

ROC study of the effect of stereoscopic imaging on assessment of breast lesions

Heang-Ping Chan,^{a)} Mitchell M. Goodsitt, Mark A. Helvie, Lubomir M. Hadjiiski, Justin T. Lydick, Marilyn A. Roubidoux, Janet E. Bailey, Alexis Nees, Caroline E. Blane, and Berkman Sahiner

Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109

(Received 12 November 2004; revised 15 January 2005; accepted for publication 22 January 2005; published 22 March 2005)

An observer performance study was conducted to evaluate the usefulness of assessing breast lesion characteristics with stereomammography. Stereoscopic image pairs of 158 breast biopsy tissue specimens were acquired with a GE Senographe 2000D full field digital mammography system using a $1.8\times$ magnification geometry. A phantom-shift method equivalent to a stereo shift angle of $\pm 3^\circ$ relative to a central axis perpendicular to the detector was used. For each specimen, two pairs of stereo images were taken at approximately orthogonal orientations. The specimens contained either a mass, microcalcifications, both, or normal tissue. Based on pathological analysis, 39.9% of the specimens were found to contain malignancy. The digital specimen radiographs were displayed on a high resolution MegaScan CRT monitor driven by a DOME stereo display board using in-house developed software. Five MQSA radiologists participated as observers. Each observer read the 316 specimen stereo image pairs in a randomized order. For each case, the observer first read the monoscopic image and entered his/her confidence ratings on the presence of microcalcifications and/or masses, margin status, BI-RADS assessment, and the likelihood of malignancy. The corresponding stereoscopic images were then displayed on the same monitor and were viewed through stereoscopic LCD glasses. The observer was free to change the ratings in every category after stereoscopic reading. The ratings of the observers were analyzed by ROC methodology. For the 5 MQSA radiologists, the average A_z value for estimation of the likelihood of malignancy of the lesions improved from 0.70 for monoscopic reading to 0.72 ($p=0.04$) after stereoscopic reading, and the average A_z value for the presence of microcalcifications improved from 0.95 to 0.96 ($p=0.02$). The A_z value for the presence of masses improved from 0.80 to 0.82 after stereoscopic reading, but the difference fell short of statistical significance ($p=0.08$). The visual assessment of margin clearance was found to have very low correlation with microscopic analysis with or without stereoscopic reading. This study demonstrates the potential of using stereomammography to improve the detection and characterization of mammographic lesions. © 2005 American Association of Physicists in Medicine. [DOI: 10.1118/1.1870172]

Key words: digital mammography, stereoscopic imaging, specimen radiograph, observer performance study, ROC methodology

I. INTRODUCTION

Mammography is currently the only recommended imaging technique for breast cancer screening. However, mammographic sensitivity is often limited by the presence of dense breast parenchyma.¹ It has been reported that the false negative rate of mammography in dense breasts can be as high as 25%.^{2,3} One of the main factors contributing to these missed cases is the camouflaging effect of the overlapping structures in the projection x-ray images. With the advent of high-resolution digital detectors for mammography, a number of new breast imaging techniques such as stereomammography,⁴⁻¹² digital tomosynthesis,¹³⁻¹⁵ and computed tomography¹⁶⁻¹⁸ are being developed in an effort to alleviate this problem. These techniques attempt to view the breast in three dimensions (3D) or to slice the breast volume into thin planes so as to reduce the superposition of breast tissue structures as imaged in two-dimensional (2D) projection mammograms. An observer performance study by

Getty *et al.*⁸ indicated that digital stereomammography improved the estimate of the probability of malignancy of mammographic lesions and allowed the detection of additional lesions that were obscured on screen-film mammograms. Rafferty *et al.*¹⁹ also demonstrated that digital tomosynthesis mammograms could reveal additional lesions obscured by dense breast tissue and improved visualization of the margins and spiculations of masses.

Stereoscopic imaging requires acquisition of a left-eye image and a right-eye image. In conventional film-based stereoradiography, two film images were obtained by shifting the x-ray source, along a direction parallel to the image plane, to the left and the right of the central axis of the imaging system. When the two film images are placed properly and viewed so that the left eye sees only the left-eye film and the right eye sees only the right-eye film, the parallax between the two images creates the depth perception. Stereoscopic imaging was utilized for various types of radiographic

examinations.^{20–25} However, it did not receive widespread acceptance in clinical practice, mainly because of the doubled film cost and increased patient exposure.²⁶ In addition, radiologists had to read the stereoradiographs with a somewhat cumbersome film stereoscope or had to be trained to read the stereoradiographs without aid using a “cross-eyed” technique.

In recent years, direct digital detectors have become available for medical imaging. Stereoradiography may become a viable approach with digital imaging because there are no additional film costs. Furthermore, digital detectors have a linear response, wider dynamic range, and higher contrast sensitivity than screen-film systems so that good-quality digital stereo image pairs may be acquired at essentially the same total radiation dose as that for a conventional single-projection screen-film image. Maidment *et al.*¹² found that human eyes can integrate the noise in the left-eye and right-eye images such that the detectability of simulated low contrast objects on a uniform noisy background in a single image was comparable to that of viewing the left- and right-eye image pair when the total dose of the latter was about $1.1\times$ of the dose of the single image. Maidment’s experimental design evaluated the efficiency of noise reduction by binocular summation without utilizing the potential additional advantage of stereo depth perception in signal detection. It is likely that this additional advantage would further reduce the total dose requirements for stereo imaging to the same as or even lower than those for a single-projection image. Digital stereoscopic images can be viewed more conveniently than stereo film radiographs because of the electronic display. Different methods for displaying digital stereoscopic images are still being developed. One common method is to display the left-eye and right-eye images alternately at a very fast refresh rate on a monitor. The images are viewed with a pair of special goggles that typically consist of liquid crystal electronic shutters. The shutters are synchronized with the display so that the left eye of the reader is allowed to see only the left-eye image and the right eye is allowed to see only the right-eye image. For high-resolution medical images such as mammograms, no commercial stereo display systems are available at present.

Stereoradiography provides structural information of the object being viewed in 3D. It has been reported that the spatial distribution of microcalcifications may be associated with the malignant or benign nature of the cluster.^{27,28} Masses may be better separated from the overlapping fibroglandular tissues in stereo than that in a 2D mammogram, making it easier to visualize the margin characteristics and determine whether spiculations are present. Therefore, stereomammography has the potential of providing additional diagnostic information that may improve the characterization of malignant and benign lesions and reduce unnecessary biopsies.

We are developing stereomammography techniques using a digital mammography system. In our previous studies, we examined the effects of stereo shift, geometric magnification, x-ray exposure, and display zooming on visual depth discrimination of crossing fibrils in stereo phantom images.^{4,5,10}

We found that a 2 mm depth discrimination could be achieved with over 90% accuracy on magnification images. We also investigated the accuracy of using a calibrated virtual cursor to measure the absolute depth of fibrils in stereoscopic images.^{6,7,11} Our results showed that the average root-mean-square errors of depth measurements in stereo images with the virtual cursor ranged from 0.2 to 1.3 mm, depending on the stereo shift angle and the imaging geometry. These studies demonstrated that stereoscopic imaging can provide both qualitative depth discrimination and quantitative measurement of fibrous structures in a breast. In the present investigation, we conducted an observer performance study using receiver operating characteristic (ROC) methodology to investigate the effects of stereoscopic reading on the accuracy of detection and characterization of mammographic lesions using images of biopsied breast tissue specimens.

II. MATERIALS AND METHODS

A. Data set

Digital stereoscopic image pairs of the breast tissue specimens were acquired with a GE Senographe 2000D full field digital mammography (FFDM) system. The study was approved by the Institutional Review Board. The GE system uses a flat panel digital detector composed of a CsI:Tl scintillator and an amorphous-Si active matrix array. The detector has a pixel size of $100\ \mu\text{m} \times 100\ \mu\text{m}$ and an output gray level resolution of 14 bits. The raw images are routinely processed with GE proprietary software and converted to 12 bit processed images. We employed a $1.8\times$ magnification geometry (no grid, 0.15 mm focal spot) and a stereo shift angle of $\pm 3^\circ$ for imaging the stereoscopic specimen radiographs.

The conventional method for stereoradiography is to move the x-ray source to the left and the right of the central ray by a chosen stereo shift angle $\pm\theta^\circ$ (or stereo shift distance $\pm w$) for acquiring the left-eye and right-eye images. In the early days of radiography, it was determined by trial and error that a total tube shift equal to 10% of the focus-to-film distance produced satisfactory stereo results.²⁶ This is equivalent to a tube shift of about $\pm 3^\circ$ ($\cong \frac{1}{2} \tan^{-1}(0.1)$). In our previous studies,^{4–7,9–11} we also found that $\pm 3^\circ$ would provide sufficient stereoscopic vision without causing excessive eye strain. The FFDM system was not designed for stereoscopic imaging. It does not have an electronic or mechanical lock mechanism to keep the x-ray tube stationary at the appropriate shift angle, nor do the collimator blades adjust to maintain complete coverage of the detector when the x-ray tube is shifted. We designed a stereo image acquisition method for phantoms and specimens in which the object is shifted instead of the focal spot. As illustrated in Fig. 1, the exposure geometry for the object relative to the focal spot when the focal spot is shifted to the left is equivalent to that when the focal spot is stationary and the object is shifted to the right by the same distance. Similarly, the geometry when the focal spot is shifted to the right is equivalent to that when the focal spot is stationary and the object is shifted to the left. A small error is caused by the slightly shorter focal-spot-to-detector distance in the object-shift geometry because the

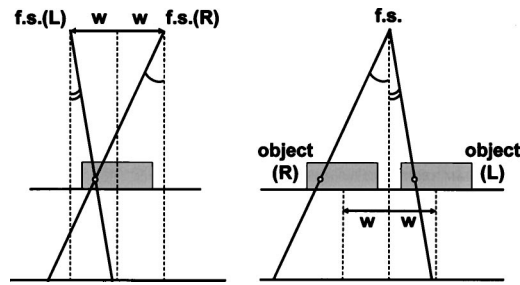


FIG. 1. Imaging geometry for acquisition of stereoscopic image pairs in magnification geometry. Left panel: a conventional “focal-spot shift” method in which the focal spot is shifted to the left and to the right of the central ray by a distance w to expose the left-eye and right-eye image. Right panel: an equivalent “object shift” method in which the object is shifted to the right and to the left of the central ray by the same distance, w . It can be seen that the image exposed by the f.s.(L) geometry is equivalent to that exposed by the object (L) geometry. Similarly, the image exposed by the f.s.(R) geometry is equivalent to that exposed by the object (R) geometry.

x-ray focal spot moves along an arc. This error is estimated to be less than 0.1% for a $\pm 3^\circ$ stereo angle shift and a fulcrum of rotation at 46 cm from the focal spot. Using the geometry of the GE system and the $\pm 3^\circ$ stereo angle shift used in this study, the object shift distance, w , can be calculated to be ± 2.4 cm from the central position. For a given stereo angle shift, the linear shift distance is the same for both the contact geometry and the magnification geometry. The phantom-shift technique was also used in our previous phantom studies.^{10,11}

To facilitate the shifting of the object in a direction parallel to the chest wall (focal spot shift direction) for the FFDM system, we built a platform using Lexan plates shown in Fig. 2. The platform has a stationary base that fits on the magnification stand. The object is placed on a sliding plate on top of the base. The sliding plate can be moved manually between two guardrails in a direction parallel to the chest wall. The central position and the left and right shift distances were marked on the stationary base. The tissue specimens could therefore be moved to the desired left and right shift locations easily and precisely. Two fiducial markers (small metal rings) were affixed to the sliding plate. Their positions in the images were later used for alignment of the left-eye and right-eye images of the stereo pairs.

Consecutive biopsied breast tissue samples that were sent to the radiology department for specimen radiographs were imaged additionally with the stereoscopic technique if the FFDM system was available. The specimens were therefore random samples without selection. Each specimen could contain microcalcifications, mass, both, or normal tissue. Some specimens were obtained with ultrasound-guided biopsy of mammographically occult masses. The normal tissue was usually a result of a second biopsy to excise additional margins if the first tissue specimen was found to have a close margin. Two sets of stereo image pairs were acquired of each sample. These were acquired in approximately orthogonal orientations, whereby the second set was obtained by rolling the sample over by approximately 90° . The exposure techniques were manually chosen by mammography technolo-

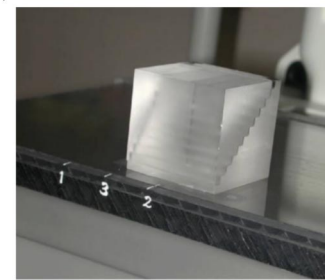
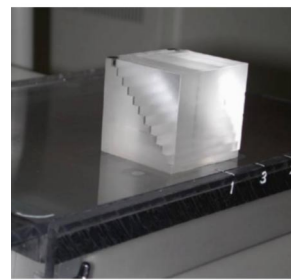


FIG. 2. The platform fits on the magnification stand of the FFDM system: (a) the sliding plate on top of the stationary base at the central position, marked as 3, (b) the sliding plate was shifted to the left position at 2.4 cm, marked as 1, and (c) the sliding plate was shifted to the right position at 2.4 cm, marked as 2. The stepwedge phantom shows where the tissue specimen would be placed.

gists. The mammography technologists were instructed to use high dose, identical techniques for the left-eye and right-eye images. The target/filter combinations were mainly Mo/Mo with Mo/Rh in some cases. The kilovoltage ranged from 24 to 27 kVp and the mAs ranged from 40 to 80 mAs, depending on the thickness of the tissue specimen.

All stereo image pairs were visually inspected for alignment and exposure by an experienced physicist. Some samples were rejected because of improper shift between the left-eye and right-eye images or improper exposure. All image pairs with good stereoscopic quality and exposure were included. This resulted in a total of 316 stereo image pairs from 158 specimens for the observer experiment. Based on pathological analysis 39.9% of the chosen samples were proven to contain malignancy. The lesion types and the number of lesions of each type for the samples used are listed in Table I. Examples of stereo image pairs of the tissue speci-

TABLE I. The lesion types and number of tissue specimens in each type.

Lesion	Malignant	Benign	Total
Mass	21	31	52
Microcalcifications	14	38	52
Both	24	9	33
No visible mass or microcalcifications	4	17	21
Total	63 (39.9%)	95 (60.1%)	158

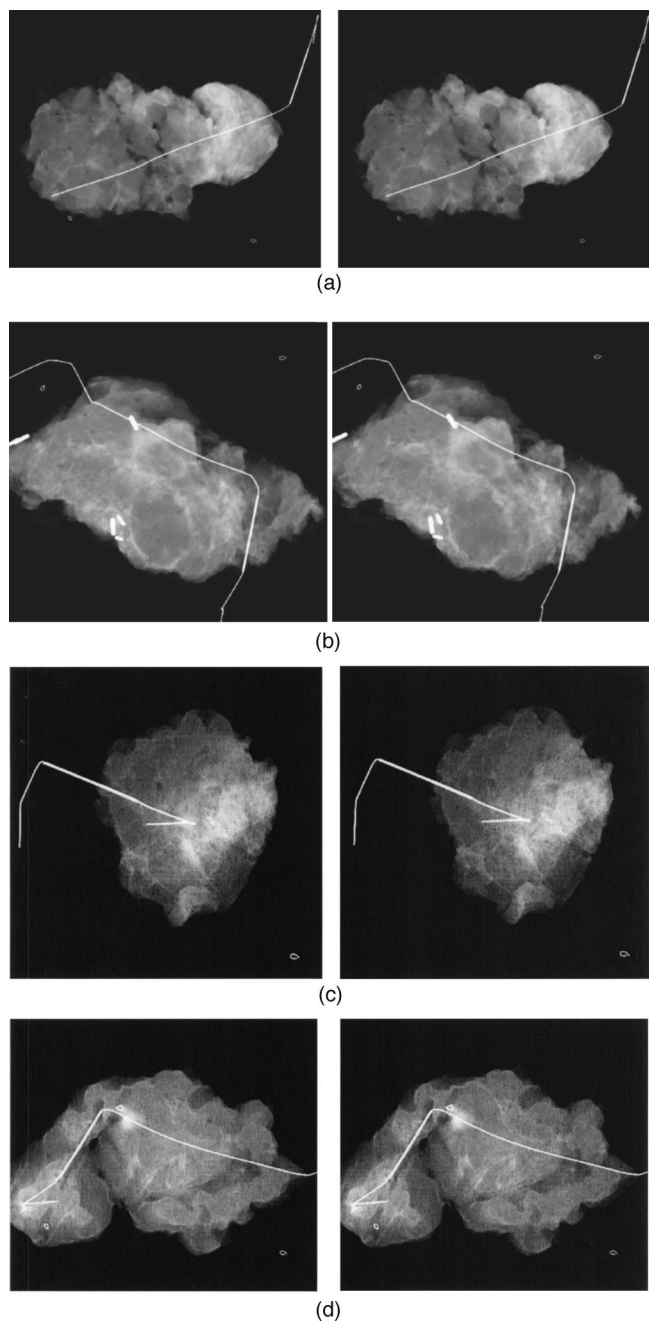


FIG. 3. Examples of stereo image pairs (left-eye image and right-eye image) of breast tissue specimens: (a) specimen with microcalcifications—invasive ductal carcinomas, (b) specimen with mass—invasive ductal carcinomas, (c) specimen with radial scar and microcalcifications—fibrocystic change, and (d) specimen with mass—fibrocystic change and fat necrosis.

mens are shown in Fig. 3.

B. Stereo image display

The images were displayed on a stereo workstation that consists of a MegaScan 8 mega-pixel CRT monitor driven by a Dome Md8-4820-LS stereoscopic board and a PC. The monitor was adjusted with a photometer to meet the DICOM grayscale standards, and the room lights were dimmed to a very low level during the observer studies. The system can

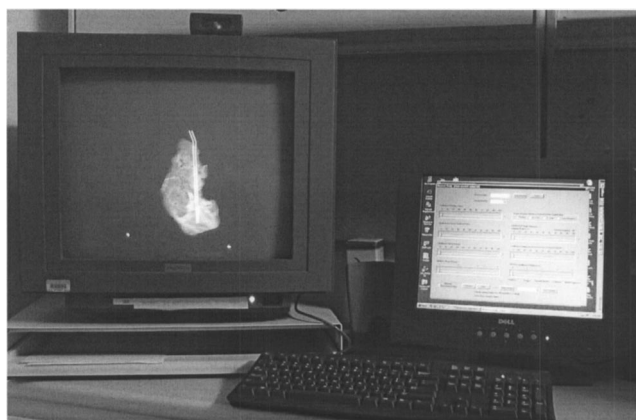


FIG. 4. Stereo display workstation composed of a MegaScan 8 mega-pixel monitor driven by a Dome Md8-4820-LS stereoscopic board and a PC. The system can display full-field (2300 × 1800 pixels) digital mammograms at a refresh rate of 120 Hz.

display full-field (2300 × 1800 pixels) digital mammograms at a refresh rate of 120 Hz. It operates in a page flipping stereoscopic mode with the left- and right-eye images displayed alternately. A pair of CrystalEyes LCD stereoscopic glasses was used for viewing the stereoscopic images. The stereo images were displayed with in-house developed software that provided functions to shift and align the left-eye and right-eye images, adjust the contrast and brightness, and store the selected alignment and windowing settings. The stereo display workstation is shown in Fig. 4. The physicist selected and saved the settings for each image pair which became the default settings when this image pair was displayed the next time. The same display conditions could therefore be used for all radiologists in the observer study. The radiologist had the option of adjusting the window settings if they deemed it necessary. The software could also switch the display to show the left-eye image alone or the right-eye image alone so that the observer could read the monoscopic image and the stereoscopic images of the same case sequentially, as designed for the observer experiment described in the following.

C. Observer performance study

A user interface was designed for the observer experiment. The user interface displayed images sequentially according to an input list. Slide bars were provided to record the observer's confidence ratings (scale of 1–100) regarding the presence of a mass, the presence of calcifications, the likelihood of malignancy of the lesion if present, the likelihood of the margin being clear. The observers were also asked to provide an assessment of malignancy in terms of the BI-RADS categories (1=negative, 2=benign, 3=probably benign, 4=suspicious, 5=highly suggestive of malignancy),²⁹ and a visual estimate of the margin clearance (0=positive margin, 1=0–2 mm, 2=2–5 mm, 3=greater than 5 mm). Five Mammography Quality Standards Act (MQSA) qualified radiologists participated in the experiment. The experiment was designed to have each observer

read the 316 specimen images in two sessions. The two views of each specimen were read independently and were arranged to be read in the two separate sessions to reduce the possibility of memorization. It may be noted that this was not equivalent to using 316 truly independent samples in the observer experiment. This increased the sample size but the possible correlation between the two views may cause a slight underestimation of the variances in the data. The reading sequence was systematically arranged in a counter-balanced design so that no specific cases were read by all observers always in the first or the second session. The case reading order was different for each observer. The observers first read the left-eye image alone as a monoscopic image and entered their assessments in all categories. The stereoscopic images were then displayed and were read with the LCD glasses. The observers were free to change their ratings in every category after reading the stereoscopic images. The observers were allowed unlimited time to read each case. They were also free to break the reading sessions into shorter ones. The radiologists were informed of the fact that the samples were randomly collected from the biopsied tissue specimens so that the proportion of malignant and benign cases would be similar to that in their clinical practice. They were therefore also aware that some specimens could be found to be negative for lesions or malignancy by pathological analysis.

Before a radiologist was recruited as observer, he/she underwent a standard Randot Circles Stereo test (Stereo Optical Co., Inc., Chicago, IL) to evaluate their stereo acuity. The reader viewed ten sets of circles on the test pattern through polarized glasses. Each set contained three circles, one of which would appear to be at a different depth from the others when viewed stereoscopically. The reader was asked to identify the circle that stood out in each of the ten sets. All radiologists participated in our observer performance study could correctly identify 9 to 10 of the circles, indicating that their level of stereopsis was at least 30 s of arc at a viewing distance of 16 in. Prior to reading the test stereo images, the observer also participated in a training session to become familiar with the reading task and the user interface.

D. Data analysis

The confidence ratings and the BI-RADS assessments of the observers were analyzed with the LABMRMC program.³⁰ The area under the ROC curve, A_z , and the partial area index above a sensitivity of 0.90, $A_z^{(0.9)}$, were used to compare the performance between monoscopic reading and monoscopic assisted with stereoscopic reading. The statistical significance of the difference in A_z between the two was estimated by the two-tailed p -value from the LABMRMC program and the Student's paired t -test. The average A_z and $A_z^{(0.9)}$ values were obtained from the average ROC curve that was derived from the average slope and intercept parameters of the individual readers' ROC curves. For the classification of malignant and benign lesions, all samples were analyzed together regardless of the lesion type.

TABLE II. Performance of radiologists in detecting microcalcifications in the tissue specimens with monoscopic (single projection) reading and with additional stereoscopic reading. The average A_z and $A_z^{(0.9)}$ were derived from the average a and b parameters of the individual ROC curves. The improvements in A_z and $A_z^{(0.9)}$ were both statistically significant with $p=0.02$ and $p=0.004$, respectively.

Radiologist	A_z		$A_z^{(0.9)}$	
	Monoscopic	With stereo	Monoscopic	With stereo
1	0.97±0.01	0.98±0.01	0.76	0.79
2	0.95±0.01	0.96±0.01	0.58	0.67
3	0.94±0.02	0.95±0.02	0.48	0.55
4	0.94±0.02	0.95±0.01	0.58	0.65
5	0.92±0.02	0.92±0.02	0.30	0.36
Average	0.95	0.96	0.57	0.63

For the analysis of the visual assessment of the margin status of the specimens in comparison with pathologists' analysis, we first combined the margin assessments from the two orthogonal views of the same specimen by taking the minimum margin clearance seen by the radiologist in the two views. This simulated the situation in which the radiologist was allowed to see the margins from the two different projections and estimated the minimum margin clearance from all visible borders, as they do in reading specimen radiographs in their routine clinical practice. The correlation of the radiologists' assessment of margin clearance with the result of pathological analysis was evaluated by the Pearson's correlation coefficient. Since pathological reports included margin assessment only for malignant lesions, only this subset of cases was used in the correlation analysis.

III. RESULTS

The radiologists' accuracy in detection of microcalcifications in the specimen by reading a single-projection image in comparison to that with additional stereoscopic reading is shown in Table II. In this ROC analysis, all samples with microcalcifications (malignant and benign) were considered to be positive cases. The samples with mass alone or without either mass or microcalcifications were treated as negative cases with respect to microcalcifications. The detection of microcalcifications in the small volume of tissue specimens appeared to be easy with or without stereoscopic reading. The A_z values for the five radiologists ranged from 0.92 to 0.97 with an average of 0.95 for monoscopic reading. Nevertheless, the radiologists still improved their performance with additional stereoscopic reading, with the A_z values ranging from 0.92 to 0.98 and an average of 0.96. The improvement, although modest, was consistent over all radiologists (the A_z value of Radiologist 5 improved from 0.918 to 0.922). The partial area index $A_z^{(0.9)}$ values for the radiologists were also high, ranging from 0.30 to 0.76 with monoscopic reading and improved to a range of 0.36 to 0.79 with

TABLE III. Performance of radiologists in detecting masses in the tissue specimens with monoscopic (single projection) reading and with additional stereoscopic reading. The average A_z and $A_z^{(0.9)}$ were derived from the average a and b parameters of the individual ROC curves. The improvements in A_z and $A_z^{(0.9)}$ both fell short of statistical significance with $p=0.08$ and $p=0.11$, respectively.

Radiologist	A_z		$A_z^{(0.9)}$	
	Monoscopic	With stereo	Monoscopic	With stereo
1	0.83±0.02	0.84±0.02	0.19	0.22
2	0.75±0.03	0.79±0.02	0.11	0.18
3	0.81±0.02	0.82±0.02	0.24	0.28
4	0.83±0.03	0.83±0.03	0.25	0.24
5	0.80±0.03	0.81±0.02	0.16	0.17
Average	0.80	0.82	0.19	0.22

additional stereoscopic reading. The improvements in A_z and $A_z^{(0.9)}$ were both statistically significant with $p=0.02$ for A_z and $p=0.004$ for $A_z^{(0.9)}$.

The radiologists' accuracy in detection of masses with the two reading conditions is compared in Table III. Similar to the ROC analysis for microcalcifications, all samples with masses were considered positive. The samples with microcalcifications alone or without either mass or microcalcifications were considered negative for masses. For monoscopic reading, the A_z values of the radiologists ranged from 0.75 to 0.83 with an average A_z of 0.80. With additional stereoscopic reading, the A_z values for four of the five radiologists improved. The A_z ranged from 0.79 to 0.84 and the average A_z was improved to 0.82. However, the improvements in both A_z and $A_z^{(0.9)}$ fell short of statistical significance with $p=0.08$ and $p=0.11$, respectively.

Table IV shows the comparison of the radiologists' assessments of the likelihood of malignancy of the tissue specimens with and without stereoscopic reading. With monoscopic reading, the A_z values of the radiologists ranged from 0.65 to 0.74. Their accuracy improved significantly ($p=0.04$) with additional stereoscopic reading to the range of

TABLE IV. Performance of radiologists in classification of malignant and benign lesions in the tissue specimens with monoscopic (single projection) reading and with additional stereoscopic reading. The average A_z and $A_z^{(0.9)}$ were derived from the average a and b parameters of the individual ROC curves. The improvements in A_z and $A_z^{(0.9)}$ were both statistically significant with $p=0.04$ and $p=0.04$, respectively.

Radiologist	A_z		$A_z^{(0.9)}$	
	Monoscopic	With stereo	Monoscopic	With stereo
1	0.72±0.03	0.74±0.03	0.07	0.09
2	0.73±0.03	0.78±0.03	0.12	0.19
3	0.74±0.03	0.74±0.03	0.09	0.11
4	0.65±0.03	0.67±0.03	0.10	0.11
5	0.68±0.03	0.70±0.03	0.10	0.13
Average	0.70	0.72	0.10	0.13

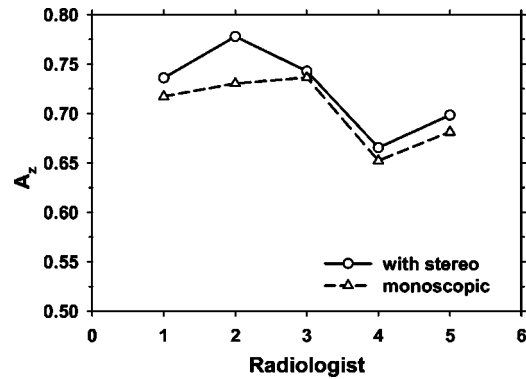


FIG. 5. The area under the ROC curves for the five radiologists for classification of malignant and benign lesions. The observers show a modest but consistent improvement in performance with additional stereoscopic viewing.

0.67 to 0.78 (Fig. 5). The partial area index $A_z^{(0.9)}$ also improved significantly ($p=0.04$) from a range of 0.07 to 0.12 to a range of 0.09 to 0.19.

Table V shows the changes in BI-RADS categories with stereoscopic reading. Since the BI-RADS assessment of categories 3 or above indicates the need of call-back for further evaluation, and categories 4 and 5 indicate a recommendation for biopsy, we summarized the changes in the BI-RADS categories across the threshold between categories 1, 2 and 3, 4, 5, and the threshold between categories 1, 2, 3 and 4, 5. By counting the number of lesions having an increase in the categories across the threshold as positive and a decrease as negative, the average number of lesions that had significant changes in BI-RADS categories over the five radiologists for the malignant lesions and the benign lesions were calculated. The results revealed that the radiologists improved their assessments of malignant lesions with stereoscopic reading. For malignant lesions, the BI-RADIS assessments for an average of 1.6 lesions (1.6/63=2.5%) per radiologist were changed from negative or benign to call-back, and an average of 2.2 lesions (2.2/63=3.5%) per radiologist were changed from categories 1, 2, and 3 to recommendation for biopsy. However, for benign lesions there were also increases in call-back and biopsy recommendations but the

TABLE V. The average number of lesions per radiologist of which the BI-RADS category was changed after stereoscopic reading. BI-RADS categories 3 or above represent a call-back and categories 4 or above represent biopsy recommendation. Positive change indicated an increase in the number of lesions from the lower to the higher categories and negative change indicated a decrease.

Change in BI-RADS assessment	Average number of lesions per radiologist	
	From categories 1, 2 to categories 3, 4, 5	From categories 1, 2, 3 to categories 4, 5
Malignant lesions	1.6	2.2
Benign lesions	1.2	0.4

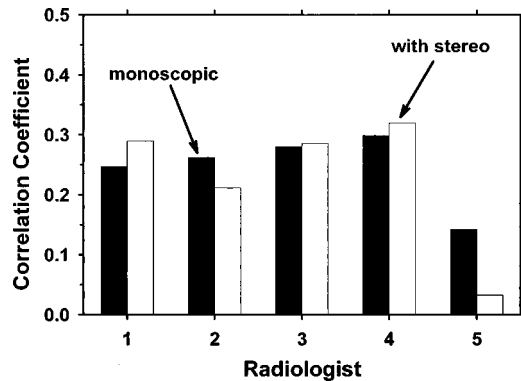


FIG. 6. The correlation coefficients between the radiologists' assessment of margin clearance and pathological analysis. The black bars were obtained with monoscopic reading, the white bars were obtained with additional stereoscopic reading.

changes were less, with an average of 1.2 ($1.2/95=1.3\%$) and 0.4 ($0.4/95=0.4\%$) lesions per radiologist for the two types of changes, respectively.

The correlation coefficients of the radiologists' assessment of margin clearance with pathological analysis are plotted in Fig. 6. The assessment of margin status visually in tissue specimens was found to be very unreliable. The correlation coefficients for all radiologists with or without stereoscopic reading were below about 0.3.

IV. DISCUSSION

The purpose of our study was to evaluate the potential advantages of stereo full-field digital mammography for the detection and characterization of breast lesions. Because of the difficulty of collecting a large data set of stereoscopic whole breast mammograms with lesions, we used stereo specimen radiographs for this preliminary study. Detection and characterization of lesions on specimen radiographs is different from similar tasks on FFDMs. Since the location of the lesion is confined to a smaller and thinner tissue sample than the whole breast, specimen radiographs should already provide superior visibility of lesion characteristics as compared to whole breast mammograms. Therefore, these are more difficult tasks for achieving improvements in the detection and characterization of the lesions. Nevertheless, our results indicate that the additional stereoscopic reading did improve the visualization of lesions and the accuracy of assessing their malignant or benign characteristics in specimen radiographs. Although the results cannot be generalized directly to reading whole breast mammograms, the potential for information gain and improvement in accuracy with stereoscopic reading have been demonstrated. In a study by Getty *et al.*⁸ comparing the characterization of mammographic lesions on film mammograms alone to that with additional reading of whole breast stereomammograms, they observed an improvement in A_z from 0.83 to 0.86. Their slightly larger improvement in A_z with whole breast mammograms than that obtained in our study appears to corroborate our expectations.

The observer performance results indicated that the data set used in this study was quite difficult even in specimen radiographs. The mass detection task was challenging even for experienced radiologists, probably because the samples contained a variety of abnormal and normal specimens including focal densities and mammographically occult masses that were imaged with ultrasound during the wire localization procedure. The characterization of malignant and benign lesions was also difficult because these lesions had been recommended for biopsy so that they all appeared to be suspicious to some degree. There were also cases in which the lesions were partially removed by core biopsy so that the appearance might not be typical. The variety of cases was included because the ROC experiment measured the relative improvement with additional stereoscopic reading for the given set of samples rather than the absolute performance of the radiologists in clinical practice.

Breast tissue specimens are routinely radiographed and read by radiologists to determine primarily if the lesion recommended for biopsy is included in the specimen and secondarily if the cancer extends to the margin in lumpectomy cases. The results are used by the surgeon to determine if additional excision is needed. The low correlation between the visual assessment of margin clearance with the pathologists' report is somewhat unexpected. It therefore indicates that visual assessment of margin status does not correspond very well with microscopic analysis. It is likely that the specimen radiograph is useful for estimating whether the lesion is far from the specimen's boundaries. However, if the lesion is close to the margin, i.e., within a few millimeters, specimen radiographs are not capable of showing whether microscopic amounts of malignant tissues are present at the boundary.

In this study we used a sequential reading method, namely, the observer first read with monoscopic viewing and provided their ratings, and this was immediately followed with stereoscopic viewing and second ratings. The second ratings therefore represented diagnostic decisions resulting from a combination of the information from the conventional monoscopic reading with that from the additional stereoscopic reading. This will likely be the reading mode used if stereoradiographs are available clinically because the left-eye and right-eye images are readily available for monoscopic viewing and because there is no need to trade off any existing benefits of conventional reading in exchange for the stereoscopic viewing. The radiologists may switch between the monoscopic and the stereoscopic images to extract complementary information or to confirm their observations. This information gain may be obtained without or with a minimal increase in patient exposure compared with current screen-film mammographic techniques. Further studies of interest include comparisons of the detection and characterization of lesions under the following sequential reading conditions: (1) monoscopic reading of either the left-eye or the right-eye image alone, (2) monoscopic readings of both the left-eye and right-eye images by switching back and forth between the two, and (3) with additional stereoscopic reading of the image pair. These comparisons will reveal if the slight shift

in the perspective obtained from monoscopic readings of both the left-eye and right-eye images will in itself provide sufficient information to improve the detection and characterization performances or if the additional stereoscopic reading with depth perception is essential. Another study of interest is a comparison of monoscopic readings of the two orthogonal views of the specimens with stereoscopic reading of one of the views or both views. This study will reveal if the 3D information obtained from orthogonal views is superior to that from stereoscopic reading of one of the views or if additional information can still be gained from stereoscopic reading of both views. Likewise, a comparison of monoscopic readings of CC view and MLO view mammograms to stereoscopic reading of the MLO view mammogram alone or both views will be an interesting study to evaluate how stereomammography may be implemented in clinical practice.

To simplify image acquisition and the observer experiment, we used the left-eye image of the stereo pair as the monoscopic image for reading. Since the stereo shift angle is only $\pm 3^\circ$, the difference in projection between an image taken at the central position (no-shift) and the left-eye (or the right-eye) image is very small. Each image of the stereo pair should be very similar to the central image. Furthermore, we instructed the technologists to use exposure techniques much higher than those used for a conventional specimen radiograph. The use of high dose techniques was intended to obtain monoscopic images of which the image quality would not be limited by quantum noise. This experimental design reduces the likelihood that the information gain with stereoscopic reading is due to the reduced noise when two monoscopic images were integrated into the stereoscopic image. Although it is difficult to perform a quantitative measurement to prove that this was indeed the case, all monoscopic images were visually evaluated and only low noise, high quality images were accepted as case samples for the observer experiment.

We adjusted the display monitor with a photometer to meet the DICOM grayscale standards. We did not attempt to take into account the attenuation by the LCD glasses in the adjustment because there are no DICOM standards for setting up a stereo display at present. The LCD glasses do degrade the perceived image quality to some extent, such as a reduction in brightness and an increase in noise. However, since the degradation would have a negative impact on stereoscopic reading, one may expect that the advantages of stereoscopic reading would be even greater than those observed in our study if the degradation could be compensated for or if better stereoscopic viewing methods (e.g., higher transmission stereo glasses or no glasses) become available in the future.

One of the expectations for developing 3D imaging techniques such as stereomammography for screening is to reduce recalls. In conventional mammography, many recalls are caused by superimposition of dense tissue mimicking masses and insignificant calcifications appearing to be clustered due to a lack of 3D spatial distribution information. In our study, analysis of the BI-RADS assessments indicates

that the increase in detection sensitivity is accompanied by a slight increase in recalls. It is not known how the reading of specimen radiographs in a laboratory experiment would translate to clinical applications. However, the observed improvements in the ROC curves indicate that there were true improvements in the performances of the radiologists with additional stereoscopic reading and that the radiologists did not simply relax the decision thresholds along their original ROC curves, which would also result in an increase in sensitivity and a decrease in specificity. The improvements in the ROC curves show the promise that, if the radiologists become more experienced with stereomammography and more confident in utilizing the additional 3D information for assessing the lesions, they may be able to adjust their decision thresholds along the resulting higher ROC curves such that the sensitivity will be gained without a tradeoff, or even with an increase, in specificity in comparison to their decisions along the lower ROC curves associated with monoscopic reading alone. Further studies will be needed to investigate if this can be realized and thus lead to a reduction in recalls.

One limitation of stereoscopic viewing is that human eyes vary in their stereo acuity, although it is believed that stereo acuity may improve with training. The radiologists participated in this study were impressed by the 3D appearance of the stereoscopic images. The image quality of our stereo display workstation is excellent without perceivable flicker. However, some of the radiologists still experienced eye fatigue if the reading time was long. These problems may be alleviated with a different display method or viewing electronics as well as improved reader ergonomic factors.

V. CONCLUSION

We have performed an observer performance study using ROC methodology to evaluate the improvement in mammographic lesion detection and characterization by stereoscopic reading. Our results indicated that statistically significant (two-tailed $p < 0.05$) improvements were achieved for detection of microcalcifications and for classification of malignant and benign lesions. The detection of masses was also improved but the improvement fell short of statistical significance. This study demonstrates the potential of using stereomammography to improve the detection and characterization of mammographic lesions.

ACKNOWLEDGMENTS

This work was supported by U.S. Army Medical Research and Materiel Command Grant Nos. DAMD17-98-1-8210, DAMD17-02-1-0214, and DAMD17-99-1-9294. The content of this publication does not necessarily reflect the position of the funding agency, and no official endorsement of any equipment and product of any companies mentioned in this publication should be inferred. The authors are grateful to Dr. Charles E. Metz for the LABMRC program.

^{a)}Electronic mail: chanhp@umich.edu

¹V. P. Jackson, R. E. Hendrick, S. A. Feig, and D. B. Kopans, "Imaging of

- the radiographically dense breast," *Radiology* **188**, 297–301 (1993).
- ²R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**, 613–617 (1992).
- ³M. G. Wallis, M. T. Walsh, and J. R. Lee, "A review of false negative mammography in a symptomatic population," *Clin. Radiol.* **44**, 13–15 (1991).
- ⁴H. P. Chan, M. M. Goodsitt, J. M. Sullivan, K. L. Darner, and L. M. Hadjiiski, "Depth perception in digital stereoscopic mammography," Atlanta, GA, 8–12 June, 2000.
- ⁵H. P. Chan, M. M. Goodsitt, K. L. Darner, J. M. Sullivan, L. M. Hadjiiski, N. Petrick, and B. Sahiner, "Effects of stereoscopic imaging technique on depth discrimination," *The Fifth International Workshop on Digital Mammography*, Toronto, Canada, 11–14 June, 2000 (Medical Physics, Madison, WI), pp. 13–18.
- ⁶M. M. Goodsitt, H. P. Chan, and L. M. Hadjiiski, "Stereomammography: Evaluation of depth perception using a virtual 3D cursor," *Med. Phys.* **27**, 1305–1310 (2000).
- ⁷M. M. Goodsitt, H. P. Chan, J. M. Sullivan, K. L. Darner, and L. M. Hadjiiski, "Evaluation of the effect of virtual cursor shape on depth measurements in digital stereomammograms," *The Fifth International Workshop on Digital Mammography*, Toronto, Canada, 11–14 June, 2000 (Medical Physics, Madison, WI), pp. 45–50.
- ⁸D. J. Getty, R. M. Pickett, and C. J. D'Orsi, *Stereoscopic Digital Mammography: Improving Detection and Diagnosis of Breast Cancer*, Berlin, 27–30 June, 2001, International Congress Series Vol. 1230 (Elsevier, Amsterdam), pp. 506–511.
- ⁹H. P. Chan, M. M. Goodsitt, L. Hadjiiski, M. A. Roubidoux, J. E. Bailey, M. A. Helvie, J. T. Lydick, and B. Sahiner, "ROC study comparing radiologists' performances in evaluating breast lesions on stereoscopic and single-projection digital specimen mammograms," *Med. Phys.* **30**, 1456 (abstract) (2003).
- ¹⁰H. P. Chan, M. M. Goodsitt, L. M. Hadjiiski, J. E. Bailey, K. Klein, K. L. Darner, and B. Sahiner, "Effects of magnification and zooming on depth perception in digital stereomammography: An observer performance study," *Phys. Med. Biol.* **48**, 3721–3734 (2003).
- ¹¹M. M. Goodsitt, H. P. Chan, K. L. Darner, and L. M. Hadjiiski, "The effects of stereo shift angle, geometric magnification and display zoom on depth measurements in digital stereomammography," *Med. Phys.* **29**, 2725–2734 (2002).
- ¹²A. D. A. Maidment, P. R. Bakic, and M. Albert, "Effects of quantum noise and binocular summation on dose requirements in stereoradiography," *Med. Phys.* **30**, 3061–3071 (2003).
- ¹³L. T. Niklason, B. T. Christian, L. E. Niklason, D. B. Kopans, D. E. Castleberry, B. H. Opsahl-Ong, C. E. Landberg, P. J. Slanetz *et al.*, "Digital tomosynthesis in breast imaging," *Radiology* **205**, 399–406 (1997).
- ¹⁴R. L. Webber, H. R. Underhill, and R. I. Freimanis, "A controlled evaluation of tuned-aperture computed tomography applied to digital spot mammography," *J. Digit Imaging* **13**, 90–97 (2000).
- ¹⁵S. Suryanarayanan, A. Karellas, S. Vedantham, S. P. Baker, S. J. Glick, C. J. D'Orsi, and R. L. Webber, "Evaluation of linear and nonlinear tomosynthetic reconstruction methods in digital mammography," *Acad. Radiol.* **8**, 219–224 (2001).
- ¹⁶V. Raptopoulos, J. K. Baum, M. Hochman, A. Karellas, M. J. Houlihan, and C. J. D'Orsi, "High resolution CT mammography of surgical biopsy specimens," *J. Comput. Assist. Tomogr.* **20**, 179–184 (1996).
- ¹⁷J. M. Boone, T. R. Nelson, and J. A. Seibert, "The potential for breast CT," *Med. Phys.* **28**, 1246 (abstract) (2001).
- ¹⁸J. M. Boone, T. R. Nelson, K. K. Lindfors, and J. A. Seibert, "Dedicated breast CT: Radiation dose and image quality evaluation," *Radiology* **221**, 657–667 (2001).
- ¹⁹E. A. Rafferty, D. Georgian-Smith, D. B. Kopans, D. A. Hall, R. Moore, and T. Wu, "Comparison of full-field digital tomosynthesis with two view conventional film screen mammography in the prediction of lesion malignancy," *Radiology* **225(P)**, 268 (2002).
- ²⁰K. Doi, N. J. Patronas, E. E. Duda, E. Geldner, and K. Dietz, "X-ray imaging of blood vessels to the brain by use of magnification stereoscopic technique," *Adv. Neurol.* **30**, 175–189 (1981).
- ²¹C. A. Kelsey, R. D. Moseley, S. A. Mettler, and D. E. Briscoe, "Cost-effectiveness of stereoscopic radiographs in detection of lung nodules," *Radiology* **142**, 611–613 (1982).
- ²²K. Doi and E. E. Duda, "Detectability of depth information by use of magnification stereoscopic technique in cerebral angiography," *Radiology* **146**, 91–95 (1983).
- ²³Y. Higashida, Y. Hirata, R. Saito, S. Doudanuki, H. Bussaka, and M. Takahashi, "Depth determination on stereoscopic digital subtraction angiograms," *Radiology* **168**, 560–562 (1988).
- ²⁴M. C. Trocme, A. H. Sather, and K. N. An, "A biplanar cephalometric stereoradiography technique," *Am. J. Orthod. Dentofacial Orthop.* **98**, 168–175 (1990).
- ²⁵J. I. Ragnarsson and J. Karrholm, "Factors influencing postoperative movement in displaced femoral neck fractures: evaluation by conventional radiography and stereoradiography," *J. Orthop. Trauma* **6**, 152–158 (1992).
- ²⁶T. S. Curry, J. E. Dowdey, and R. C. Murry, *Christensen's Physics of Diagnostic Radiology*, 4th ed. (Lea & Febiger, Philadelphia, PA, 1992).
- ²⁷A. D. A. Maidment, M. Albert, E. F. Conant, C. W. Piccoli, and P. A. McCue, "Prototype workstation for 3-D diagnosis of breast calcifications," *Radiology* **201(P)**, 556 (1996).
- ²⁸E. F. Conant, A. D. Maidment, M. Albert, C. W. Piccoli, S. A. Nussbaum, and P. A. McCue, "Small field-of-view digital imaging of breast calcifications: method to improve diagnostic specificity," *Radiology* **201(P)**, 369 (1996).
- ²⁹*American College of Radiology Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas)* (American College of Radiology, Reston, VA, 2003).
- ³⁰D. D. Dorfman, K. S. Berbaum, and C. E. Metz, "ROC rating analysis: Generalization to the population of readers and cases with the jackknife method," *Radiology* **27**, 723–731 (1992).