Comparative Performance of Multi-view Stereoscopic and Mammographic Display Modalities for Breast Lesion Detection

by

Lincoln Jon Webb

Medical Physics Graduate Program
Duke University

Date:_____

Approved:

_____
Ehsan Samei, Supervisor

_____
Joseph Lo

_____
Sua Yoo

Thesis submitted in partial fulfillment of
the requirements for the degree of Master's of Science in the
Medical Physics Graduate Program
of Duke University

2010

ABSTRACT

Comparative Performance of Multi-view Stereoscopic and Mammographic Display Modalities for Breast Lesion Detection

by

Lincoln Jon Webb

Medical Physics Graduate Program
Duke University

Date:_____

Approved:

_____
Ehsan Samei, Supervisor

_____
Joseph Lo

_____
Sua Yoo

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master's of Science in the
Medical Physics Graduate Program
of Duke University

2010

Abstract

Rationale and Objectives: Mammography is known to be one of the most difficult radiographic exams to interpret. Mammography has important limitations including superposition of normal tissue that can obscure a mass, chance alignment of normal tissue to mimic a true lesion, and the inability to derive volumetric information. It has been shown that stereomammography can overcome these deficiencies by showing that layers of normal tissue lay at different depths. If standard stereomammography (i.e. a single stereoscopic pair consisting of 2 projection images) can significantly improve lesion detection, how will multi-view stereoscopy (MVS), where many projection images are used, compare to mammography? The aim of this study was to assess the relative performance of MVS compared to mammography for breast mass detection.

Materials and Methods: The MVS image sets consisted of the 25 raw projection images acquired over an arc of approximately 45° using a Siemens prototype breast tomosynthesis system. The mammograms were acquired using a commercial Siemens FFDM system. The raw data was taken from both of these systems for 27 cases and realistic simulated mass lesions were added to duplicates of the 27 images at the same local contrast. The images with lesions (27 mammography and 27 MVS) and the images without lesions (27 mammography and 27 MVS) were then post-processed to provide comparable and representative image appearance across the two modalities. All 108

image sets were shown to five full-time breast imaging radiologists in random order on a state-of-the-art stereoscopic display. The observers were asked to give a confidence rating for each image (0 for lesion definitely not present, 100 for lesion definitely present). The ratings were then compiled and processed using ROC and variance analysis.

Results: The mean AUC for the five observers was 0.614±0.055 for mammography and 0.778±0.052 for multi-view stereoscopy. The difference of 0.164±0.065 was statistically significant with a p-value of 0.0148.

Conclusion: The differences in the AUCs and the p-value suggest that multi-view stereoscopy has a statistically significant advantage over mammography in the detection of simulated breast masses. This highlights the dominance of anatomical noise compared to quantum noise for breast mass detection. It also shows that significant lesion detection can be achieved with MVS without any of the artifacts associated with tomosynthesis.

Contents

List of Tables

List of Figures

Acknowledgements

This study was conducted with partial financial support from a grant from the Prevent Cancer Foundation. I would also like to thank my advisor, Ehsan Samei, for all his help and support and all the others who have helped me along the way: Joseph Lo, Sua Yoo, Jay Baker, Ruth Walsh, Connie Kim, Mary Scott Soo, Karen Johnson, and Sujata Ghate.

1. Introduction

Mammography is currently the clinical standard for breast cancer screening in non-high risk patients. It is also regarded as one of the most difficult radiographic exams to interpret. Two-dimensional projection radiographic images are hindered by the inability to ascertain volumetric information, such as the three-dimensional shape and the exact location of a lesion. This limitation affects mammography even more, which reflects different nonuniform layers of glandular and fatty tissues, combined with the nonuniformity of breast size and composition across subjects. In mammography, a true lesion can be obscured by the superposition of the overlying or underlying tissue, resulting in a false negative. Superposition of normal tissues can also lead to false positives, where the chance alignment of tissue can mimic a true lesion[1].

Stereomammography (consisting of two images taken 6-10 degrees apart) has been shown to largely overcome these limitations and even help readers detect new lesions that are not visible in standard mammograms[2]. More recently, a large clinical trial at Emory University has shown that stereomammography has significantly reduced false positive lesion detections by 46% and significantly increased true positive lesion detections by 23% over standard mammography[1]. If a single stereoscopic image pair significantly improves lesion detection, what kind of improvement can be made if multiple projections are taken and paired together, such as in multi-view stereoscopy

(MVS)? MVS is different from stereomammography in that instead of a single stereoscopic image pair, multiple pairs of stereograms are employed.

The purpose of this study was to investigate how MVS compares to standard mammography in the detection of breast masses. Our MVS image sets were acquired with a prototype breast tomosynthesis system. This system is much like a standard full field digital mammography (FFDM) system, except that the x-ray tube rotates about an axis 6 cm above the detector surface over an arc of about 45° and acquires 25 projection images. Instead of performing a tomosynthesis reconstruction on the raw projections, the images were processed to form a set of 25 images that appeared like noisy mammograms. Of the 25 projections, the first was paired with the fourth, the second with the fifth and so on, resulting in a set of 22 stereo pairs. The corresponding raw mammogram data, acquired on a digital mammography unit, were obtained and processed in the same way as the MVS images. The raw images were then duplicated, simulated lesions[3] were added to the duplicates, and all the images were processed using the same algorithm. The images were then viewed monoscopically and stereoscopically by five expert breast imaging radiologists using a stereoscopic display. Confidence ratings to assess the relativistic merit of MVS were then compiled and processed using receiver operating characteristic (ROC) analysis.

2. Materials and Methods

2.1 Image Acquisition

Using an IRB-approved protocol, Duke University's RAI Labs has acquired breast tomosynthesis images of over 300 subjects using a prototype breast tomosynthesis system (Siemens Mammomat *Novation*^{TOMO}, Siemens Medical Solutions USA, Inc., Malvern, PA). The subjects are also imaged with a commercial digital mammography system (Siemens Mammomat *Novation*^{DR}, Siemens Medical Solutions USA, Inc., Malvern, PA). Our study employed a subset of data from this database. The data have a pixel pitch of 0.070 mm for the mammographic images and 0.085 mm for the tomosynthesis projection images, and the tomosynthesis images were acquired with 1×2 pixel binning (the larger acquisition pixel size is along the anode-cathode axis). To ensure that no actual cancers were present in the images, the tomosynthesis database was scoured for normal cases using subjects' mammography history, including any diagnoses and biopsy results. Once suitable tomosynthesis cases were identified, the mammogram database was searched to obtain as many of the corresponding raw mammograms as possible. In the end, 27 suitable raw mammograms were obtained and their corresponding 27 tomosynthesis images were extracted from the database, all of which were either right or left medial lateral oblique (MLO) views. All 54 of these raw image sets were then duplicated in preparation for lesion simulation.

As mentioned previously, although the MVS image sets were acquired in 1×2 binning mode, they are noisy as compared to the mammograms because the total dose for a tomosynthesis exam (i.e. the total dose from all 25 projections) is divided amongst the projections. To account for this difference in dose between modalities, we assumed dose to be a function of the tube current-time product (mAs) and the compressed breast thickness since dose increases with increasing mAs and breast thickness. The mean compressed breast thickness and mAs for each modality are given in Table 1 along with their standard deviations. The average tomosynthesis breast thickness was 0.85 mm thicker than the average mammography breast thickness. This difference in thickness was considered to have a negligible effect on dose[4]. As for mAs, the total mean mAs for the MVS image sets was 31% higher than the mean for the mammograms, which means that each projection of an MVS set was acquired with about 5% of the dose used to acquire a mammogram, corresponding to a 95% reduction in dose per projection. In summary, although MVS has less dose per projection, the total dose for the entire image set is 31% higher than that for the mammogram. Both mathematical observer models[5,6] and actual observer studies[7] have shown that even a 50% reduction in dose has a minimal effect on the detection of malignant masses in mammographic studies. We therefore assumed that the difference in dose between the two modalities would have a negligible effect on the detection of the simulated malignant masses.

Table 1: Mean thickness and mAs for both modalities with their standard deviations.

	thickness mean (mm)	thickness std (mm)	mAs mean	mAs std
mammography	46.89	13.19	116.03	47.46
multi-view stereo	47.74	13.35	151.89	51.59

2.2 Image Processing

2.2.1 Lesion Simulation

Simulated, realistic, malignant masses were added to the 54 duplicate raw image sets using the malignant masses developed by Saunders et al[3]. The masses of different shapes ranged in size from 5.3-15.4 mm along the largest dimension. The lesions were all added at a certain local contrast level using the following logic. We know that local contrast C can be given in terms of maximum signal (or lesion) exposure E_s and background exposure E_b via

$$C = \frac{E_s - E_b}{E_b}$$

Eq. (1)

which can be rewritten as

$$\frac{E_s}{E_b} = C + 1.$$

Eq. (2)

We assume that the pixel value P of the detector has a linear relationship with exposure E as

$$P = \alpha E + \beta,$$

Eq. (3)

where α is a constant that describes the slope of the P vs. E relationship, and β is the vertical offset, or y intercept, of that line. With the β values (from the experiment below), the natural log of Eq. (2) and substituting in Eq. (3) for exposure E leads to

$$\ln(P_s - \beta) - \ln(P_b - \beta) = \ln(C + 1), \quad \text{Eq. (4)}$$

which is the equation used to determine what the maximum pixel value of a simulated lesion (P_s) should be for a given contrast C in a given image with pixel value P_b. All masses were added at a local contrast of C = 9.76-10.4%, which was determined by experimentation to give a good level of mass conspicuity in both image types.

The above formulation is convenient as the contrast of the added lesions is independent of the image pixel value. However, it relies on a linear relationship between pixel values and exposure. To verify this relationship, we measured the system response in the range of kVp and breast thickness applicable to our data set. The phantom used consisted of a varying number of 20mm slabs of standard 50% glandular/50% fat acrylic. The mean background pixel value was plotted against the measured exposure (in mR). All the lines were linear with $R^2 > 0.998$. Table 2 tabulates the different kVp values and phantom thicknesses used with their corresponding α and β values.

For lesion simulation, both the image of the simulated lesion mask and the image of the breast into which the lesion was to be inserted were logged. The proper value for

Table 2: α and β values for different kVp and thickness.

kVp	thickness (mm)	α (pixel value/mR)	β (pixel value)
28	20	109.8	6.37
28	40	120.28	5.2
28	60	131.57	0.83
30	40	128.84	2.74
30	60	140.43	2.05
30	80	145.05	3.15

the maximum added pixel value, corresponding to $\ln(P_s - \beta) - \ln(P_b - \beta)$, was determined using Eq. (4) and the pixel values of the logged image of the lesion were scaled accordingly. A location for the centroid of the lesion was selected, and the scaled and logged image of the lesion was added to the logged image of the breast. The resulting image was then exponentiated and visually inspected. In the case of the mammograms, the pixel size was small (i.e. 70 μm) compared to the MVS projections (i.e. 85 μm). As such, the size of the added lesion was adjusted accordingly to create the same physical size for the simulated lesion in both modalities.

For the MVS image sets, there was an additional complexity: realistic lesion movement had to be characterized so the lesion could be added at the correct location in each of the 25 projections. By using the geometry of the tomosynthesis system, it was approximated that the lesion centroid location should be moved X pixels away from the lesion centroid location selected in the center projection, which corresponds to the same location selected in the mammogram, as:

$$X = \frac{d}{s}\tan(1.8n)$$

Eq. (5)

where n = -12,-11,…,11,12 (0 corresponds to the center projection), s is the pixel size (in mm), and d is the depth at which the lesion is to be inserted in the volume (also in mm). With the corresponding shift with lesion location from image to image, the lesions were added randomly within the breast volume at approximately mid-plane of each breast. The constants in Eq. (5) come from the geometry of the system

2.2.2 Histogram Analysis and Segmentation

Once the lesions were inserted into the 54 duplicated raw image sets, all 108 raw image sets needed to be processed before they could be presented to a radiologist. The lesions had to be added to raw data because we could not ensure that the simulated lesions would have had the same original contrast had they been inserted into the post-processed mammograms. Also, there is no post-processing algorithm on the tomosynthesis unit for the MVS image sets. Therefore, it was necessary to develop a post-processing algorithm because the algorithms used by FFDM units are proprietary and in this study we needed to provide a consistent presentation of images independent of the modality.

The images were first inverted and the histogram for each image was analyzed. The portion of the histogram that represented anatomy was the only part used; this

portion was then spread linearly over all pixel values (0-65535 for uint16 images). For the MVS sets, all of the non-anatomic structures in each projection that make it difficult to process a series of images, such as the compression paddle, were cropped out. Care was taken to ensure that the image of the breast in each of the projections was the same size so that histogram equalization would have the same effect on each projection.

2.2.3 Histogram Equalization

Before equalization processing could take place, the bad pixels in the image had to be removed as they could have an adverse effect on the equalization process. Bad pixels had values 300 more than any other and were always at the very bottom of the image against the chest wall, and there was almost never more than one in an image. These values were removed and set equal to the value ten pixels above it.

The histogram equalization algorithm used was a contrast-limited adaptive histogram equalization[8] (CLAHE) algorithm implemented in Matlab (version 7.7.0, The MathWorks, Inc., Natick, MA). In CLAHE processing, each image is divided into smaller regions called tiles, and each tile's contrast is enhanced so that the histogram of each tile matches a given distribution. The tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast in homogeneous areas can be limited to avoid noise amplification by clipping the histogram of that tile. There were a number of inputs required for this function, which were all determined by

experimentation: the number of tiles was set to [5×5], the number of bins was 2^{16} (since it's a 16-bit image), the distribution was 'exponential,' the distribution shape parameter alpha was set to 6, and the histogram clip limit varied from 0.0001 and 0.003 depending on the image.

2.2.4 Smoothing and Sigmoid LUT

Because the resulting images were exceptionally noisy, especially the MVS image sets, a convolution with a small 3×3 smoothing kernel was performed on each image, including the mammograms. The center value for the kernel was 0.2, and the others on the periphery were 0.1. This kernel greatly diminished the noise present in the images and provided a slight improvement in lesion visualization across both modalities.

After smoothing, a sigmoid look-up table was applied to the images to optimize the grayscale of the image for the stereoscopic display. The domain of the sigmoid was 0-65535 and the range was 0-3800. The value 3800 was chosen because it was near the maximum pixel value the monitors could display. To determine the sigmoid curve for a given image, nonlinear regression was used to fit a function of the form

$$y = y_o + \frac{a}{\left(1 + e^{-(x-x_0)/b}\right)} \qquad \text{Eq. (6)}$$

to five points: two fixed endpoints (0, 0) and (65535, 3800), a center point (x_o, $65535/2$), a point to the left of center (x_1, 0.15×3800), and a point to the right of center (x_2,

0.85×65535). Eq. (6) is the standard equation for a sigmoid curve, where $a = 3800$ describes the height of the curve, $y_o = 0$ describes its vertical offset, $b = 500$ describes the general shape of the curve, and x_o is the center of the sigmoid. For mammographic images, x_o was set to be the location of the image histogram peak on the x-axis. For most MVS image sets, x_o for each projection was determined to be the x-axis location of that projection's histogram peak. There were some MVS sets where this formalism yielded images of poor quality, so a single x_o value for the entire image set was determined by experimentation. The slope of the linear portion of the sigmoid curve was generally in the range of 0.3 to 0.6 and x_1 and x_2 were determined accordingly. Two examples of the resulting mammograms and their corresponding MVS set are shown in Figures 1 and 2. For the MVS sets, only two stereo pairs are shown: the center projection and the two pairing images to either side of it.

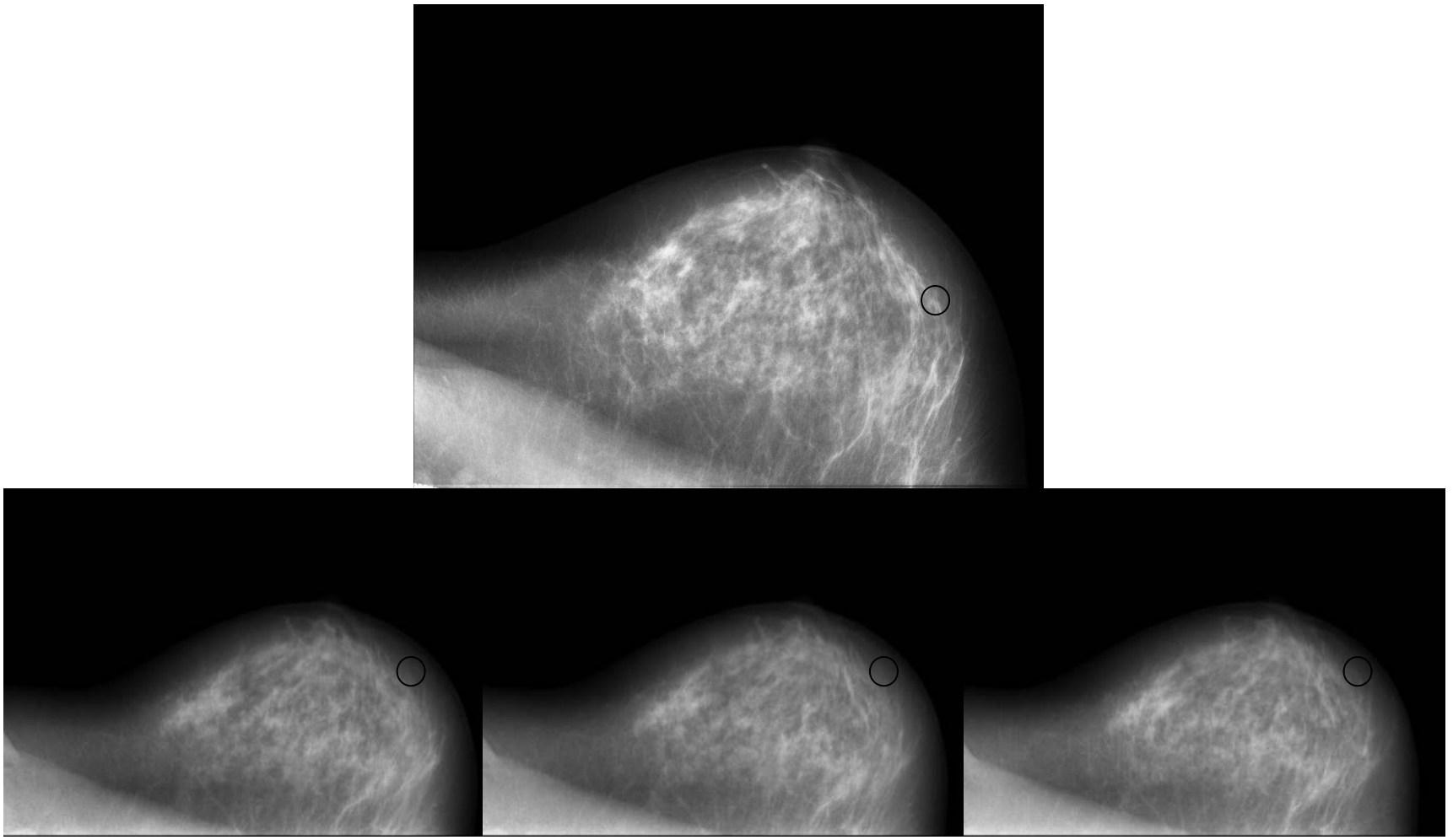

Figure 1: Two examples of mammographic images (top) and three projections of their corresponding MVS set (bottom three). Lesions are shown with circles.

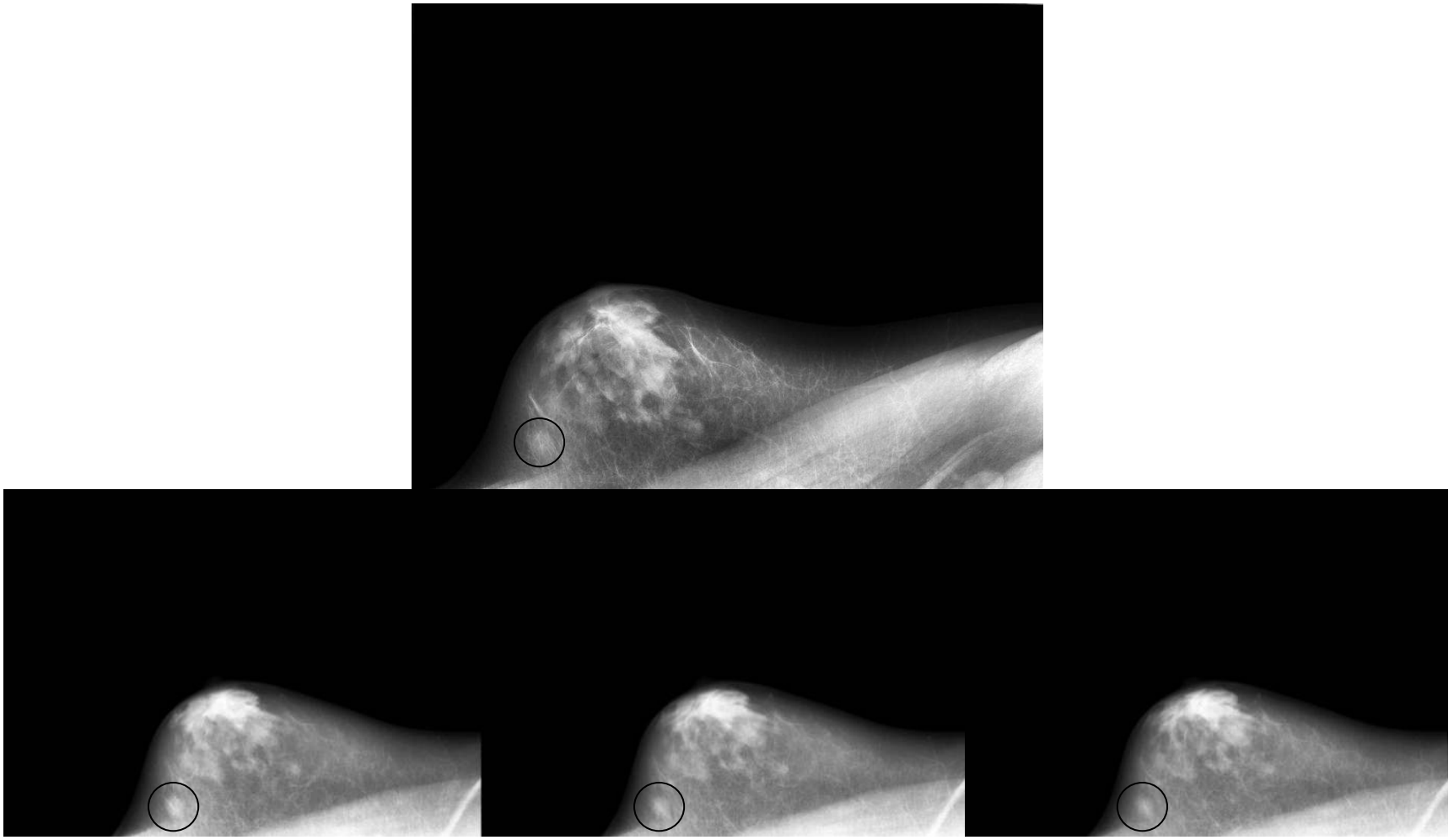

Figure 2: Two more examples of mammographic images (top) and three projections of their corresponding MVS set (bottom three). Lesions are shown with circles.

2.3 Graphical User Interface

In all, there were 108 image sets to display from 27 patients. There were 27 negative mammograms, 27 positive mammograms, 27 negative MVS sets, and 27 positive MVS sets. Each mammogram was a DICOM image about 16MB in size and each MVS image set (which consisted of 25 DICOM images) was 481MB. Images this large are hard to manipulate and often take up most of the RAM on the workstation, which slows the system down. Therefore, each MVS set was converted into a single three-dimensional MAT file, which reduced the file size to about 100MB. All images were then transferred to the stereoscopic display workstation to minimize memory issues as well as to minimize loading and image rendering time on the stereo display. The stereoscopic display, shown in Figures 3 and 4, is a state-of-the-art device consisting of two five-

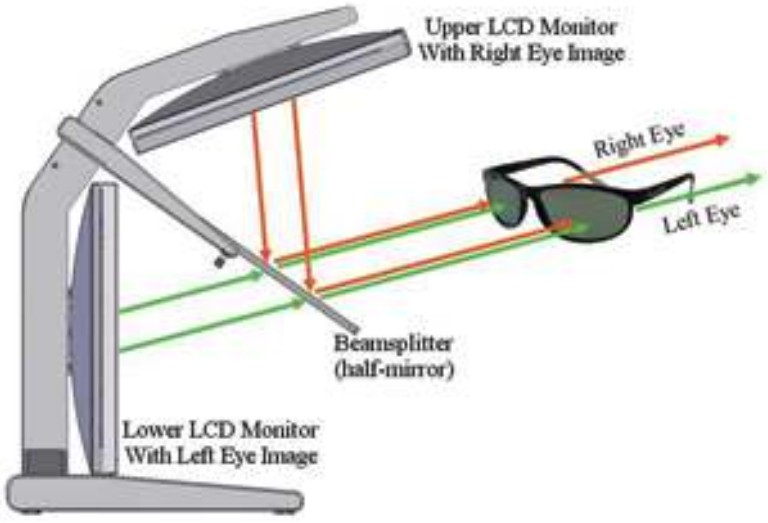

Figure 3: A schematic showing how the stereoscopic image is formed[9].

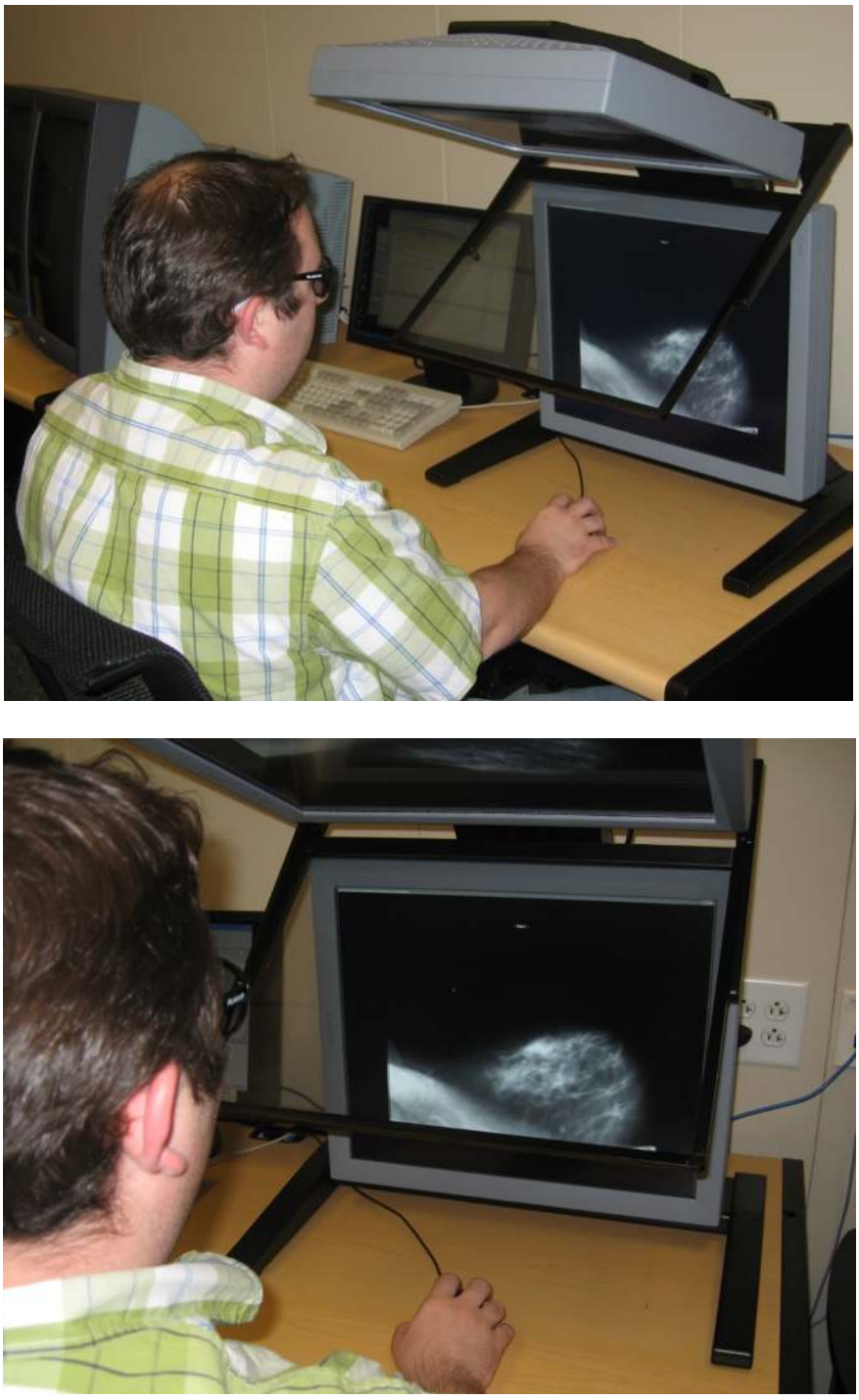

Figure 4: The stereoscopic display, note the blurry superimposed image in the bottom frame.

megapixel medical displays (Planar Dome C5iGRAY, Planar Systems, Inc., Beaverton, OR) separated by about 110° with a semi-transparent mirror bisecting the angle. As is illustrated by the schematic in Figure 3, the image from the upright display is transmitted through the mirror and the image from the display mounted above is reflected off the mirror. The two monitors are cross-polarized, so user must wear special glasses with cross-polarized lenses. This enables each eye to see an image from only one of the monitors. The 'cyclopean eye' of the mind will then fuse these two images and give a sense of depth to the user[10].

A graphical user interface (GUI) was developed in Matlab which displayed stereo pairs on the stereoscopic display and allowed the user to scroll back and forth through the set of projection stereo pairs within a set and to move from one set to the next. The GUI also allowed the user to change the window width and window level across an image set. Display size for the images for the two modalities was adjusted considering the difference in pixel size to achieve the same physical display size on the display monitor. The set of 25 projections was acquired over an arc of approximately 45°, which corresponds to a separation of about 1.8° between each consecutive projection image. The ideal angle of separation between a stereoscopic pair is about 6°. Although the perceived depth of the object of interest increases with greater angular separation between the two images, most people find it uncomfortable to view stereo pairs that are separated by more than 8-9° of arc[10]. Since we were limited by this increment of 1.8°

between images, we decided to pair images that were about 5.4° apart (i.e. pair and display images 1 and 4, 2 and 5 … 22 and 25) since it was the closest increment to the ideal angle of separation, which meant that each MVS image set consisted of 22 stereo pairs. All images were rotated such that the chest wall was at the bottom and the nipple was at the top. This was done because of the acquisition geometry in the tomosynthesis unit. If the images were kept in the standard vertical orientation, there would be no horizontal parallax necessary for viewing a stereoscopic image[10].

This same GUI was used to display the mammograms. The same mammogram image was displayed on both monitors of the stereoscopic display to keep the brightness of the mammogram the same as that of the MVS stereo pairs. Additionally, we wanted to keep viewing conditions as similar as possible across the two image types. If the mammogram were to be viewed on a standard medical display we would have to account for the reduction in luminance due to the polarized lenses as well as the difference in brightness between the two display modalities.

2.4 Observer Study and ROC Analysis

Five expert breast imaging radiologists were each shown all 108 image sets. First, each observer was shown a multi-view stereo set and asked to get familiar with it and try to get a sense of depth in the image, and then the same set was shown again but with the images on the monitors swapped, thus swapping the image that each eye sees for a

given stereo pair and inverting the observer's sense of depth. The observers then chose the orientation they preferred and all of the sets in the study were shown in the same orientation. Additionally, they were briefly introduced to the stereo display, shown how it works, and were shown a series of six training image sets. In the training the observer was shown a positive mammogram and then the corresponding positive MVS image set. The simulated lesion was pointed out to the observer in each image set so he or she could get a sense of how a lesion looks in the multi-view stereoscopy modality.

After training, the observers were shown the 108 image sets and asked to detect masses by giving a confidence score for each image ranging from 0 to 100, where 0 confidence means the observer is 100% confident that a lesion is absent (definitely negative) and 100 confidence means that the observer is 100% confident a lesion is present (definitely positive). As discussed in the beginning of Section 2.3, there are four image sets per case (one positive and one negative mammogram, and one positive and one negative MVS set), all of which were shown to each observer in a random order. Observers were able to view all the images and give their confidence scores in a single two-hour session. The confidence scores were then compiled and processed using the DBM MRMC 2.1 software package[11-19] that employs PROPROC curve fitting methodology to determine the receiver operating characteristic (ROC) curve and the area under the curve (AUC) for each observer-modality combination. The ROC results

were compared across modalities and across observers, and they were compared for statistical significance of differences.

ROC curves plot the true positive fraction (TPF) versus the false positive fraction (FPF). The TPF is equal to sensitivity, which is defined as TP/(TP+FN). Sensitivity measures the proportion of actual positives that are correctly identified. If we have a sensitivity of 100% (or 1 on the y-axis of the ROC curve), then the test recognizes all of the actual positives, or in our case, it would correctly identify all of the images with a mass. The FPF is equal to 1-specificity, where the specificity is defined as TN/(TN+FP). Specificity measures the proportion of negatives that are correctly identified. A specificity of 100% (or 0 on the x-axis of the ROC curve) means that the test recognizes all actual negatives, or in our case, it would correctly identify all of the images where a lesion is not present. Using these definitions, we see that a perfect test would yield a point in the upper left corner of the ROC plot, which represents 100% sensitivity (no false negatives) and 100% specificity (no false positives). In contrast, if the observer were randomly assigning a confidence score or randomly guessing whether or not a given image is positive or negative, this would yield a point along a diagonal line from the bottom left to the top right corners. This diagonal line separates the ROC space: points above represent good classification and points below represent poor classification. One metric that can be derived from the ROC is the AUC, or the area under the ROC curve. An AUC of 0.5 corresponds to random classification (as the area under the diagonal line

would be 0.5), and the higher the AUC the better the test performs, where an AUC of 1.0 is perfect detection. The AUC is corresponds to the accuracy of the test, or the percent correct, (TP+TN)/(TP+TN+FP+FN).

Take, for example, a point at (0.9, 1.0), such as in Figure 6 where we have a high TPF and FPF. At this point, we have 100% sensitivity, meaning the test is correctly identifying positive cases. However, we also have a specificity of ~10%, meaning that the test is not correctly identifying actually negative cases, yielding a high number of false positives. This is a good operating point for a test like mammography because this trade-off is something that is recognized: the radiologist wants to catch all cancers, even at the expense of having some false positives. We would rather err on the side of caution. We could also look at a point such as (0,0.2). Here we have a sensitivity of 20%, meaning that the test is only correctly identifying 20% of the positive cases, but we also have a specificity very close to 100%, meaning the test is correctly identifying nearly all negative cases.

3. Results

The ROC results are shown in Figures 5 and 6 for mammography and multi-view stereoscopy, respectively. The AUC for each individual curve is also listed in the figures. The mean AUC for the five observers was 0.614±0.055 for mammography and 0.778±0.052 for multi-view stereoscopy. The difference between the two mean AUCs was 0.164±0.065. The ROC difference was calculated to have a p-value of 0.0148 (or 0.0072 if using random readers, random cases analysis) and a 95% confidence interval of (-0.295,-0.034).

Table 3 shows the results in more depth. The table includes the AUC values shown in Figures 5 and 6 as well as the difference between the two for each observer. All the columns in Table 3 that represent a difference were taken by subtracting the MVS value from the value for mammography. The fifth column shows the change in FPF at the 95% sensitivity line between the two modalities for each observer, which is valuable in characterizing how MVS performs in the region of the curve at which breast radiographers operate. The last column is a similar metric, but shows the difference in the partial AUC values. The partial AUCs were calculated for each observer above the 95% sensitivity line. The results indicate more variability across observers for the stereo modality, possibly due to the novelty of the new technique. However, for each observer, there was a statistically significant improvement for mass detection for multi-view stereoscopy compared to mammography.

Table 3: AUC values for the two modalities, the difference between them, the change in FPF between modalities at 95% sensitivity, and the difference in partial AUC (>95% sensitivity) between modalities.

	Mammo AUC	MVS AUC	ΔAUC	ΔFPF at 95% sens.	ΔpAUC (>95% sens.)
Observer 1	0.627	0.714	-0.087	0.267	-0.0114
Observer 2	0.586	0.788	-0.201†	0.023	0.0003
Observer 3	0.599	0.808	-0.209†	0.197	-0.0065
Observer 4	0.616	0.782	-0.166	0.106	-0.0026
Observer 5	0.640	0.798	-0.158	0.187	-0.0055
All	0.614	0.778	-0.164†	0.156	-0.0051

† indicates $p < 0.05$

The lesion simulation and image processing algorithms discussed in Section 2.2 are also outcomes of this research. One feature of this algorithm is that it processes a series of images, such as the MVS image sets, using the same parameters, thus ensuring a constant image presentation across the projections. Figure 7 shows examples of images that were processed using the algorithm presented in this study as compared to the proprietary Siemens algorithm, available commercially.

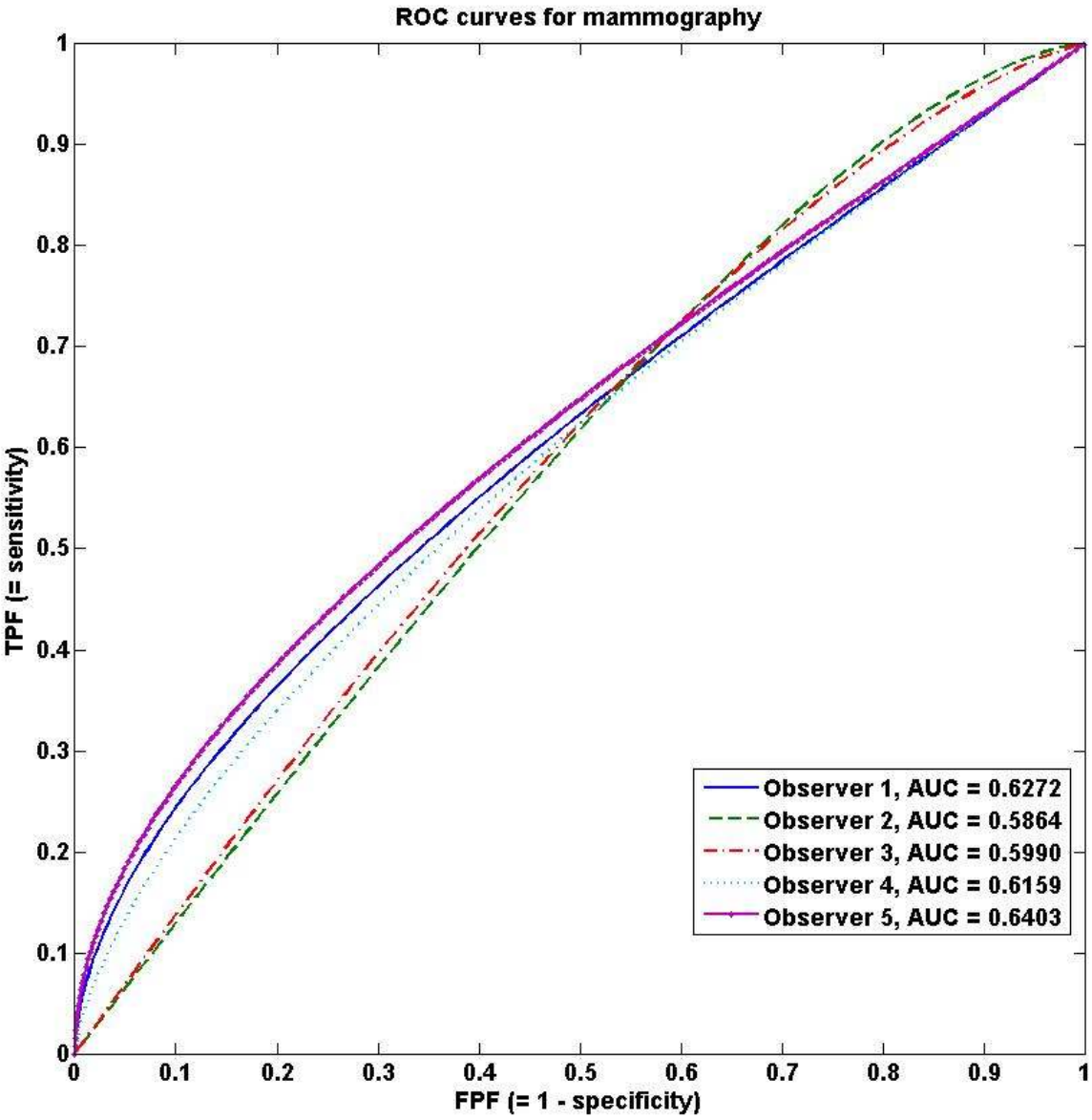

Figure 5: ROC curves for mammography.

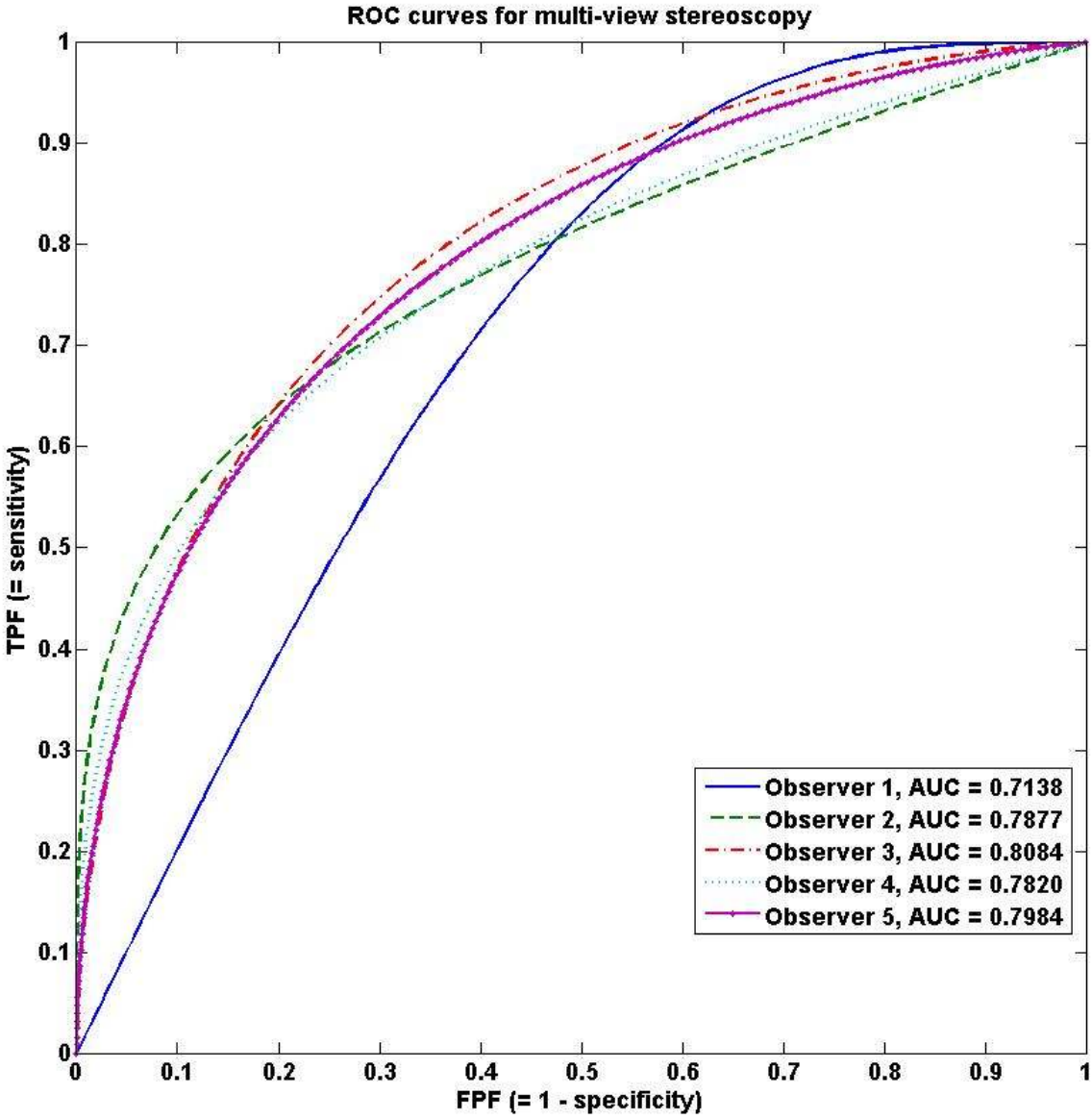

Figure 6: ROC curves for multi-view stereoscopy.

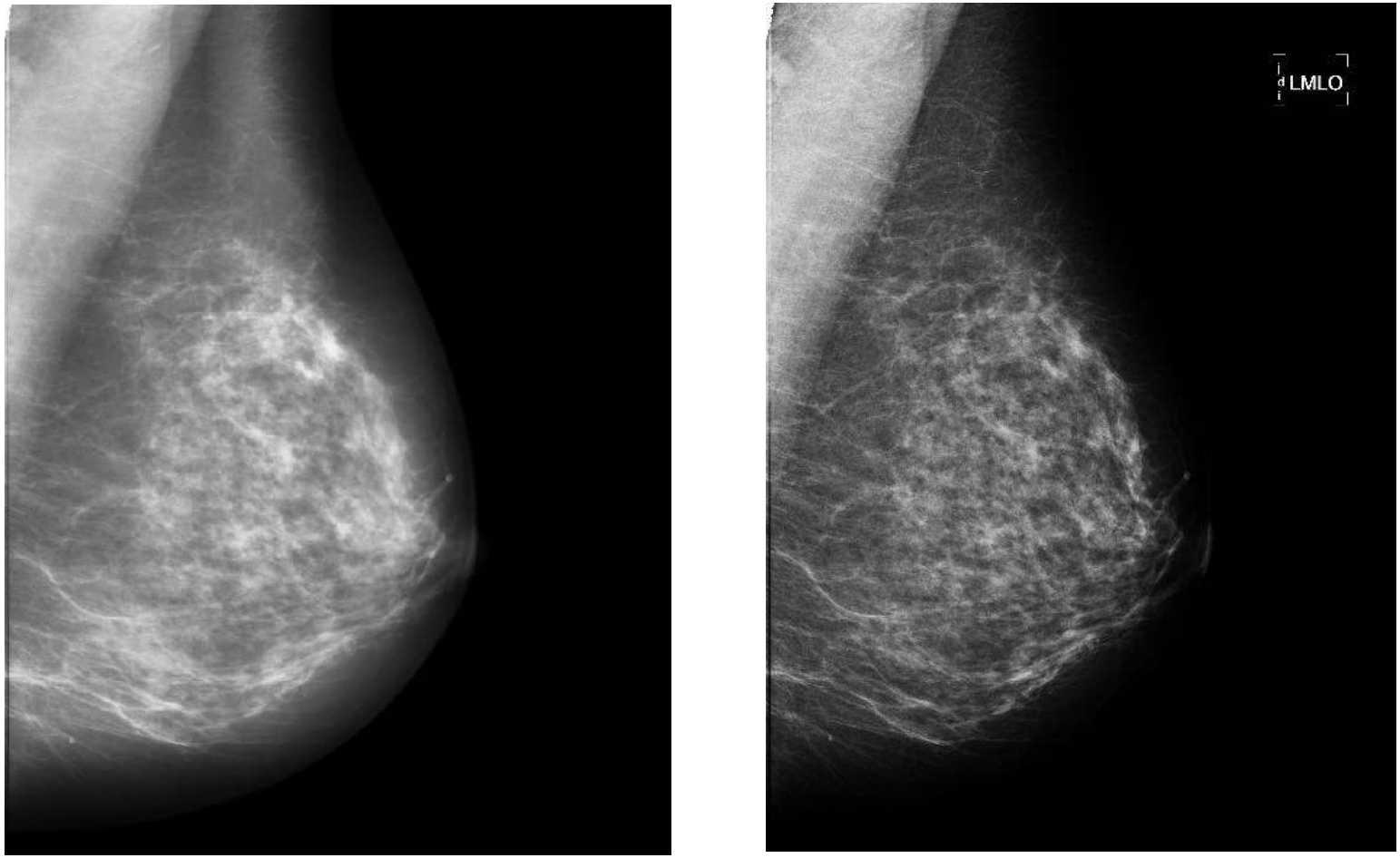

Figure 7: Examples of images processed with the algorithm presented in this study (left) and Siemens' commercial algorithm (right).

4. Discussion

There are two obviously distinct groups of curves between Figures 5 and 6, suggesting that multi-view stereoscopy has a statistically significant advantage over mammography in lesion detection. The significance is reinforced by the low p-value, which shows that the two modalities are not equal. Although the dose used to acquire the MVS image sets was slightly higher than that used to acquire the mammograms, this doesn't explain the significant increase in lesion detection using that modality. Also, the higher AUC values for multi-view stereoscopy are contrary to the fact that the MVS image sets were acquired with 1×2 acquisition binning and that the images have a lower resolution than that of the mammographic images. It is also contrary to the fact that each projection of the MVS sets was acquired with about 5% of the dose of the mammograms.

Even though the MVS image sets are significantly more noisy (because of the low exposures used to create each projection image), they provide the observer with more volumetric information than a single stereo pair or standard mammogram with which to make a diagnosis. These results agree with what was found by Chawla et al[20], where the optimal acquisition geometry for a multi-projection breast imaging system was investigated under equivalent total dose conditions. Adjusting for the 31% difference in total dose between mammography and MVS in our study, the data from Chawla et al[20,21] indicate 0.07-0.11 improvement in AUC for 25 projections as compared to mammography (see Figure 7 of reference 20 and Figure 11(a) of reference 21). In our

study, we found a corresponding ΔAUC of 0.164. The similarity of magnitude between these comparative results is encouraging and highlights the added value of using angular information. The differences (0.07-0.11 versus 0.164) may be attributed to the fact that the lesions used by Chawla et al[20] were less anthropomorphic. But more importantly, we implemented a particular visualization of the angular information through stereoscopy, which evidently provides a more visually efficient method for visualization of three-dimensional information.

Another method for visualizing three-dimensional breast information is through tomosynthesis. As mentioned previously, the MVS image sets are the raw projection data taken from a tomosynthesis unit. Normally these raw data are reconstructed into a three-dimensional tomosynthesis image set which is viewed in cross sectional slices. However, tomosynthesis images are more numerous to interpret (typically more than 50 high-resolution slices per breast). Moreover, they exhibit out-of-plane blur artifacts that are common to the tomosynthesis technique. MVS provides improved lesion detection, as does tomosynthesis, but it is not hindered by the artifacts associated with tomosynthesis. MVS may also be a more intuitive method of visualizing the breast volume than tomosynthesis. Notwithstanding this conclusion, a follow up study is warranted to compare the relative performance of MVS and tomosynthesis for breast cancer detection.

Currently, most standard stereomammography, such as what was implemented in the clinical trial at Emory, requires twice the dose of a normal mammogram[1]. In this study the MVS image sets were produced with a dose about only 31% higher than a normal mammogram and lesion detection was still significantly improved. Emory reported that stereomammography has a 46% reduction in false positive lesion detection ($p < 0.0001$) and a 23% increase in true positive lesion detection ($p < 0.05$) over mammography. For stereomammography this corresponds to an AUC of 0.687 and an AUC of 0.446 for standard mammography for an improvement in AUC of 0.241, which is better than our value of 0.164. However, in our study the AUC values for MVS and mammography were higher than what was found by Emory. An important note is that the Emory study focused only on lesions and did not differentiate between masses and microcalcifications.

It has been suggested that if stereomammography were done with half-dose image pairs, it would have a negligible effect on lesion detection since the effective SNR increases by a factor of $\sqrt{2}$ because of binocular summation of the two images[22]. This binocular summation effect has also impacted the SNR of the MVS sets and mammograms that were used in this study. Related to this effect is the relative importance of quantum and anatomical noise in the detection of breast masses. As shown by Samei et al[7], decreasing radiation dose in digital mammography by as much as 50% would have a minimal effect on the detection of masses. However, this dose

decrease would have a more noticeable effect on the detection of microcalcifications, the discrimination between benign and malignant masses, and interpretation time.

One obstacle encountered in this study was that some observers thought some of the simulated lesions appeared to be benign, which resulted in the observer giving a low confidence score to the image. However, in Saunders' paper[3] the radiologist observers judged that the malignant lesions were realistic and generally rated simulated lesions similarly to real lesions. An important caveat is that Saunders' study was done with 2-D mammograms and it is possible that his simulated lesions begin to look more benign in an MVS display. Also, in our study some observers tried to classify the lesions as real or simulated because they were told the masses were simulated. Because of this, we emphasized that the task at hand was mass detection and not classification (i.e. not determining real vs. simulated or malignant vs. benign). We should further add that this study did not simulate deformity of the mass as a function of projection angle, as the deformation was deemed negligible at the small angular ranges employed in this study and the lesions appeared realistic by our radiologists. Also, this study only dealt with the detection of masses. Other diagnostic tasks including the detection of microcalcifications, the detection of architectural distortion, and classification of lesions into benign and malignant are objectives that merit future studies. Related to this, another limitation of this study is that we were trying to detect simulated masses, not "real" masses. Sometimes with mammography, a radiologist may not see a mass but

may suspect one if there are changes in the breast architecture such as distortion of the Cooper's ligaments or a puckering of the tissue due to desmoplastic changes from cancer. We cannot really look for these changes with a simulation. Finally, we should add that we had no way to measure the false positive rate, or the rate of recall due to the observer's memory of a similar image seen previously in the study.

5. Conclusions

The aim of this study was to assess the relative performance of multi-view stereoscopy (MVS) compared to mammography for breast mass detection. The mean AUC for the five observers was 0.614±0.055 for mammography and 0.778±0.052 for MVS. The difference of 0.164±0.065 was statistically significant with a p-value of 0.0148, which implies MVS offers a significant advantage over mammography in the detection of breast masses. This highlights the dominance of anatomical noise compared to quantum noise for breast mass detection and how useful three-dimensional information can be to make a correct diagnosis. It also shows that significant lesion detection can be achieved with MVS, thus circumventing the artifacts associated with tomosynthesis and the extra dose associated with standard stereomammography.

References

1 Getty, D. J., D'Orsi, C. J. & Pickett, R. A. Stereoscopic digital mammography: Improved accuracy of lesion detection in breast cancer screening. *Lect Notes Comput Sc* **5116**, 74-79 (2008).

2 Getty, D. J., Pickett, R. M. & D'Orsi, C. J. Stereoscopic digital mammography: Improving detection and diagnosis of breast cancer. *Cars 2001: Computer Assisted Radiology and Surgery* **1230**, 538-544 (2001).

3 Saunders, R., Samei, E., Baker, J. & Delong, D. Simulation of mammographic lesions. *Acad Radiol* **13**, 860-870, doi:10.1016/j.acra.2006.03.015 (2006).

4 Saunders, R. S. & Samei, E. The effect of breast compression on mass conspicuity in digital mammography. *Med Phys* **35**, 4464-4473, doi:10.1118/1.2977600 (2008).

5 Chawla, A. S., Saunders, R., Abbey, C., Delong, D. & Samei, E. Analyzing the effect of dose reduction on the detection of mammographic lesions using mathematical observer models - art. no. 614601. *Proc. of SPIE* **6146**, 614601, doi:10.1117/12.656378 (2006).

6 Chawla, A. S., Samei, E., Saunders, R., Abbey, C. & Delong, D. Effect of dose reduction on the detection of mammographic lesions: A mathematical observer model analysis. *Med Phys* **34**, 3385-3398 (2007).

7 Samei, E., Saunders, R. S., Baker, J. A. & Delong, D. M. Digital mammography: Effects of reduced radiation dose on diagnostic performance. *Radiology* **243**, 396-404, doi:10.1148/radiol.2432061065 (2007).

8 Zuiderveld, K. J. in *Graphic Gems IV* 474–485 (Academic Press Professional, 1994).

9 *Planar 3D: Stereoscopic Display Technology*, <http://www.planar3d.com/3d-technology/stereoscopic-101/> (2010).

10 Radiological Society of North America. Scientific Assembly (89th : 2003 : Chicago Ill.), Samei, E. & Flynn, M. J. *2003 syllabus : advances in digital radiography : categorical course in diagnostic radiology physics*. (Radiological Society of North America, 2003).

11 Dorfman, D. D., Berbaum, K. S. & Metz, C. E. Receiver Operating Characteristic Rating Analysis - Generalization to the Population of Readers and Patients with the Jackknife Method. *Invest Radiol* **27**, 723-731 (1992).

12 Dorfman, D. D., Berbaum, K. S., Lenth, R. V., Chen, Y. F. & Donaghy, B. A. Monte Carlo validation of a multireader method for receiver operating characteristic discrete rating data: Factorial experimental design. *Acad Radiol* **5**, 591-602 (1998).

13 Hillis, S. L. & Berbaum, K. S. Power estimation for the Dorfman-Berbaum-Metz method. *Acad Radiol* **11**, 1260-1273, doi:10.1016/j.acra.2004.08.009 (2004).

14 Hillis, S. L. & Berbaum, K. S. Monte Carlo validation of the Dorfman-Berbaum-Metz method using normalized pseudovalues and less data-based model simplification. *Acad Radiol* **12**, 1534-1541, doi:10.1016/j.acra.2005.07.012 (2005).

15 Hillis, S. L., Obuchowski, N. A., Schartz, K. M. & Berbaum, K. S. A comparison of the Dorfman-Berbaum-Metz and Obuchowski-Rockette methods for receiver operating characteristic (ROC) data. *Stat Med* **24**, 1579-1607, doi:10.1002/Sim.2024 (2005).

16 Hillis, S. L. A comparison of denominator degrees of freedom methods for multiple observer ROC analysis. *Stat Med* **26**, 596-619, doi:10.1002/Sim.2532 (2007).

17 Hillis, S. L., Berbaum, K. S. & Metz, C. E. Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. *Acad Radiol* **15**, 647-661, doi:10.1016/j.acra.2007.12.015 (2008).

18 *Kurt Rossmann Laboratories for Radiologic Image Research*, <http://xray.bsd.uchicago.edu/krl/> (2003).

19 *Medical Image Perception Laboratory*, <http://perception.radiology.uiowa.edu/> (2006).

20 Chawla, A. S., Samei, E., Saunders, R. S., Lo, J. Y. & Baker, J. A. A mathematical model platform for optimizing a multiprojection breast imaging system. *Med Phys* **35**, 1337-1345, doi:10.1118/1.2885367 (2008).

21 Chawla, A. S., Lo, J. Y., Baker, J. A. & Samei, E. Optimized image acquisition for breast tomosynthesis in projection and reconstruction space. *Med Phys* **36**, 4859-4869, doi:Doi 10.1118/1.3231814 (2009).

22 Maidment, A. D. A., Bakic, P. R. & Albert, M. Effects of quantum noise and binocular summation on dose requirements in stereoradiography. *Med Phys* **30**, 3061-3071, doi:10.1118/1.1621869 (2003).