Evaluation of a Multi-Scale Enhancement Protocol for Digital Mammography.

Ralf Mekle[a], Andrew Laine[a, b], Suzanne Smith[b], Cory Singer[b], Tova Koenigsberg[b], and Marc Brown[b]

[a] Department of Biomedical Engineering, Columbia University
[b] Department of Radiology, Columbia-Presbyterian Medical Center

ABSTRACT

We have carried out a receiver operating characteristics (ROC) study for the enhancement of mammographic features in digitized mammograms. The study evaluated the benefits of multi-scale enhancement methods in terms of diagnostic performance of radiologists. The enhancement protocol relied on multi-scale expansions and non-linear enhancement functions. Dyadic spline wavelet functions (first derivative of a cubic spline) were used together with a sigmoidal non-linear enhancement function[1, 2]. We designed a computer interface on a softcopy display and performed an ROC study with three radiologists, who specialized in mammography. Clinical cases were obtained from a national mammography database of digitized radiographs prepared by the University of South Florida (USF) and Harvard Medical School.
Our study focused on dense mammograms, i.e. mammograms of density 3 and 4 on the American College of Radiology (ACR) breast density rating, which are the most difficult cases in screening, were selected. To compare the performance of radiologists with and without using multi-scale enhancement, two groups of 30 cases each were diagnosed. Each group contained 15 cases of cancerous and 15 cases of normal mammograms. Conventional ROC analysis was applied, and the resulting ROC curves indicated improved diagnostic performance when radiologists used multi-scale non-linear enhancement.

Keywords: Spline wavelets, ROC analysis, contrast enhancement, digital mammography, softcopy display.

1. INTRODUCTION

Recently, research has focused on the development of digital displays and softcopy workstations for digital mammography. Limited spatial resolution, luminance, and dynamic range cannot be solved simply by hardware improvements or computer programming alone. A possible solution of these problems is the application of multi-scale contrast enhancement techniques derived from non-linear models.
Radiologists are mostly familiar with films where the Modulation Transfer Function (MTF) is approximately equal to 2^8 gray levels of contrast resolution. However, images acquired with digital detectors can record at least 2^{12} distinct gray levels of intensity and are now commercially available. The wealth of dynamic range within these digital acquisition systems provides strong evidence that the SNR can be increased in digital mammography. For expert radiologists the human visual system can detect at most 2^7 shades of gray. These considerations motivate the need for judicious methods of processing of digital radiographs that can optimize the bandwidth of the human visual system. We have designed enhancement software that is well adapted for this purpose and provides a "data mining" tool to map and make visible selected "quantum levels" of information living within the wide range of contrast resolution provided by digital detectors.

Medical imaging is a field where quantitative accuracy and qualitative fidelity are paramount. In any image enhancement process distortion of the original image and artifacts are not affordable. Multidimensional feature enhancement via wavelet analysis has been previously demonstrated on mammograms[2, 3, 4] and is a powerful tool for processing digital medical images without artifacts. The enhancement process adjusts multi-scale coefficients at some particular spatial-frequency scale by increasing, decreasing or resetting their values. Each image is then reconstructed with modified coefficients. This simple enhancement technique relies on the idea that features of interest in a given radiograph are detectable at a particular scale and can be amplified, whereas noise and less clinically interesting features may live at other levels of analysis whose visual appearance can be diminished or eliminated in a reconstructed image. Further results and detailed descriptions of these methods can be found in[5, 6, 7, 8, 9].
Surprisingly, there have been very few studies carried out to evaluate the benefits of multi-scale enhancement methods in terms of diagnostic performance. Our study aimed at providing quantitative evidence of these benefits. ROC analysis[10] is most commonly used in medical imaging for such purposes, though alternative statistical approaches can be found as well[11].

Wavelet Applications in Signal and Image Processing VIII, Akram Aldroubi, Andrew F. Laine, Michael A. Unser, Editors, Proceedings of SPIE Vol. 4119 (2000) © 2000 SPIE · 0277-786X/00/$15.00

ROC curves have been compared to evaluate the visibility of malignancies[12], mass detection techniques[13] and algorithms for computer-aided diagnosis (CAD) that use neural networks[14].

The paper is organized as follows: In Section 2 we describe a protocol for multi-scale non-linear contrast enhancement. After a short overview of the use of multi-scale expansions for contrast enhancement we discuss the dyadic spline wavelet selected, its implementation, and how a non-linear enhancement function is applied to multi-scale coefficients. Section 3 addresses the design of a graphical user interface (GUI) that was developed to carry out the ROC study including high-performance displays and specialized hardware for softcopy display of digital mammograms. Next, the ROC study itself together with its results and subsequent data analysis is discussed in Section 4. After a discussion of the results of the study, conclusions and possible directions of future work are presented in Section 5.

2. ENHANCEMENT PROTOCOLS FOR DIGITAL MAMMOGRAPHY

2.1. Contrast Enhancement via Multi-scale Expansions: A Short Overview

We summarize below, the advantages of the use of overcomplete multi-scale representations for adaptive contrast enhancement of digital mammograms. Critically sampled multi-scale representations are not suitable for detection and enhancement tasks because of aliasing effects introduced during downsampling of the analysis[15, 16]. However, overcomplete representations avoid such aliasing artifacts and offer the desirable property for image enhancement, of being shift invariant[17]. Indeed, this property ensures that the spatial locations of any mammographic finding within an image are preserved across all scales. Thus, in our approach the transform coefficient matrix size at each scale remains the same as the original spatial resolution of the digital mammogram, since there is no downsampling across each level of analysis.

Overcomplete multi-scale analysis and reconstruction algorithms using dyadic scales previously developed in[2] and[18] were used as an initial choice of analysis function for our enhancement protocol. The implementation was carried out using several lowpass filters and highpass filters with localized frequency support. At each level of the multi-scale expansion an input image is decomposed into a coarse approximation and detailed structures. The coarse approximation is the output from applying a lowpass filter, and the detailed structures are obtained from highpass filtering. A gain or enhancement function modifies the matrices of coefficients that have been isolated by the filters at each level and may boost coefficients at some scales and/or attenuate others. The filter bank implementation of enhancement processing by an expansion-reconstruction algorithm for 2 levels of analysis is schematically illustrated in Figure 1.

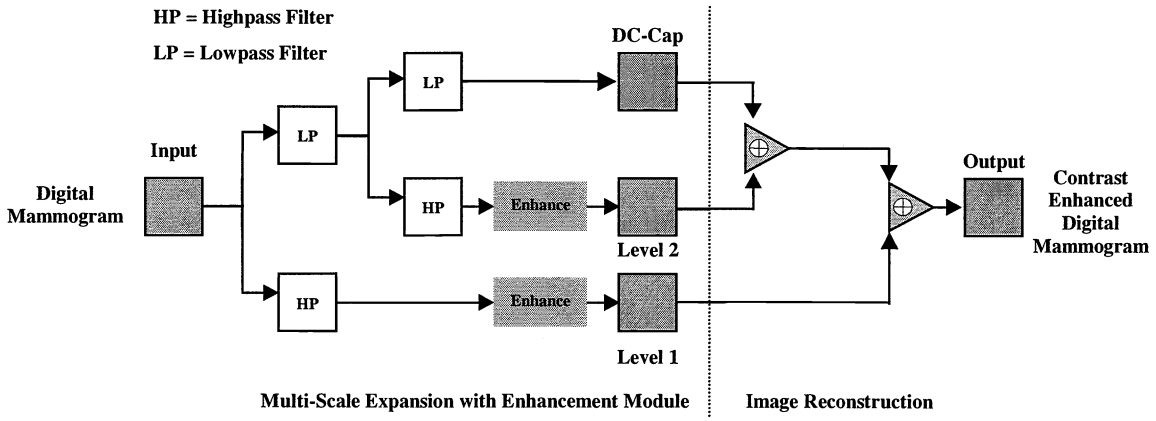


Figure 1: Multi-scale analysis with non-linear contrast enhancement: Schematic of filter bank implementation. In the left part multi-scale expansion with enhancement for 2 levels of analysis is shown, and reconstruction is presented (in a simplified manner) in the right part.

The modified matrices of coefficients are simply "plugged in" during reconstruction producing a "focused" subband enhancement. As shown above, the enhancement function can be implemented independently of a particular set of filters and easily incorporated into a filter bank to provide the benefits of multi-scale enhancement[1, 19].

2.2. High Speed Implementation to Support Interactive Processing

Similar to orthogonal and biorthogonal discrete wavelet transforms[20], the discrete dyadic wavelet transform can be implemented within a hierarchical filtering scheme. Let an input signal $x(n)$ be real, $x(n) \in l^1(Z)$, $n \in [0, N-1]$. Filtering a finite input signal might cause artifacts at the boundaries. A common remedy for such a problem is realized by constructing a

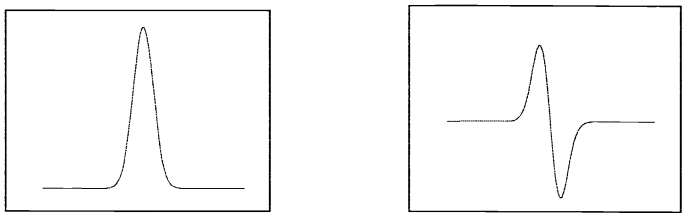
mirror extended signal $x_{me}(n)$ to be supported in [-N, N-1]. In[1] it is shown how a mirror extension is particularly elegant solution in conjunction with symmetric/antisymmetric filters. The optimized circular convolution described in[1] was implemented in native "ANSI C" to speed up performance for multi-scale decomposition and image reconstruction. This algorithm was incorporated into the graphical user interface (GUI) developed during the preparation of the study.

As a further goal, we envision developing feature specific enhancement protocols. An enhancement protocol would consist of a multi-scale expansion of a mammogram by a specific basis and an associated non-linear enhancement function that is best matched to a specific type of lesion, e.g. microcalcifications. For the study under consideration, a dyadic spline wavelet function was used as the basis, and a non-linear sigmoidal function was applied as the enhancement function. Both are described in greater detail next.

2.3. Dyadic Spline Wavelet Algorithm

The wavelet transform of a signal $f(x)$ at scale s and position x is defined by $W_s f(x) = f * \psi_s(x)$, where $\psi_s(x) = \frac{1}{s}\psi(\frac{x}{s})$, and $\psi(x)$ is the mother wavelet of zero average. To allow fast numerical implementation of discrete wavelet transforms, Mallat and Zhong[21] introduced a dyadic wavelet transform, where the scale parameter varies only along the dyadic sequence $\{2^j\}$ with $j \in Z$. The 2-D dyadic wavelet transform partitions plane orientations into two bands. This means that there are two channels of analysis along orthogonal x and y directions. The wavelet transform of a 2D signal $f(x,y)$ at scale 2^j has two components defined by: $W^1_{2^j} f(x,y) = f * \psi^1_{2^j}(x,y)$ and $W^2_{2^j} f(x,y) = f * \psi^2_{2^j}(x,y)$, with $\psi^d_{2^j}(x,y) = \frac{1}{2^{2^j}}\psi^d(\frac{x}{2^j}, \frac{y}{2^j})$, $d=1,2$.

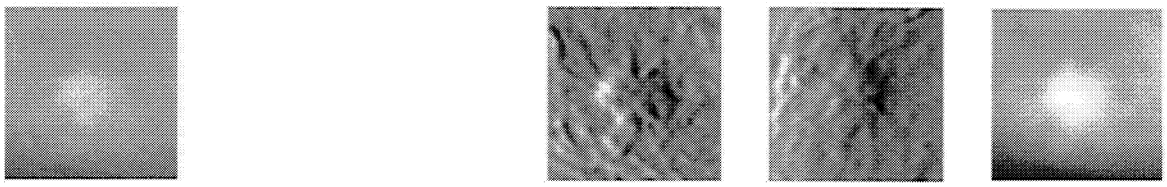
We used the quadratic spline wavelet function $\psi(x)$ defined by Mallat and Zhong in[21] of compact support and continuously differentiable. It is the derivative of a cubic spline smoothing function $\theta(x)$. These functions are displayed in Figure 2 below.



(a) (b)

Figure 2: (a) Cubic spline smoothing function $\theta(x)$. (b) Quadratic spline wavelet $\psi(x)$ of compact support defined as the derivative of the smoothing function.

In this context, the wavelet transform $W^d_{2^j} f$ of the signal f is proportional to the derivative of the signal smoothed at the scale 2^j. The coefficients of modulus maxima detection are then equivalent to an adaptive sampling that finds signal variation points in the two orthogonal directions x and y. As images represent finite energy signals measured at some finite resolution, we cannot compute the wavelet transform at scales below the limit set by this resolution. We applied this analysis at dyadic scales varying from 1 (original signal) to the limit imposed by acquisition (digitizer sampling rate). Figure 3 shows an example for one level of an overcomplete wavelet expansion of a spiculated mass at a dyadic scale.



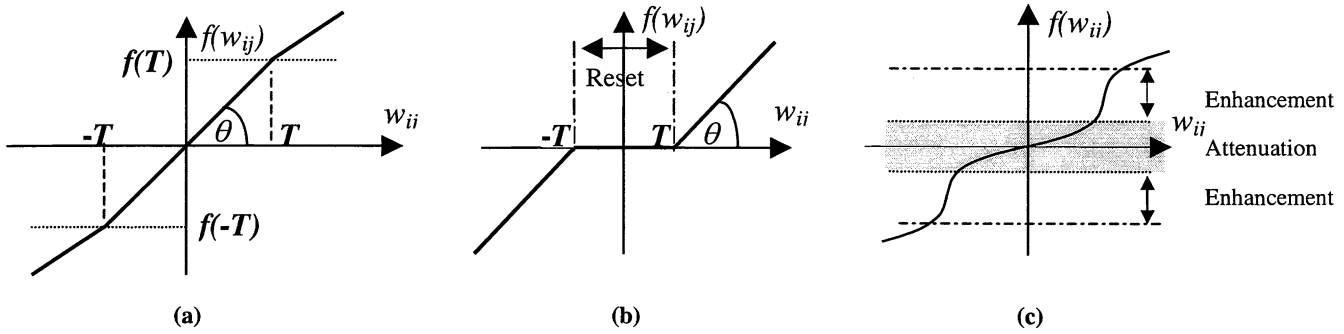
(a) (b) (c) (d)

Figure 3: Level 5 of an overcomplete dyadic wavelet expansion of a spiculated mass. (a) Original image. (b) Horizontal details. (c) Vertical details. (d) Approximation image.

2.4. Non-Linear Enhancement Function

Modification of selected analysis coefficients within a certain scale can make more obvious indiscernible or barely seen mammographic features[7]. Contrast enhancement was achieved by applying a non-linear function to transform coefficients at selected scales. This operation results in local attenuation or amplification of coefficients. Enhancement functions must be

cumulative and monotonically increasing in order to preserve the order of intensity information in the original image and to avoid artifacts[22]. Figure 4(a) provides a very simple example of a piecewise linear enhancement function. The parameter w_{ij} represents a multi-scale coefficient. Multi-scale coefficients w_{ij} are modified by an enhancement function $f(w_{ij})$. T is the threshold of the function. Depending on the value of the angle θ there will be an attenuation or amplification of coefficients. Figure 4(b) displays a hard-thresholding function for denoising. Unfortunately, these two particular functions have the disadvantage of being discontinuous at the threshold value T. This could result in an abnormal distribution of coefficient values in the output. For this reason, smoother functions, like sigmoids, are preferable and were used in this study. Figure 4(c) shows an example of such a smooth function as described in[2].



(a) (b) (c)

Figure 4: (a) A simple piecewise linear enhancement function. (b) Hard-thresholding function. (c) A sigmoidal non-linear enhancement function.

The analytical formulation of the sigmoidal enhancement function as designed in[2] is the following:

$$f(\mathrm{w_{ij}}) = a\left[sigm\left(c(\mathrm{w_{ij}} - b)\right) - sigm\left(-c(\mathrm{w_{ij}} + b)\right)\right]$$
$$a = \frac{1}{sigm\left(c(1-b)\right) - sigm\left(-c(1+b)\right)},\ 0 < b < 1; \quad sigm(y) = \frac{1}{1+e^{-y}} \tag{1}$$

Parameters b and c control the threshold and the amount of enhancement (gain), respectively. This enhancement function is continuous, monotically increasing, and has a continuous first derivative. This ensures that application of the function will not introduce new discontinuities of coefficients in the transform domain.

From Figure 4(c) we see that this enhancement function decreases the value of the coefficients around zero, which is equivalent to a denoising action, while it may increase values of coefficients outside this range, equivalent to enhancement. This type of enhancement function, in 'steps', offers a very rich and flexible paradigm to carry out non-linear dynamic analysis of coefficients within a specific scale[23].

In general, non-linear estimators are signal dependent and behave differently for different realizations of a signal. In this context, Johnstone and Donoho have shown that by considering the signal as deterministic, thresholding of wavelet coefficients gives a nearly optimal estimation of piecewise smooth functions[24]. Thresholding of wavelet coefficients performs an adaptive smoothing of the image by averaging the noisy areas and preserving or enhancing coefficients in areas of sharp transitions. For example, fine microcalcifications represent high frequency information of the image. Consequently, smooth amplification of coefficients within this particular spatial frequency range (in combination with possibly decreasing information in other spatial frequencies) will enhance these features of interest. Similar analysis can be done to enhance by focus low spatial frequency features such as masses. Since the computation of enhancement parameters uses data dependent information such as local or global coefficient variance, digital and digitized radiographs acquired under different imaging conditions are best processed independently to achieve optimal enhancement. In our work we used both coefficient variance computed with respect to a selected region-of-interest (ROI) and user input (see Section 3.2) to adapt the threshold and gain parameters.

3. DEVELOPMENT OF A GRAPHICAL USER INTERFACE (GUI)

3.1. Motivation

Running such an enhancement algorithm in batch mode might be sufficient for single experiments. However, adjustment of parameters tied to a data dependent enhancement function is slow because of the repeated need to decompose and reconstruct from modified coefficients. A more desirable solution would be to observe the results of modified multi-scale coefficients interactively and to continue the enhancement procedure, until results are visually satisfactory or the decision is made that no further improvement can be achieved. In addition, with introducing fixed enhancement protocols into a clinical screening

paradigm, the algorithm must be simple, fast, and user-friendly, i.e. usage of the algorithm should be familiar to the radiologist and intuitive. Since each radiologist may have preferences with respect to contrast in mammograms, it must be possible to adjust parameter settings to individual preferences. Thus, we designed a graphical user interface to facilitate carrying out such a study and to create a softcopy display prototype. We call this application a "test bed" softcopy display tool. Its first version was employed for the ROC study described in the next section.

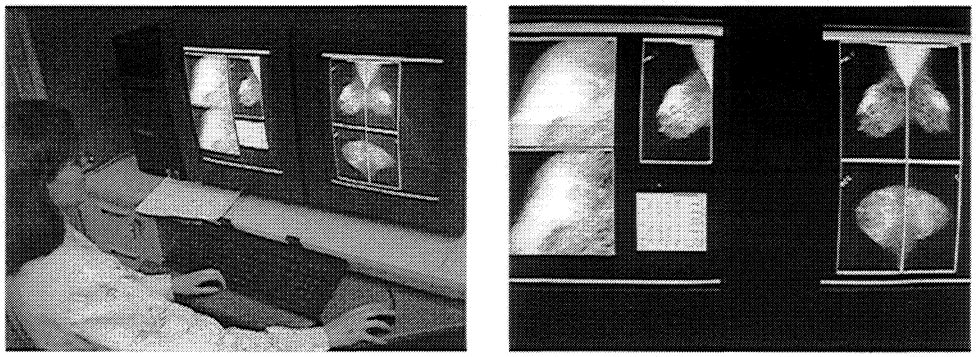
3.2. Design and Implementation

The graphical user interface (GUI) developed for this study was written in Visual C++ 6.0. The code for the wavelet expansion and image reconstruction was written in native "ANSI C" to speed up execution. The prototype interface was primarily designed to process raw 16-bit data. Data was obtained from a national mammography database of digitized radiographs provided by the University of South Florida (USF, "Digital Database for Screening Mammography" (DDSM)). Our database of digitized mammograms at the time of the study contained 586 selected cases of malignant lesions, biopsy proven, and 437 cases of normal breasts. 559 cases of dense breasts (density of 3 and 4 on the American College of Radiology (ACR) breast density rating) with 266 normals and 293 cancers, referred by radiologists as the most challenging cases, were included in our testing database.

Images from the mammography database were digitized from film at resolutions of 40 to 50 μm. Image line length varied between 2000 and 3000 pixels, and number of rows from 4000 to 5900 pixels. Depending on the scanner utilized for digitization the contrast resolution was either 12 bits or 16 bits per pixel resulting in 15-50 megabytes per view.

To handle this large amount of data and to provide the diagnosing radiologist with as much information as possible, all four views (right and left medial-lateral (RMLO, LMLO) and right and left cranial-caudal (RCC, LCC)) of a case were loaded into memory and displayed as downsampled images on display screen, which consisted of two high-resolution MegaScan monitors each with a screen size of 2048 by 2560 pixels. Specialized framebuffers allowed a display of 2^{10} gray levels (see Section 3.3). The four views were aligned to assist the radiologist to look for asymmetries. In addition, one view could be selected, and a viewport could display a selected region of interest (ROI) at full (original) resolution from a selected mammogram. Thus, the original mammogram could also be examined through this viewport, if desired. More importantly, suspicious areas could be captured in this viewport and processed through enhancement via the multi-scale expansion method described in Section 2.

Figure 5(a) shows Dr. Koenigsberg, one of three radiologists who participated in this investigation, during the ROC study. Figure 5(b) depicts a typical screen display of the GUI showing additional viewports described above.



(a) (b)

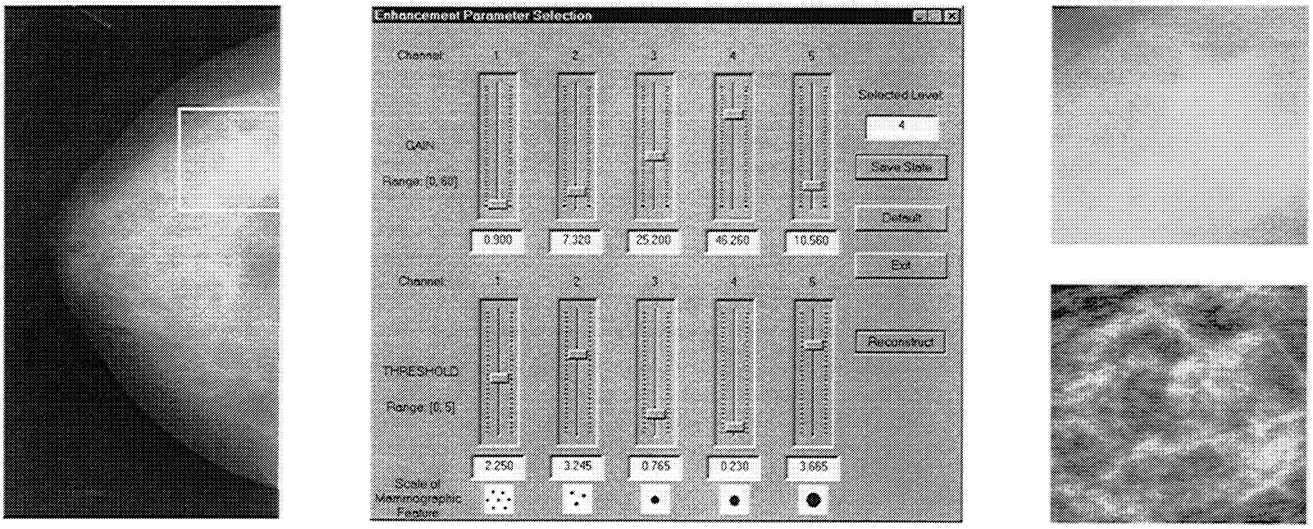
Figure 5: (a) Tova Koenigsberg, M.D., using the GUI during the preliminary ROC study described above. (b) Typical screen display used during the ROC study: four original digitized mammograms of one case on the right monitor, and a selected view, the GUI interface for parameter adjustments, original and enhanced ROI are shown on the left monitor.

As mentioned in Section 2.4 the shape of the enhancement function can be changed through modification of the two parameters gain and threshold. Therefore, each parameter could be adjusted by sliders for each level (subband) of the multi-scale expansion (see Figure 6(b)). On release of the slider button, a reconstruction "event" was "triggered", and a resulting image displayed in an output window. For example, reconstruction of a 512 by 512 matrix for five levels of decomposition (5 subbands) took 5 to 6 seconds. However, reconstruction time can certainly be improved to achieve true real-time performance, by employing faster algorithms.

After processing, enhanced images could be saved together with information about the location of an ROI to facilitate evaluation of a particular diagnosis for each case in comparison with the "ground truth" provided in the USF database. All suspicious areas in a case could be carefully examined by sequentially choosing multiple ROIs.

Figure 6(b) shows the test bed interface as an illustration. Selected subband coefficients at a particular level could be strongly suppressed by choosing large thresholds (> 2) and small gains (< 1), which can be desirable for the elimination of (structured and acquisition) noise, or normal benign anatomical (fibroglandular) structures.

Since the size of digital mammograms is quite large, an ROI (fixed at either 512 x 512 or 1024 x 1024) within the original image was chosen to avoid computing over regions that did not contain suspicious areas. This is also shown in Figure 6, where Figure 6(a) exhibits an original digitized mammogram with a 512 x 512 ROI that contains a possible mass. Figure 6(c) and Figure 6(d) display this ROI before and after enhancement via non-linear modification of multi-scale coefficients, respectively.



(a) (b) (d)

Figure 6: (a) Original mammogram with selected ROI containing a mass, (b) Multi-Scale Contrast Enhancement (MSCE) GUI, (c) Original ROI, and (d) Enhanced ROI.

3.3. Display and Hardware Settings

The enhancement protocol was executed on an IBM IntelliStation Z Pro Professional Workstation Type 6865. This machine had two Intel Pentium II Xeon microprocessors (450 MHz), 512 Mbytes of RAM and was equipped with 36 Gbytes of hard disk space. Windows NT 4.0 with service pack 4 was the operating system.

To explore the richness of information quantized at 16-bit per pixel (bpp) grayscale data, the IBM workstation was equipped with two Metheus P1540 Graphics controllers. These are high-resolution display subsystems with a digital-to-analog converter (DAC) capable of 1024 shades of gray An extended hardware palette could be accessed through specialized "C" function calls allowing the display of mammograms at 12-bit resolution. Two high-resolution MegaScan monitors were attached to this workstation providing a dual headed display on a single logical frame buffer or virtual desktop of 4096×2560 pixels. Both monitors were calibrated to correct for non-linearity through gamma correction. Finally, lighting conditions were controlled for the ROC study to model reading room conditions.

4. DESCRIPTION OF THE RECEIVER OPERATING CHARACTERISTICS (ROC) STUDY

The first receiver operating characteristics (ROC) study focused on overcomplete dyadic wavelets for enhancement of mammographic features in digitized mammograms. Specifically, dyadic spline wavelet functions were used together with a sigmoidal non-linear enhancement function explicitly described in Section 2. The ROC study included three radiologists specialized in mammography. The Director of the Breast Imaging Center at Columbia-Presbyterian Medical Center, Dr. Suzanne Smith, assisted in the selection of cases.

4.1. Selection of Cases

To measure the benefits of diagnosing digitized mammograms with enhancement through multi-scale expansions, we focused on dense mammograms, i.e. mammograms of density 3 and 4 on the American College of Radiology (ACR) breast density rating, which are the most difficult cases in screening. In general, the enhancement protocol aimed at improving the detection and localization of mammographic features, such as microcalcifications, masses, and spicular lesions without introducing "false-positives".

To compare the performance of radiologists with and without using the enhancement tool, two groups of 30 cases each were presented. Each group contained 15 cases of cancerous and 15 cases of normal mammograms. As mentioned above, a national mammography database provided "ground truth" (mostly through biopsy) for the selected cases. The selection was carried out very carefully under the guidance of a mammographer (Dr. Smith). Due to time constraints the number of cases included was limited for this initial study.

4.2. Paradigm of Diagnosis of Study

For each case presented to the radiologist, the enhancement procedure followed was the following:

Paradigm A: Without Enhancement:

The radiologist made a diagnosis based only on the four original displays and the viewport. No processing of regions-of-interest (ROIs) was allowed.

Paradigm B: With Enhancement:

The radiologist selected an ROI in one of the views and could apply multi-scale enhancement. Four levels of coefficients were computed. The radiologist then evaluated the quality of an enhanced ROI and adjusted the equalizer sliders of a channel to improve the visual quality of suspicious regions. Once he/she was satisfied with the visual result or if he/she judged that additional benefit could not be achieved, he/she made a diagnostic decision.

A diagnosis included specifying all lesions found and assigning a BI-RAD scale to each breast and the case. In addition, the radiologist was asked to choose a level of confidence (LOC) for each positive diagnosis, i.e. cancer is present, on an integer scale from 1 (definitely negative) to 5 (definitely positive). The value for the LOC was used in the analysis of data to decide whether a lesion was classified as malignant or benign (please see discussion of LOC ratings in Section 4.4).

4.3. ROC Data

Table 1 and Table 2 summarize the data acquired during the study. Group1 comprises the set of cases, where the radiologists were allowed to take advantage of the enhancement protocol, whereas Group 2 contains those cases, where no processing could be applied. Each of the tables shows the case numbers, the case designation and total number (#) of lesions for each case according to the mammography database (DB), and for each of the three mammographers the BI_RAD rating and LOC values. The BI_RAD rating was chosen from the standard categories 0-5 with 0 meaning that additional information for a more confident diagnosis was needed. In such cases, the radiologists were asked to select a BI_RAD rating different from 0, if they were asked to make a diagnosis without any additional information. This number is shown in parentheses for such cases.

In each table both groups are sorted into actually-negative cases (normals with "0" lesions) and actually-positive cases (cancers with, at least "1" malignant lesion), since this is required for subsequent analysis of the data.

4.4. ROC Analysis: General Principles

The most widely used method to objectively evaluate the performance of a diagnostic system or the difference in performance between two diagnostic systems is ROC analysis. It compares radiologists' image-based diagnoses with known states of disease and health. In ROC analysis, performance of a diagnostic system is described by the indices of "sensitivity" and "specificity", where "sensitivity" can be expressed as the true-positive fraction (TPF) and "specificity" by the true-negative fraction (TNF) of a diagnosis[10]. In a complimentary way, the false-negative fraction (FNF) and the false-positive fraction (FPF) can be defined as FNF = 1-TPF and FPF = 1-TPF, respectively. Due to this dependence, it is only necessary to measure one pair of indices, and frequently TPF and FPF are used (as in our study).

The underlying model for ROC analysis is the use of probability density distributions of a radiologist's confidence in a positive diagnosis for a particular diagnostic task for true positive and true negative patients. An ROC curve can be generated from pair values for TPF and FPF based on selection of a confidence threshold[10]. ROC curves that indicate better decision performance are positioned higher in the unit square spanned by FPF and TPF axes. The area under the ROC curve, A_z provides a useful summary index for the inherent discrimination performance of a diagnostic system. Thus, A_z is the average value of sensitivity or specificity, respectively of a corresponding ROC curve, if the specificity or sensitivity, respectively of the system is selected randomly between 0.0 and 1.0[10].

In practice, data for an ROC analysis is obtained by providing a set of rating categories to the radiologist. For a rating scale we chose discrete values from 1 to 5 for the level of confidence (LOC) in a positive diagnosis. The meaning of these values was as follows: (1) definitely or almost definitely negative, (2) probably negative, (3) possibly positive, (4) probably positive, and (5) definitely or almost definitely positive. With this choice the value for the LOC is similar to the standard BI_RAD rating scale used in screening.

Group1	(with Enhancement)							
			Mammographer 1		Mammographer 2		Mammographer 3	
Case #	Database	DB Total # of Lesions	BI RAD	LOC	BI RAD	LOC	BI RAD	LOC
2	A 0058	0	4	3	1	1	3	2
5	A 0069	0	1	2	1	1	1	1
6	A 0041	0	3	2	1	1	1	1
7	A 0077	0	3	2	2	1	2	1
9	A 0064	0	2	2	2	1	2	2
13	A 0067	0	0(3)	2	1	1	0(3)	3
15	A 0080	0	0(3)	3	2	1	2	1
16	A 0089	0	3	3	1	1	1	2
19	A 0062	0	2	2	1	1	2	1
21	A 0057	0	2	2	1	1	0(3)	3
24	A 0072	0	1	2	1	1	1	1
25	A 0070	0	1	2	0(3)	2	1	2
26	A 0068	0	1	2	1	1	2	1
28	A 0039	0	3	2	1	1	0(4)	3
30	A 0092	0	3	2	1	1	1	1
1	B 3044	1	4	4	4	4	4	3
3	B 3073	1	3	2	3	2	4	3
4	B 3006	1	5	5	5	5	5	5
8	B 3032	1	0(3)	2	5	4	4	4
10	B 3107	1	5	4	4	4	5	4
11	C 0060	1	0(3)	3	0	3	0(4)	3
12	B 3057	1	4	4	5	4	4	4
14	B 3078	1	5	4	5	4	0(4)	3
17	B 3033	1	0(3)	2	0	2	0(3)	3
18	B 3031	1	0(4)	4	5	4	0(3)	3
20	B 3076	1	0(3)	3	0	3	0(5)	4
22	B 3058	1	5	5	5	5	4	4
23	B 3079	1	2	2	1	1	1	1
27	B 3047	1	3	2	0(4)	3	0(4)	3
29	C 0008	1	0(3)	3	3	3	0(4)	3

Table 1: ROC data for three mammographers for Group 1, i.e. with Enhancement enabled.

Group2	(without Enhancement)							
			Mammographer 1		Mammographer 2		Mammographer 3	
Case #	Database	DB Total # of Lesions	BI RAD	LOC	BI RAD	LOC	BI RAD	LOC
3	A 0015	0	2	2	1	1	1	1
4	A 0034	0	2	2	0(3)	2	0(3)	3
5	A 0112	0	2	1	1	1	0(4)	3
8	A 0020	0	2	2	1	1	2	2
9	A 0003	0	3	2	1	1	1	1
13	A 0030	0	2	2	1	1	0(3)	2
15	A 0009	0	2	2	1	1	2	2
16	A 0037	0	2	2	1	1	1	2
17	A 0099	0	0(3)	2	1	1	2	1
18	A 0116	0	0(3)	3	1	1	1	1
21	A 0035	0	0(3)	2	0(4)	3	0(3)	3
23	A 0018	0	2	2	1	1	1	1
24	A 0022	0	2	2	1	1	0(3)	3
27	A 0005	0	0(3)	2	0(3)	2	1	2
30	A 0016	0	2	2	1	1	1	2
1	B 3003	1	1	2	1	1	5	5
2	B 3389	1	2	2	1	1	1	1
6	B 3009	1	0(4)	4	0(3)	2	0(4)	3
7	C 0309	1	4	4	1	1	0(4)	3
10	C 0142	1	0(3)	3	0(3)	2	1	2
11	B 3016	1	0(4)	4	0(3)	2	4	4
12	B 3382	1	2	2	1	1	3	2
14	B 3134	1	5	4	4	4	5	5
19	B 3005	3	0(3)	3	3	3	0(4)	4
20	C 0127	1	0(3)	3	0(4)	3	0(4)	4
22	C 0015	1	0(4)	4	0(4)	4	5	5
25	B 3007	1	3	3	4	3	4	4
26	B 3012	1	5	5	5	5	0(4)	3
28	B 3380	1	0(4)	4	4	4	0(4)	4
29	C 0358	1	5	5	5	4	0(4)	4

Table 2: ROC data for three mammographers for Group 2, i.e. without enhancement.

To generate an ROC curve from discrete data requires assumptions about the functional form of the curve. The "binormal" model has been widely used in medical imaging. This model includes two adjustable parameters, denoted as "*a*" and "*b*", and assumes that each conventional ROC curve has the same functional form as that implied by two "normal" (i.e., Gaussian) decision variable distributions with generally different means and standard deviations. With the binormal model, a maximum-likelihood parameter estimation scheme is then used to generate an ROC curve that best represents the data. If two different diagnostic systems are to be evaluated, the statistical difference of an apparent difference between measured ROC curves is of interest. Testing differences between ROC curves is well described in the literature[25].

4.5. Results from ROC Analysis and Discussion

In our study, ROC analysis was possible, since the "ground truth" for each case was provided by the mammography database. An initial analysis of the data counted the number of false-positives and true-positives in each group of cases. Before a lesion was considered being diagnosed as malignant or benign, the LOC value was thresholded[10]. The threshold value influences the shape of the ROC curve and its interpretation. For example, if the threshold for the level of confidence was chosen to be 3, meaning that lesions with a LOC greater or equal to 3 were considered as malignant, then the average TPF was found to be 0.667 with enhancement, and TPF = 0.569 without enhancement. This observed increase in sensitivity is encouraging, though it was accompanied by a slight increase in the fraction of false-positives (0.222 compared to 0.178). The latter is not too surprising, *since the applied enhancement protocol only used dyadic spline wavelets* with the non-linear sigmoidal enhancement function, which is certainly not optimal for all types of lesions. We believe that dyadic wavelet expansions are best used to enhance microcalcifications. If the analysis of the data only focused on microcalcifications, then we observed TPF = 0.417 with enhancement compared to TPF = 0.222 without enhancement. No increase or decrease in FPF was noticed! The last finding supports the promise for future research to design specific enhancement protocols for each mammographic feature. Table 3 summarizes initial results of the ROC study using the single basis function described earlier in Section 2.3.

With Enhancement (all Types of Lesions)		Without Enhancement (all Types of Lesions)	
TPF	**FPF**	**TPF**	**FPF**
0.667	0.233	0.569	0.178

Table 3: Results of preliminary ROC study: TPF refers to the fraction of true-positives and FPF to the fraction of false-positives.

A more thorough analysis of the data was undertaken by using the *ROCKIT* software developed by a research group led by Charles Metz at the University of Chicago[26, 27]. This software package was written to analyze data from ROC studies and to generate corresponding ROC curves. More specifically, the purpose of *ROCKIT* is to calculate maximum-likelihood estimates of the parameters of a conventional "binormal" model for the input data, to calculate maximum-likelihood estimates of the parameters of a "bivariate binormal" model for data from two potentially correlated diagnostic tests and, thus, to estimate the binormal ROC curves implied by those data and their correlation; and to calculate the statistical significance of the difference between two ROC curve estimates using any one of three distinct statistical tests (bivariate , area, and TPF test). Moreover, three types of input data are allowed for statistical testing of the differences between ROC curves: unpaired (uncorrelated), fully paired (correlated), and partially-paired test results[27, 26]. *ROCKIT* assumes that the input data follow normal distributions after some unknown monotonic transformation[10]. ROC curves measured in a broad variety of fields demonstrate this "binormal" form[28, 29]. The assumption may be satisfied even when the raw data have multimodal and/or skewed distributions[27, 26].

Using the *ROCKIT* software the analysis was first applied independently to the datasets for Group1 and Group 2 for each of the three radiologists. Unfortunately, this approach did not allow us to compare the diagnostic performance for the two diagnostic systems (softcopy display with and without enhancement), since the data was found to be degenerate. In this case, the result of the ROC analysis would be a straight line with a constant value for TPF, and, therefore the software aborts processing to avoid meaningless output. According to the authors of the software, a degenerate data distribution can be found, if the number of samples is too small or in datasets with many tied values[27].

Since the number of cases could not be increased after conducting the study, and in order to obtain more complete results, we decided to apply the analysis to the union of data from all three radiologists. This was justified by the fact that all three radiologists came from the same population with a similar level of experience. Thus, their performance should be similar under the same conditions, and the data could be treated as independent samples (unpaired data). Nevertheless, we are well aware that the statistical significance of the results must be interpreted carefully. For future ROC studies we plan to increase the number of cases, in order to avoid such a problem.

For the analysis Group 1 (with enhancement) was set as Condition 1 and Group 2 (without enhancement) was considered as Condition 2. The resulting ROC curves for data analyzed as unpaired are shown in Figure 7. Finally, the most important results of ROC analysis, the binormal parameters *a*, *b*, and the area under the ROC curve A_z with their corresponding standard errors, 95% confidence intervals, and correlation of a and b are summarized for unpaired data in Table 4.

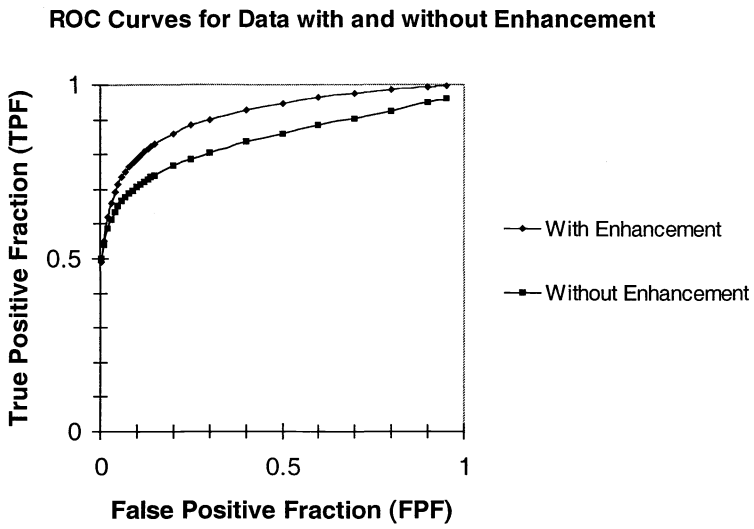


Figure 7: ROC curves for data with Condition 1 (with enhancement) and Condition 2 (without enhancement) analyzed as unpaired data (independent analysis).

Condition 1 (With Enhancement)			Condition 2 (Without Enhancement)		
Binormal Parameter a	**Binormal Parameter b**	**Area under ROC Curve A_z**	**Binormal Parameter a**	**Binormal Parameter b**	**Area under ROC Curve A_z**
1.6183	0.6393	0.9136	1.0813	0.4208	0.8405
Standard Error a	**Standard Error b**	**Standard Error A_z**	**Standard Error a**	**Standard Error b**	**Standard Error A_z**
0.3162	0.2093	0.0325	0.2329	0.1307	0.0475
95% Confidence Interval for a	**95% Confidence Interval for b**	**95% Confidence Interval for A_z**	**95% Confidence Interval for a**	**95% Confidence Interval for b**	**95% Confidence Interval for A_z**
(0.9986, 2.2381)	(0.2291, 1.0495)	(0.8312, 0.9615)	(0.6247, 1.5379)	(0.1647, 0.6770)	(0.7301, 0.9162)
	Correlation(a, b)			**Correlation(a, b)**	
	0.6544			0.4989	

Table 4: Binormal parameters *a*, *b*, area under ROC curve A_z with their corresponding standard errors, 95% confidence intervals, and correlation(*a*, *b*) for Condition 1 (with enhancement) and Condition 2 (without enhancement) analyzed as unpaired data (independent analysis).

As seen from the analysis, the value for the area under the ROC curve A_z was by 8.7% larger for Condition 1 than it was for Condition 2. In all cases the standard error for A_z was between 0.03 and 0.05, which was rather small. Though the 95% confidence intervals for A_z overlapped, there was a clear tendency that diagnostic performance improved *with enhancement* in comparison with diagnosis *without enhancement*. All ROC curves lay high in the unit square of FPF and TPF, which corresponded to accurate diagnostic performances in general, but the curve for Condition 1 was positioned slightly higher (see Figure 7).

The latter observation and the increase of the summary index A_z within statistical errors encourage us to further pursue the application of enhancement protocols for mammographic screening. We are aware of the fact that there always are inherent sources of variability in the index A_z, such as a "case-sample" component due to random variations in the difficulty of the cases included in an ROC experiment, a "between-reader" component due to random variations in the skills of the observers participating in the experiment, and a "within-reader" component associated with each reader's inability to reproduce her/his diagnosis of every case on repeated readings[10]. In addition, we were not able to analyze the data for each radiologist separately due to data degeneracy as mentioned above. The latter has diminished the statistical significance of our results obtained from the analysis of all data combined, since not all samples were completely independent.

5. CONCLUSIONS AND FUTURE WORK

We have reported on the first receiver operating characteristics (ROC) study to evaluate the benefits of contrast enhancement via overcomplete multi-scale expansions of mammograms was successfully completed. The study was carried out in collaboration with radiologists at the Breast Imaging Center in Columbia-Presbyterian Medical Center and the Biomedical Imaging Laboratory of Columbia University.

In continuation of our previous work in digital mammography, an enhancement protocol using a dyadic spline wavelet as the basis for multi-scale expansion and an associated non-linear sigmoidal enhancement function was designed. Suspicious areas (ROIs) of digitized mammograms were decomposed onto a multi-scale basis to obtain coefficients at distinct subbands. Coefficients were modified by applying a non-linear sigmoidal function. Two parameters could be adjusted to change the nature of enhancement. Image reconstruction from modified coefficients occurred in nearly real time through an interactive interface running on a high-resolution digital mammography workstation. To visualize raw data of digitized mammograms at the highest possible contrast and spatial resolutions, 16-Bit BARCO/Metheus framebuffers together with a dual headed high-resolution MegaScan grayscale monitor were utilized in hardware. We incorporated specialized software function calls to directly access the video framebuffer for fast/smooth image display and update.

To quantify the performance of our multi-scale based processing technique in terms of overall sensitivity and specificity, an ROC study was designed and conducted with three radiologists from Columbia-Presbyterian Medical Center specialized in mammography. Conventional ROC curves were generated and significant statistical parameters determined. The area under the ROC curve A_z was used as a summary index to quantify overall specificity and sensitivity of the two diagnostic systems[10]. Unfortunately, it was not possible to analyze datasets for each of three mammographers separately due to data degeneracy. Nevertheless, analyzing all the data together yielded a slight increase (8.7%) in the area A_z for diagnosis with enhancement compared to diagnosis without. Despite the limited statistical significance of this result, it encourages us to further investigate the application of multi-scale methods for contrast enhancement of mammograms. More extensive ROC studies with a larger number of cases are planned to further evaluate the benefits of such processing techniques.

Ancillary to statistical results, we received very positive feedback from the participating radiologists, who expressed great interest in using the interactive display tool and acknowledged a marked improvement in image quality, when enhancement was applied.

The current enhancement protocol works best for the detection/enhancement of microcalcifications. Future directions of work include the expansion of the choice of enhancement protocols to a menu of feature specific enhancement algorithms tailored for each mammographic feature, such as microcalcifications, masses, and spicular lesions, e.g. the application of brushlet functions[30] to mammograms with spicular lesions. In addition, the investigation of a range of optimal enhancement parameters and the optimization of our interface software tool comprise further projects. Our "dream" is to present a clinical interface, where specific enhancement protocols can be selected by a physician by only "pushing a button on the screen". We envision that through such a clinical interface the diagnostic performance of radiologists in screening digital mammograms could be substantially improved, both in terms of cost and quality.

6. ACKNOWLEDGEMENT

This work was supported by the Breast Cancer Research Program of the Department of Defense U.S. Army Medical Research and Material Command, Award Number DAMD17-93-J-3003 and the Whitaker Foundation.

7. REFERENCES

[1] I. Koren and A. Laine, "A discrete dyadic wavelet transform for multidimensional feature analysis," in *Time Frequency and Wavelets in Biomedical Signal Processing*, *IEEE Press series in biomedical engineering*, M. Akay, Ed. Piscataway, NJ: IEEE Press, 1998, pp. 425-448.

[2] A. F. Laine, S. Schuler, J. Fan, and W. Huda, "Mammographic feature enhancement by multiscale analysis," *IEEE Transactions on Medical Imaging*, vol. 13, pp. 725-740, 1994.

[3] D. Brzakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammograms," *IEEE Transactions on Medical Imaging*, vol. 9, pp. 233-241, 1991.

[4] H. Yoshida, W. Zhang, W. Cai, K. Doi, R. M.Nishikawa, and M. L. Giger, "Optimizing wavelet transform based on supervised learning for detection of microcalcifications in digital mammograms," Proceedings of the IEEE International Conference on Image Processing, IEEE Comput. Soc. Press, Washington, D.C., pp. 152-155, 1995.

[5] R. N. Strickland and H. I. Hahn, "Wavelet transforms for detecting microcalcifications in mammograms," *IEEE Transactions on Medical Imaging*, vol. 15, pp. 218-229, 1996.

[6] W. Qian, L. P. Clarke, M. Kallergi, and H.-D. Li, "Tree-structured nonlinear filter and wavelet transform for microcalcification segmentation in mammography," Biomedical Image Processing and Biomedical Visualization, Proceedings of the SPIE, vol. 1905, San Jose, CA, pp. 520-590, 1995.

[7] D. Chen, C.-M. Chang, and A. Laine, "Detection and enhancement of small masses via precision multiscale analysis," Computer Vision - ACCV '98, Proceedings of the Third Asian Conference on Computer Vision, vol. 1, Springer-Verlag, Hong Kong, PRC, pp. 192-199, 1998.

[8] N. Karssemeijer and G. M. t. Brake, "Detection of stellate distortions in mammograms," *IEEE Transactions on Medical Imaging*, vol. 15, pp. 611-619, 1996.

[9] A. F. Laine, J. Fan, and W. Yang, "Wavelets for contrast enhancement of digital mammography," *IEEE Engineering in Medicine and Biology Society Magazine*, vol. 14, pp. 536-550, 1995.

[10] C. E. Metz, "ROC methodology in radiologic imaging," *Investigative Radiology*, vol. 21, pp. 720-733, 1986.

[11] B. J. Betts, J. Li, A. Aiyer, S. M. Perlmutter, P. C. Cosman, R. M. Gray, R. A. Olshen, and al., "Image Quality," Stanford University, Stanford, Final Report Final Report For U.S. Army Medical Research and Material Command, Fort Detrick, Maryland, November 18 1998.

[12] H. Nab, N. Karssemeijer, L. Van Erning, and J. Hendriks, "Comparison of digital and conventional mammography: a ROC study of 270 mammograms," *Medical Informatics*, vol. 17, pp. 125-131, 1992.

[13] W. Qian, L. Li, and L. P. Clarke, "Image feature extraction for mass detection in digital mammography: Influence of wavelet analysis," *Medical Physics*, vol. 26, pp. 402-408, 1999.

[14] H.-P. Chan, B. Sahiner, R. Wagner, and N. Petrick, "Effects of sample size on classifier design for computer-aided diagnosis," Medical Imaging 1998: Image Processing, vol. 3338, pt.1-2, SPIE- the International Society for Optical Engineering, San Diego, CA, USA, pp. 845-858, 1998.

[15] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proceedings of the IEEE*, vol. 84, pp. 626-638, 1996.

[16] M. Holschneider and R. Kronland-Martinet, "A real-time algorithm for signal analysis with the help of the wavelet transform," presented at Wavelets: Time-frequency Methods and Phase Space, Springer Verlag, Berlin, Germany, pp. 286-304, 1990.

[17] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory*, vol. 38, pp. 587-607, 1992.

[18] C.-M. Chang and A. F. Laine, "Enhancement of mammograms from oriented information," Proceedings of the IEEE International Conference on Image Processing, vol. 3, IEEE Comput. Soc., Santa Barbara, CA, pp. 524-527, 1997.

[19] P. G. Tahoces, J. Correa, M. Souto, and C. G. and, "Enhancement of chest and breast radiographs by automatic spatial filtering," *IEEE Transactions on Medical Imaging*, vol. 10, pp. 330-335, 1991.

[20] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Siam, 1992.

[21] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 710-732, 1992.

[22] A. F. Laine, J. Fan, and S. Schuler, "A framework for contrast enhancement by dyadic wavelet analysis," in *Digital Mammography*, S. M. A. A. G. Gale, D. R. Dance, and A.Y. Cairns, Ed. Amsterdam, The Netherlands: Elsevier, 1994, pp. 91-100.

[23] W. B. Richardson Jr., "Nonlinear filtering and multiscale texture discrimination for mammograms," Mathematical Methods in Medical Imaging, D.C. Wilson and J.N. Wilson, Eds., vol. 1768, San Diego, CA, pp. 293-305, 1992.

[24] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425-455, 1994.

[25] B. J. McNeil and J. A. Hanley, "Statistical approaches to the analysis of receiver operating characteristic (ROC) curves," *Medical Decision Making*, vol. 4, pp. 137-150, 1984.

[26] C. E. Metz, "ROCKIT 0.9B," Beta Version ed. Chicago, IL: Department of Radiology, University of Chicago, 1998.

[27] C. E. Metz, B. A. Herman, and C. A. Roe, "Statistical comparison of two ROC curve estimates obtained from partially-paired datasets," *Medical Decision Making*, vol. 18, pp. 110-121, 1998.

[28] J. A. Hanley, "The robustness of the "binormal" assumptions used in fitting ROC curves," *Medical Decision Making*, vol. 8, pp. 197-203, 1988.

[29] K. O. Hajian-Tilaki, J. A. Hanley, L. Joseph, and J.-P. Collet, "A comparison of parametric and nonparametric approaches to ROC analysis of quantitative diagnostic tests," *Medical Decision Making*, vol. 17, pp. 94-107, 1997.

[30] F. Meyer and R. R. Coifman, "Brushlets: A tool for directional image analysis and image compression," *Applied and computational harmonic analysis*, vol. 4, pp. 147-187, 1997.