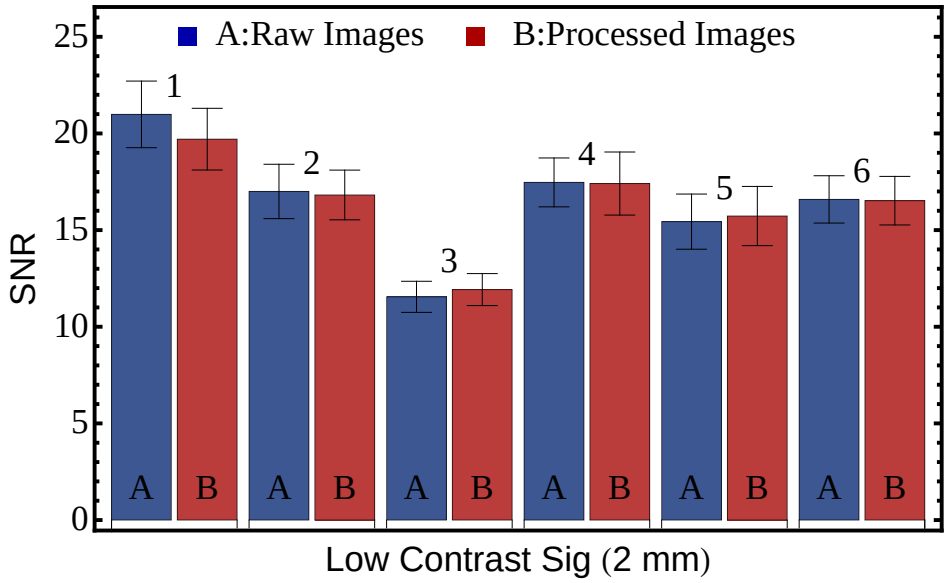


Fig. 4.13: The CHO SNR for low contrast signals with diameter 2 mm at six different locations.

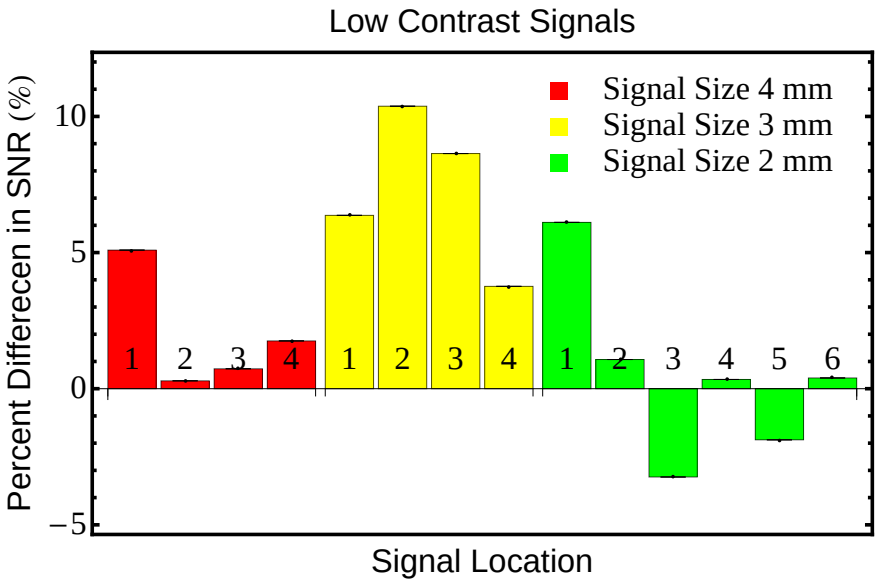


Fig. 4.14: The percent difference of the CHO SNR between raw images and processed images of low contrast signals. Positive difference indicates higher CHO SNR of the raw images, and negative difference indicates higher SNR of the processed images. Three colors represent different signal sizes, numbers represent the corresponding locations. The average percentage error of the percent difference was 13%.

4.4 Discussion

The purpose of post-acquisition image processing in FFDM is twofold: to present mammograms in a way that is familiar with radiologists that were trained with film-based images and also to maximize the diagnostic information extracted by radiologists. In this thesis, we focused on answering the narrow question on whether image processing in FFDM degrades useful information in clinical images. In this preliminary work we developed a phantom and and implemented a methodology to investigate this question and implemented on a single FFDM system. For the methodology to be fully validated it should be tested on different systems with different image processing algorithms under varying conditions. Furthermore, human observer studies should be considered to investigate the effects of such processing algorithms on humans.

In order to analyze the effects of image processing algorithms, an image-based Hotelling observer is much more suitable than the system-model-based Hotelling observer that we proposed in the previous Chapters. The system-model-based Hotelling observer has been developed based on an assumption of linear system performance, however, such assumption will likely not be satisfied because most processing algorithms are non-linear.

We used a LG CHO model to reduce the dimensionality of the image-based Hotelling observer model. We know from literature that with enough LG channels, the LG CHO approximates the performance of the ideal observer and we have demonstrated in Fig.(4.4) that the LG CHO SNR becomes stable after using ten or more than ten channels.

Low contrast signals are intended to be used for evaluating the possible reduction of signal detectability by the noise smoothing algorithm. We therefore chose low contrast signals so that the signals would be hardly detectable in the signal signal plus background images (the input to the image processing algorithms), but are easily detectable after subtracting the background (the input to the model observer), as shown in Fig.(4.6). The CHO SNR for low contrast signals with and without the processing algorithms shows no significant difference within the error bars. It demonstrates that the non-linear processing

of the GE Senographe DS does not reduce signal information from the processed images. However, even though the difference of the CHO SNR between the raw images and the processed images is not significant, the CHO SNR tends to be sightly higher for the raw (un-processed) images.

High contrast signals are designed for evaluating edge enhancing algorithms that are potentially applied to enhance these signals. High contrast signals are readily detectable in the signal present images. The CHO SNR for high contrast signals also shows no significant difference between the processed and unprocessed images. The non-linear post-acquisition processing applied to high contrast signals does not reduce the signal information significatly, however the trends show slight non-significant increase CHO SNR in the unprocessed, raw images.

The CHO SNRs for the same signal size and contrast are significantly different at different locations because the system noise is Poison distributed and thus proportional to x-ray intensity over the signal regions. Even though the CHO SNR was calculated using the difference signal $\Delta\mathbf{S}$ that is independent on background structure, the system noise at different locations is still different depending on the background attenuation.

4.5 Conclusions

In this Chapter, we implemented the LG Channelized Hotelling observer model to investigate the post acquisition image processing algorithms used in the GE Senographe DS FFDM system. We studied the effects of image processing on both high and low contrast signals with a realistic anthropomorphic background. We found that there is no significant difference of the LG CHO-SNR between the processed and unprocessed images. We thus concluded that, for the FFDM system we studied, image processing does not adversely affect signal detectability of the system as evaluated by the LG CHO-SNR for the specific SKE/BKE task we perfromed. This study, however, does not evaluate whether human observer detectability affected by image processing algorithms. Our evaluation results were

limited to the GE Senographe DS system, however, the methodology is general and should be applied to investigate image processing used in other clinical FFDM systems in the future. Furthermore, the methodology should be validated against other methods based on unkown location of the signal and for background unkonwn, or known statistically.

5. SUMMARY AND CONCLUSIONS

5.1 Summary

In this dissertation, we proposed methodologies for evaluating both the image acquisition performance and image processing algorithms of clinical FFDM systems.

In the evaluation of image acquisition performance of clinical FFDM systems, we developed a system-model-based Hotelling observer method, which would be complementary to the current Class II Special Controls Guidance for FFDM systems published by the U.S. FDA. In Chapter 2, we have demonstrated the application of the methodology on a bench-top system that models projection mammography in spatial domain, with using the covariance matrix and H matrix to characterize the system noise and deterministic properties respectively. By placing a phantom in the FOV, the scatter from patients, the focal spot unsharpness and the magnification factor have been taken into account in the assessment. To make the methodology clinically practical, we extended it into Fourier frequency domain and applied it on two clinical FFDM systems, which are the GE Senographe DS and the Hologic Selenia, with taking limited number of images as shown in Chapter 3.

In the evaluation of image processing algorithms in clinical FFDM, we implemented an image based method that uses the LG channelized Hotelling observer model in Chapter 4 to quantify image quality before and after applying the image processing. We have demonstrated the application of this method on the GE Senographe DS system. Moreover, even though we only applied this method on one clinical FFDM system, the method is general and has the potential to be used to evaluate image processing algorithms used in other FFDM systems as well as in other imaging modalities.

In Chapters 2 and 3, we proposed to use a system-model-based Hotelling observer method in the evaluation of image acquisition performance of clinical FFDM systems. In this approach that we introduced for the first time, we built an empirical model of the FFDM system, with estimating the generalized definitions of system noise and deterministic properties when placing a uniform phantom assembly in the FOV. This method provides a complete description of the image acquisition performance of clinical FFDM systems with taking into account not only the detector blur, but also scatter blur, focal spot unsharpness and noise correlations. This approach provides a clinically practical means to not only evaluate FFDM image acquisition performance at a given exposure, but also to predict system system performance at any other exposures with different types of signals without the needs of collecting additional images when the access to clinical FFDM systems is limited. In addition, CD curves can be derived from the Hotelling observer SNR with simulating a series of signals that have different sizes and thicknesses. It leads to potential prediction of human observer performance. In Chapter 3, we have demonstrated this prediction by adjusting the SMHO CD curve with an efficiency value. The results reached an agreement with actual human studies provided in published literatures.

In Chapter 4, we implemented the CHO model to investigate the effects of image processing in clinical FFDM systems. We proposed to use an image based method that accounts for the effects of non-linear/adaptive image processing. We designed signal phantoms to be used with an anthropomorphic background phantom in order to trigger potential image processing used in GE Senographe DS system. We compared image quality before and after the image processing, and we concluded that the image processing has a trend of reducing system detectability for both high and low contrast signals, but the difference is within error bars. Even though we have only demonstrated this method on one clinical FFDM system, the method is general, and has the potential to be used for evaluating image processing in other FFDM systems as well as in other modalities.

5.2 Assumptions and Limitations

The purpose of this research is to develop an approach to evaluate the image acquisition part of FFDM systems. We incorporated the effects of focal spot unsharpness, the effects of scatter from patients and the effects of magnification to image quality. Other effects that might affect image quality but were not investigated here include the ghosting effect during the image acquisition, the motion effect, etc. These effects can be discussed in future research.

Furthermore, some assumptions were made for the system properties. These are limitations on the application not on the concepts of the methodology. In Chapters 2 and 3, uniform backgrounds were given by the phantom assembly. The background region can be therefore separated into subregions, and the sub-ROIs can be treated as if they are equivalent to each other. We ended up having much more ROIs than the number of background images that we obtained. This assumption is not valid for phantoms with non-uniform backgrounds. The background ROI selection has to be location specific in that case, which requires a larger number of background images to be taken. Another assumption is that the 2D system GMTF can be estimated from the 1D GMTF measured along x and y directions. This assumption is limiting when a relatively large focal spot is used since it produces a non-symmetric GMTF. It is also limiting when the system grid removes different amounts of scatter along directions other than the x and y axes. Our signal simulation is also limiting when the actual signals are produced with variabilities in their thicknesses.

5.3 Future Work

In this dissertation, we have only considered the use of the Hotelling observer model, which gives the best possible system performance. Even though we have demonstrated the potential of using this observer model to predict human performance, human observer efficiency comparing to this observer is relatively low. In the future, we can implement other observer

models that better match with human performance, such as the non-prewhitening matched filter, in an effort to predict human performance. And we can perform actual human studies, and compare the results with both the Hotelling observer and the non-prewhitening matched filter in generating CD curves for evaluating image acquisition performance of clinical FFDM systems.

In the evaluation of FFDM image acquisition, we have shown the application of the system-model-based Hotelling observer method on two different FFDM technologies. In the future, we can use this method to evaluate other FFDM technologies, and provide a complete comparison of FFDM image acquisition performance between different available technologies as well as between different manufacturers.

In the evaluation of image processing, we implemented the CHO model to evaluate non-linear image processing used in clinical FFDM. This method also has the potential to be able to evaluate non-linear processing used in CT reconstruction algorithms. GE claims that they developed an ASIR iterative reconstruction method for CT that can remarkably reduce dose without affecting image quality. Using this CHO method, we have the potential to evaluate this iterative reconstruction method, link radiation dose to an objective assessment of image quality and provide validation to GE's aforementioned statement.

We used an anthropomorphic phantom with realistic background structures in the evaluation of image processing in FFDM. Here, we added the background in the attempt to trigger image processing that would be otherwise turned off. We considered an SKE/BKE task without accounting for the effects of anatomic background noise on signal detectability. In the future, we could implement a search-based model observer that searches for signals with unknown locations in a statistically known background. This detection task would link observer based evaluation one step closer to diagnostic performance of FFDM systems.

APPENDIX

A. HOTELLING OBSERVER MAXIMIZES THE SNR OF LINEAR OBSERVERS

The Hotelling observer maximizes the performance of all linear observers including mathematical model observers and humans[29].

The test statistic of a linear observer is:

$$t(\mathbf{g}) = \mathbf{u}^{\mathrm{t}}\mathbf{g}, \tag{A.0.1}$$

where $\mathbf{u}^{\mathrm{t}}$ is the linear observer template and $\mathbf{g}$ is a given image. According to Eq.1.2.2, the linear observer SNR satisfies:

$$\mathrm{SNR}^2 = \frac{\left(\mathbf{u}^{\mathrm{t}}\Delta\mathbf{S}\right)^2}{\mathbf{u}^{\mathrm{t}}\mathbf{K}\mathbf{u}}. \tag{A.0.2}$$

The test statistic template of the Hotelling observer is:

$$t(\mathbf{g}) = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\mathbf{g}, \tag{A.0.3}$$

and therefore the Hotelling observer SNR satisfies:

$$\mathrm{SNR}^2 = \Delta\mathbf{S}^{\mathrm{t}}\mathbf{K}^{-1}\Delta\mathbf{S}. \tag{A.0.4}$$

In order to prove that the Hotelling observer SNR maximizes SNR of all linear observers, we need to prove the following inequality:

$$\Delta \mathbf{S}^{\mathrm{t}} \mathbf{K}^{-1} \Delta \mathbf{S} \geq \frac{\left(\mathbf{u}^{\mathrm{t}} \Delta \mathbf{S}\right)^2}{\mathbf{u}^{\mathrm{t}} \mathbf{K} \mathbf{u}}. \tag{A.0.5}$$

After multiplying the term $\mathbf{u}^{\mathrm{t}}\mathbf{K}\mathbf{u}$ to both sides of the inequality, we need to prove that:

$$\Delta \mathbf{S}^{\mathrm{t}} \mathbf{K}^{-1} \Delta \mathbf{S} \,.\, \mathbf{u}^{\mathrm{t}} \mathbf{K} \mathbf{u} \geq \left(\mathbf{u}^{\mathrm{t}} \Delta \mathbf{S}\right)^2. \tag{A.0.6}$$

Eq.A.0.6 can be rewritten as:

$$\Delta \mathbf{S}^{\mathrm{t}} \mathbf{K}^{-\frac{1}{2}} \mathbf{K}^{-\frac{1}{2}} \Delta \mathbf{S} \,.\, \mathbf{u}^{\mathrm{t}} \mathbf{K}^{\frac{1}{2}} \mathbf{K}^{\frac{1}{2}} \mathbf{u} \geq \left(\mathbf{u}^{\mathrm{t}} \mathbf{K}^{-\frac{1}{2}} \mathbf{K}^{\frac{1}{2}} \Delta \mathbf{S}\right)^2. \tag{A.0.7}$$

Because of the Cauchy - Schwartz inequality, Eq.A.0.7 is true, and therefore Eq.A.0.5 is true[29].

BIBLIOGRAPHY

[1] D. V. Vakil and R. W. Morgan, "Etiology of breast cancer. II. Epidemiologic aspects," *Can Med Assoc J.* **109**(3), pp. 201–6, 1973.

[2] Early Breast Cancer Trialists Collaborative Group, "Tamoxifen for early breast cancer: an overview of the randomised trials," *Lancet.* **351**(1), p. 145167, 1998.

[3] R. Peto, J. Boreham, M. Clarke, C. Davies, and V. Beral, "UK and USA breast cancer deaths down 25% in year 2000 at ages 2069 years ," *Lancet.* , 2000.

[4] American Cancer Society, "Cancer Facts & Figures 2008 ," *Atlanta: American Cancer Society.* , 2008.

[5] A. C. Society, *Breast Cancer Facts and Figures 2011-2012*, Atlanta: American Cancer Society, Inc., 2011.

[6] H. Lynch, *Genetics and breast cancer*, New York: Van Nostrand Reinhold., 1981.

[7] B. Armstrong and R. Doll, "Environmental factors, cancer incidence, and mortality in different countries, with special reference to dietary practices," *Int J Cancer.* **15**, pp. 617–31, 1975.

[8] J. Kovi and M. Heshmat, "Incidence of cancer in negroes in Washington DC and selected African cities," *Am J Epidemiol.* **96**, pp. 401–13, 1971.

[9] G. McCormick and R. Moon, "Effect of pregnancy and lactation on growth of mammary tumors induced by 7, 12 DMBA," *BrJ Cancer.* **19**, p. 160, 1965.

[10] H. T. LYNCH, W. A. ALBANOT, M. A. LAYTON, W. J. KIMBERLING, and J. F. LYNCH, "Breast cancer, genetics, and age at first pregnancy," *Journal of Medical Genetics.* **21**, pp. 96–98, 1984.

[11] J. A. Harvey, B. T. Nicholson, and M. A. Cohen, "Finding Early Invasive Breast Cancers: A Practical Approach," *Ridiology.* **248**(1), pp. 61–76, 2008.

[12] S. A. Feig, "Estimation of currently attainable benefit from mammographic screening of women aged 40 to 49 years," *Cancer* **75**(10), p. 2412, 1995.

[13] S. A. Feig, "Decreased breast cancer mortality through mammographic screening: results of clinical trials," *Radiology* **167**, pp. 659–665, 1988.

[14] L. Humphrey, M. Helfand, B. Chan, and S. Woolf, "Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force," *Ann Intern Med.* **137**, pp. 347–60, 2002.

[15] Institute of Medicine, "Saving women lives: integration and innovation: a framework for progress in early detection and diagnosis of breast cancer," *Washington, D.C.:National Academies Press.* , 2005.

[16] S. Heywang-Krunner, I. Schreer, W. Heindel, and A. Katalinic, "Imaging studies for the early detection of breast cancer," *Dtsch Arztebl Int.* **105**(31-32), pp. 541–7, 2008.

[17] K. Kerlikowske, D. Grady, J. Barclay, E. Sickles, and V. Ernster, "Effect of age, breast density, and family history on the sensitivity of first screening mammography," *JAMA.* **276**, pp. 33–8, 1996.

[18] S. A. Feig and M. J. Yaffe, "Digital Mammography," *RadioGraphics* **18**, pp. 893–901, 1998.

[19] W. T. Yang, C. Lai, G. J. Whitman, W. A. M. Jr., M. J. Dryden, A. C. Kushwaha, A. A. Sahin, D. Johnston, P. J. Dempsey, and C. C. Shaw, "Comparison of Full-Field Digital Mammography and Screen-Film Mammography for Detection and Characterization of Simulated Small Masses," *AJR.* **187**(6), pp. W576–W581, 2006.

[20] M. J. Yaffe and J. G. Mainprize, "Detectors for Digital Mammography," *Technology in Cancer Research and Treatment.* **3**(4), pp. 309–324, 2004.

[21] J. M. Lewin, C. J. D'Orsi, R. E. Hendrick, L. J. Moss, P. K. Isaacs, A. Karellas, and G. R. Cutter, "Clinical Comparison of Full-Field Digital Mammography and Screen-Film Mammography for Detection of Breast Cancer," *AJR* **179**(3), pp. 671–677, 2002.

[22] E. D. Pisano, R. E. Hendrick, M. J. Yaffe, J. K. Baum, S. Acharyya, J. B. Cormack, L. A. Hanna, E. F. Conant, L. L. Fajardo, L. W. Bassett, C. J. D'Orsi, R. A. Jong, M. Rebner, A. N. A. Tosteson, and C. A. Gatsonis, "Diagnostic Accuracy of Digital versus Film Mammography: Exploratory Analysis of Selected Population Subgroups in DMIST," *Radiology* **246**, pp. 376–383, 2008.

[23] R. M. Nishikawa, S. Acharyya, C. Gatsonis, E. D. Pisand, E. B. Cole, H. S. Marques, C. J. D'Orsi, D. M. Farria, K. M. Kanal, M. C. Mahoney, M. Rebner, M. J. Staiger, and F. the Digital Mammography Image Screening Trial investigator group, "Comparison of Soft-copy and Hard-copy Reading for Full-Field Digital Mammography," *Radiology.* **251**(1), pp. 41–49, 2009.

[24] A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of Human Visual Signal Discrimination," *Science* **214**, pp. 93–94, 1981.

[25] A. Burgess and R. Wagner, "Human signal detection performance for noisy medical images ," *Proceedings of IEEE Computer Society International Workshop on Medical Imaging* , 1982.

[26] K. J. Myers, H. H. Barrett, M. C. Borgstrom, D. D. Patton, and G. W. Seeley, "Effect of noise correlation on detectability of disk signals in medical imaging," *JOSA.* **2**(10), pp. 1752–1759, 1985.

[27] L. B. Lusted, "Signal Detectability and Medical Decision-Making," *Science* **171**(3977), pp. 1217–1219, 1971.

[28] C. E. Metz, "ROC Methodology in Radiologic Imaging," *Invest. Radiol.* **21**(9), pp. 720–733, 1986.

[29] H. H. Barrett and K. J. Myers, *Foundations of Image Science*, Wiley Interscience, New Jersey, 2004.

[30] A. E. Burgess, "Comparison of receiver operating characteristic and forced choice observer performance measurement methods," *Med. Phys.* **22**(5), pp. 643–655, 1995.

[31] ICRU, *Medical Imaging-the Assessment of Image Quality.Report 54.*, International Commission of Radiation Units and Measurements, Bethesda, MD, 1996.

[32] J. Y. HH Barrett, J. Rolland, and K. Myers, "Model Observers for Assessment of Image Quality," **90**(21), pp. 9758–9765, 1993.

[33] R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers, "Hotelling trace criterion and its correlation with human-observer performance," **4**(5), pp. 945–953, 1987.

[34] B. D. Gallas, "Variance of the channelized-hotelling observer from a finite number of trainers and testers," in *Medical Imaging 2003: Image Perception, Observer Performance, and Technology Assessment*, **5034**, p. 100, SPIE, 2003.

[35] C. U.S. Department Of Health and Human Services, Food and Drug Administration, *Class II Special Controls Guidance Document: Full Field Digital Mammography System*, 2010.

[36] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt, and J. M. Boone, *The Essential Physics of Medical Imaging (2nd edition)*, Lippincott Williams & Wilkins, Philadelphia, 2002.

[37] J. Sandrik and R. Wagner, "Absolute measures of physical image quality: Measurement and application to radiographic magnification," *Med. Phys.* **9**, p. 540, 1982.

[38] C. E. Metz and K. Doi, "Transfer function analysis of radiographic imaging systems," *Phys. Med. Biol.* **24**, p. 1079, 1979.

[39] M. L. Giger and K. Doi, "Investigation of basic imaging properties of digital radiography. Part 1: modulation transfer function," *Med. Phys.* **11**, pp. 287–295, 1984.

[40] J. C. Dainty and R. Shaw, *Image Science*, Academic, London, 1974.

[41] R. M. G. R. J. Jennings, P. W. Quinn and T. R. Fewell, "Evaluation of x-ray sources for mammography," *SPIE Med. Imag.* **1896**, pp. 259–268, Proc. SPIE, 1993.

[42] R. J. Jennings, H. Jafroudi, R. M. Gagne, T. R. Fewell, P. W. Quinn, D. E. Steller Artz, J. J. Vucich, M. T. Freedman, and S. K. Mun, "Storage phosphor-based digital mammography using a low-dose x-ray system optimized for screen-film mammography," **2708**, pp. 220–232, SPIE, 1996.

[43] J. T. D. III, D. L. Ergun, L. Rutz, D. A. Hinshaw, H. Blume, and D. C. Clark, "DQE(f) of four generations of computed radiography acquisition devices ," *Med. Phys.* **22**(10), pp. 1581–1593, 1995.

[44] R. M. Nishikawa and M. J. Yaffe, "Signal-to-noise properties of mammographic filmCscreen systems," *Med. Phys.* **12**(1), pp. 32–39, 1985.

[45] R. M. Gagne, B. D. Gallas, and K. J. Myers, "Toward objective and quantitative evaluation of imaging systems using images of phantoms," **33**(1), pp. 83–95, 2006.

[46] I. S. Kyprianou, B. Gallas, A. Badano, , S. Park, H. Liu, and K. J. Myers, "Noise and signal detection in digital x-ray detectors using the spatial definition of SNR," in *Medical Imaging 2009: Physics of Medical Imaging*, M. J. Flynn and J. Hsieh, eds., **7258**, p. 725819, SPIE, 2009.

[47] K. C. Young, J. J. H. Cook, and J. M. Oduko, "Comparison of software and human observers in reading images of the CDMAM test object to assess digital mammography systems," in *Medical Imaging 2006: Physics of Medical Imaging*, M. J. Flynn and J. Hsieh, eds., **6142**, p. 614206, SPIE, 2006.

[48] C.-Y. J. Yang and R. V. Metter, "The variability of software scoring of the CDMAM phantom associated with a limited number of images," in *Medical Imaging 2007: Physics of Medical Imaging*, J. Hsieh and M. J. Flynn, eds., **6510**, p. 65100C, SPIE, 2007.

[49] L. Fletcher-Heath and R. V. Metter, "Quantifying the performance of human and software CDMAM phantom image observers for the qualification of digital mammography systems," in *Medical Imaging 2005: Physics of Medical Imaging*, M. J. Flynn, ed., **5745**, pp. 486–498, SPIE, 2006.

[50] I. Kyprianou, A. Badano, B. Gallas, and K. J. Myers, "Singular value description of a digital radiographic detector: theory and measurements," *Med. Phys.* **35**(10), pp. 4744–4756, 2008.

[51] K. C. Young, J. J. H. Cook, and J. M. Oduko, "Efficiency of the human observer compared to an ideal observer based on a generalized NEQ which incorporates scatter and geometric unsharpness: evaluation with a 2AFC experiment," in *Medical Imaging 2005: Image Perception, Observer Performance, and Technology Assessment*, M. J. Flynn and J. Hsieh, eds., **5749**, p. 251, SPIE, 2005.

[52] A. Schnitzler, "Image detector model and parameters of the human visual system," *JOSA.* **63**(11), pp. 1357–1368, 1973.

[53] R. F. Wagner, "Toward a unified view of radiological imaging systems. Part II: Noisy images," *Med. Phys.* **4**, pp. 279–296, 1977.

[54] R. F. Wagner, "Fast Fourier digital quantum mottle analysis with application to rare earth intensifying screen systems," *Med. Phys.* **4**, pp. 157–162, 1977.

[55] R. F. Wagner, G. T. Barnes, and B. S. Askins, "Effect of reduced scatter on radiographics information content and patient exposure: A quantitative demonstration," *Med. Phys.* **7**, pp. 13–18, 1980.

[56] E. P. Muntz, "Analysis of the significance of scattered radiation in reduced dose mammography, including magnification effects, scatter suppression, and focal spot and detector blurring," *Med. Phys.* **6**, pp. 110–117, 1979.

[57] J. M. Boone, B. A. Arnold, and J. A. Seibert, "Characterization of the point spread function and modulation transfer function of scattered radiation using a digital imaging system," *Med. Phys.* **13**(2), pp. 254–256, 1986.

[58] V. N. Cooper(III), J. M. Boone, J. A. Seibert, and C. J. Pellot-Barakat, "An edge spread technique for measurement of the scatter-to-primary ratio in mammograph," *Med. Phys.* **27**(5), pp. 845–853, 2000.

[59] K. Doi and K. Rossman, "The effect of radiographic magnification on blood vessel imaging with various screen-film systems," *Med. Phys.* **1**(5), pp. 257–261, 1974.

[60] C. C. Shaw, X. Liu, M. Lemacks, J. X. Rong, and G. J. Whitman, "Optimization of MTF and DQE in magnification radiography - a theoretical analysis," *Proc. SPIE* **3977**, pp. 466–475, 2000.

[61] A. Ganguly, S. Rudin, D. R. Bednarek, K. R. Hoffmann, and I. S. Kyprianou, "Microangiography for Neuro-vascular Imaging, Part I: Experimental Measurements and Feasibility," *Med. Phys.* **30**(11), pp. 3018–3028, 2003.

[62] I. S. Kyprianou, "A method for total x-ray imaging system evaluation application to a microangiographic detector for neurovascular procedures," *PhD thesis. University of New York at Buffalo.* , pp. 22–25, 2004.

[63] I. S. Kyprianou, S. Rudin, D. R. Bednarek, and K. R. Hoffmann, "Generalizing the MTF and DQE to include x-ray scatter and focal spot unsharpness: application to a new microangiographic system," *Med. Phys.* **32**(2), pp. 613–26, 2005.

[64] E. Samei, N. T. Ranger, A. MacKenzie, I. D. Honey, J. T. Dobbins(III), and C. E. Ravin, "Detector or System? Extending the Concept of Detective Quantum Efficiency to Characterize the Performance of Digital Radiographic Imaging Systems," *Radiology* **249**, pp. 926–937, 2008.

[65] H. Liu, I. S. Kyprianou, A. Badano, K. J. Myers, R. J. Jennings, S. Park, R. V. Kaczmarek, and K. Chakrabarti, "SKE/BKE Task-based methodology for calculating Hotelling observer SNR in mammography," in *Medical Imaging 2009: Physics of Medical Imaging*, M. J. Flynn and J. Hsieh, eds., **7258**, p. 72581D, SPIE, 2009.

[66] P. Monnin, N. W. Marshall, H. Bosmans, F. O. Bochud, and F. R. Verdun, "Image quality assessment in digital mammography: part II. NPWE as a validated alternative for contrast detail analysis," *Phys. Med. Biol.* **56**, pp. 4221–4238, 2011.

[67] A. Noel and F. Thibault, "Digital detectors for mammography: the technical challenges," *European Radiology.* **14**(11), pp. 1990–1998, 2004.

[68] F. Jeunehomme, R. Iordache, S. Muller, G. Mawdsley, and M. Yaffe, "Controlling graylevel variation in contrast-enhanced digital mammography: design of a calibration procedure," *SPIE Med. Imag.* **5030**, pp. 338–348, SPIE, 2003.

[69] J. E. Gray and J. A. Princehorn, "HTC Grids Improve Mammography Contrast," *White Paper* , W-BI-HTC(9/04).

[70] E. D. Pisano and M. J. Yaffe, "Digital Mammography," *Radiology.* **234**, pp. 353–362, 2005.

[71] R. Visser and N. Karssemeijer, "Manual CDCOM version 1.5: software for automated readout of CDMAM 3.4 images," 2007.

[72] D. R. Dance, C. L. Skinner, K. C. Young, J. R. Beckett, and C. J. Kotre, "Additional factors for the estimation of mean glandular breast dose using the UK mammography dosimetry protocol," *Phys. Med. Biol.* **45**, pp. 3225–3240, 2000.

[73] D. R. Dance, "Monte Carlo calculation of conversion factors for the estimation of mean glandular breast dose," *Phys. Med. Biol.* **35**(9), pp. 1211–1219, 1990.

[74] R. F. Wagner and D. G. Brown, "Unified SNR analysis of medical imaging systems," **30**(6), pp. 489–518, 1985.

[75] J. T. Dobbins(III), D. L. Ergun, L. Rutz, D. A. Hinshaw, H. Blume, and D. C. Clark, "DQE(f) of four generations of computed radiography acquisition devices," *Med. Phys.* **22**(10), pp. 1581–1593, 1995.

[76] G. K. Yadava, I. S. Kyprianou, S. Rudin, D. R. Bednarek, and K. R. Hoffmann, "Generalized performance evaluation of x-ray image intensifier compared with a microangiographic system," *SPIE Med. Imag.* **5745**, p. 419, SPIE, 2005.

[77] G. K. Yadava, S. Rudin, A. T. Kuhls-Gilcrist, and D. R. Bednarek, "Generalized Objective Performance Assessment of a New High-Sensitivity Microangiographic Fluoroscopic (HSMAF) Imaging System," *Proc Soc Photo Opt Instrum Eng.* **69130U**, pp. 1–11, 2008.

[78] "International Standard- MEDICAL ELECTRICAL EQUIPMENT CHARACTERISTICS OF DIGITAL X-RAY IMAGING DEVICES," *IEC 62220-1* , p. 18, 2003(E).

[79] S. Suryanarayanan, A. Karellas, S. Vedantham, H. Ved, S. P. Baker, and C. J. DOrsi, "Flat-Panel Digital Mammography System: Contrast-Detail Comparison between Screen-Film Radiographs and Hard-Copy Images," *Radiology* **225**, pp. 801–807, 2002.

[80] S. Rivetti, N. Lanconelli, R. Campanini, M. Bertolini, G. Borasi, A. Nitrosi, C. Danielli, L. Angelini, and S. Maggi, "Comparison of different commercial FFDM units by means of physical characterization and contrast-detail analysis," *Med. Phys.* **33**(11), pp. 4198–4209, 2006.

[81] K. C. Young, A. Alsager, J. M. Oduko, H. Bosmans, B. Verbrugge, T. Geertse, and R. V. Engen, "Evaluation of software for reading images of the CDMAM test object to assess digital mammography systems," *SPIE Med. Imag.* **6913**, p. 69131C, SPIE, 2008.

[82] K. C. Young, J. J. H. Cook, J. M. Oduko, and H. Bosmans, "Comparison of software and human observers in reading images of the CDMAM test object to assess digital mammography systems," *SPIE Med. Imag.* **6142**, p. 614206, SPIE, 2006.

[83] S. Rivetti, N. Lanconelli, R. Campanini, M. Bertolini, G. Borasi, A. Nitrosi, C. Danielli, L. Angelini, and S. Maggi, "Comparison of different commercial FFDM units by means of physical characterization and contrast-detail analysis," *Med. Phys.* **33**(11), pp. 4198–4209, 2006.

[84] A. E. Burgess, "The Rose model, revisited ," *JOSA.* **16**(3), pp. 633–646, 1999.

[85] M. J. Tapiovaara and R. Wagner, "SNR and DQE analysis of broad spectrum X-ray imaging ," *Phys. Med. Biol.* **30**(6), pp. 519–529, 1985.

[86] J. Bozek, M. Mustra, K. Delac, and M. Grgic1, "A Survey of Image Processing Algorithms in Digital Mammography," *Rec. Advan. in Mult. Sig. Process. and Commun.* **231**, pp. 631–657, 2009.

[87] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, 2002.

[88] T. Matsubara, H. Fujita, and T. Endo, "Development of mass detection algorithm based on adaptive thresholding technique in digital mammograms," pp. 391–396, The 3rd International Workshop on Digital Mammography, 1996.

[89] L. Breiman, J. H. Fredman, R. A. Olshen, and C. J. Stone, *Classification And Regression Trees*, Chapman and Hall, New York, NY, 1984.

[90] N. Szekely, N. Toth, and B. Pataki, "A Hybrid System for Detecting Masses in Mammographic Images," *IEEE Transactions on Instrumentation and Measurement* **55**(3), pp. 944–951, 2006.

[91] N. Petrick, H. P. Chan, B. Sahiner, and D. Wei, "An Adaptive Density Weighted Contrast Enhancement Filter for Mammographic Breast Mass Detection," *IEEE Transactions on Medical Imaging* **15**(1), pp. 59–67, 1996.

[92] F. Fauci, S. Bagnasco, R. Bellotti, D. Cascio, S. C. C. F. D. Carlo, G. D. Nunzio, M. E. Fantacci, G. Forni, A. Lauria, E. L. T. adn R. Magro, G. L. Masala, P. Oliva, M. Quarta, G. Raso, A. Retico, and S. Tangaro, "Mammogram Segmentation by Contour Searching and Massive Lesion Classification with Neural Network," **5**, pp. 2695–2699, 2004.

[93] S. Osher and L. I. Rudin, "Feature-oriented image enhancement using shock filters," *SIAM J. Numer. Anal.* **27**(4), 1990.

[94] M. Nitzberg and T. Shiota, "Nonlinear Image Filtering With Edge And Corner Enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(8), pp. 826–833, 1992.

[95] I. Pollak, A. S. Willsky, and H. Krim, "Image Segmentation and Edge Enhancement with Stabilized Inverse Diffusion Equations," *IEEE Transactions on Image Processing* **9**(2), pp. 256–266, 2000.

[96] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **7**(2), pp. 165–177, 1985.

[97] J. Lee, "Digital Image Smoothing And The Sigma Filter," *Computer Vision, Graphics, and Image Processing* **24**(2), pp. 255–269, 1983.

[98] P. M. Narendra, "A Separable Median Filter for Image Noise Smoothing," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **3**(1), pp. 20–29, 1981.

[99] L. Alvarez, P. L. Lions, and J.-M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion, II," *SIAM J. Numer. Anal.* **29**(3), 1992.

[100] K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," *JOSA. A.* **4**(12), pp. 2447–2457, 1987.

[101] H. H. Barrett, C. K. Abbey, and B. D. Gallas, "Stabilized estimates of Hotelling-observer detection performance in patient-structured noise," *SPIE Med. Imag.* **3340**(27), SPIE, 1998.

[102] S. Park, H. H. Barrett, M. A. Kupinski, and K. J. Myers, "Performance of a channelized-ideal observer using Laguerre-Gauss channels for detecting a Gaussian signal at a known location in different lumpy backgrounds," *Proc. SPIE* **6146**, p. 61460P, 2006.

[103] S. Park, H. H. Barrett, E. Clarkson, M. A. Kupinski, and K. J. Myers, "Channelized-ideal observer using Laguerre-Gauss channels in detection tasks involving non-Gaussian distributed lumpy backgrounds and a Gaussian signal," *J Opt Soc Am A Opt Image Sci Vis.* **24**(12), pp. B136–B150, 2007.

[104] B. D. Gallas and H. H. Barrett, "Validating the use of channels to estimate the ideal linear observer," *JOSA. A.* **20**(9), pp. 1725–1738, 2003.