

An observer study comparing spot imaging regions selected by radiologists and a computer for an automated stereo spot mammography technique

Mitchell M. Goodsitt,^{a)} Heang-Ping Chan, Justin T. Lydick, Chaitanya R. Gandra, Nelson G. Chen, Mark A. Helvie, Janet E. Bailey, Marilyn A. Roubidoux, Chintana Paramagul, Caroline E. Blane, Berkman Sahiner, and Nicholas A. Petrick^{b)}
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109-0030

(Received 22 October 2003; revised 16 February 2004; accepted for publication 15 March 2004; published 27 May 2004)

We are developing an automated stereo spot mammography technique for improved imaging of suspicious dense regions within digital mammograms. The technique entails the acquisition of a full-field digital mammogram, automated detection of a suspicious dense region within that mammogram by a computer aided detection (CAD) program, and acquisition of a stereo pair of images with automated collimation to the suspicious region. The latter stereo spot image is obtained within seconds of the original full-field mammogram, without releasing the compression paddle. The spot image is viewed on a stereo video display. A critical element of this technique is the automated detection of suspicious regions for spot imaging. We performed an observer study to compare the suspicious regions selected by radiologists with those selected by a CAD program developed at the University of Michigan. True regions of interest (TROIs) were separately determined by one of the radiologists who reviewed the original mammograms, biopsy images, and histology results. We compared the radiologist and computer-selected regions of interest (ROIs) to the TROIs. Both the radiologists and the computer were allowed to select up to 3 regions in each of 200 images (mixture of 100 CC and 100 MLO views). We computed overlap indices (the overlap index is defined as the ratio of the area of intersection to the area of interest) to quantify the agreement between the selected regions in each image. The averages of the largest overlap indices per image for the 5 radiologist-to-computer comparisons were directly related to the average number of regions per image traced by the radiologists (about 50% for 1 region/image, 84% for 2 regions/image and 96% for 3 regions/image). The average of the overlap indices with all of the TROIs was 73% for CAD and 76.8% \pm 10.0% for the radiologists. This study indicates that the CAD determined ROIs could potentially be useful for a screening technique that includes stereo spot mammography imaging. © 2004 American Association of Physicists in Medicine. [DOI: 10.1118/1.1737492]

Key words: stereomammography, spot mammography, observer study

I. INTRODUCTION

It has been estimated that early detection could save the lives of about 30–50% of women who develop breast cancer.^{1,2} X-ray mammography is the most powerful screening tool we have for detecting breast cancer. However, it has limitations, especially in imaging breasts containing large dense tissue regions.^{3–5} Optimal perceptibility in these regions may not be attained even with the new full-field digital mammographic systems due to the higher noise levels associated with these poorly penetrated regions and due to the inability to separate overlapping tissues. Alternative views and spot mammography are known to be beneficial in these situations. For example, Hayes *et al.* have reported that magnification and spot compression techniques improved mammographic specificity in 50.8% of the screening cases that were recalled for assessment at their center.⁶ Specifically, sixteen ‘equivocal’ diagnoses became ‘normal’ or ‘benign’ and 15 of these patients avoided surgical biopsy. Twelve ‘equivocal’ diagnoses became ‘malignant,’ which helped surgical planning, and in all 12 cases, histology confirmed the diagnosis of malignancy.” Hayes *et al.* concluded that special views are

“necessary for the complete mammographic assessment of many screen-detected abnormalities.” Furthermore, Faulk and Sickles found, in a study of the efficacy of spot compression-magnification and tangential views in the mammographic evaluation of palpable breast masses, that spot compression–magnification views depicted 97% of the masses; whereas, standard views depicted 87%.⁷ Tangential views also depicted 87% of the masses with some of the masses that were detected in tangential views not being detected in the standard views and *vice versa*. In addition, they found that use of special views enabled radiologists to correctly predict benign or malignant status in 77% of cases as compared with correct prediction in 69% of cases with only standard views. It is important to note that the supplemental imaging techniques like spot and tangential views are only employed in diagnostic work-up studies. They are not employed at screening.

Spot imaging is performed either in contact or magnification mode. The radiologist first examines a contact mammogram, and identifies a suspicious region for spot imaging. The technologist then utilizes the identified location on the

mammogram to reposition the breast to her best ability such that the suspicious region is in the center of the field. She then compresses this region using a special, smaller spot compression paddle. The x-ray beam is also collimated to a smaller field size. The goal is to mechanically separate a suspicious lesion from the surrounding tissue for improved perception in the spot mammogram. One problem with this method is that the positioning of the breast involves some guesswork so the desired lesion may not necessarily be imaged at all, or it may not be imaged optimally.

We have been developing an automated version of the spot imaging technique. Our initial idea was to employ a computer aided detection (CAD) program to determine the locations of suspicious dense regions within a full-field digital mammogram, and to take a second digital mammogram of only those regions using automated collimation and automated spot compression along with a more penetrating exposure. This second separate "spot mammogram" would be taken within seconds of the full-field mammogram while the breast is maintained in the same position, but compressed more to improve tissue separation. We developed instrumentation to implement this method⁸ including (1) a stretched Mylar membrane device that is placed between the large conventional paddle and the breast and acts to restrain the breast during the changeover to the smaller spot compression paddle, (2) an x - y translator for positioning the spot paddle at the suspicious region, and (3) a secondary collimator that restricts the x-ray beam to the suspicious region area.⁸ We also performed experiments with a compressible breast phantom that contained simulated masses. We found that when we spot-compressed a particular simulated mass in the phantom, it moved laterally out from underneath a dense overlying region and became visible in the spot mammogram. However, we were unable to reproduce the same result when we repeated the experiment multiple times. This was disappointing, but it made us aware that it can be difficult to position the spot paddle to produce the desired shearing force on a lesion, and even when the force is in the right direction, it may not be sufficient to move the lesion far enough out from overlying or underlying dense tissue to be seen. This prompted us to think of an alternative spot compression method—one that would be easier to implement, involve less equipment modification and produce better distinction between overlapping lesions.

The new technique is stereo spot digital mammography. It is very similar to the above technique except there is no changeover to or positioning of a spot paddle and no need for a breast restraining device to hold the breast in the same position while switching to the spot paddle. Instead, after the suspicious region is identified in the full-field image by the CAD program, a stereo pair of images is immediately acquired with the x-ray beam automatically collimated to image that region. The additional equipment required is a secondary collimator to restrict the x-ray beam to the desired region and a stereo workstation that would be located in the radiologist reading room. The x-ray tube or the focal spot would also have to shift by about 3 to 6 degrees in the left and right directions to generate the left- and right-eye images

of the stereo pair. Stereoscopic image acquisition and display will enable radiologists to view the suspicious regions in three dimensions. This will reduce the tissue superposition problem inherent in conventional single projection mammography. Furthermore, in comparison with conventional spot compression, the automated technique should produce more accurate spot imaging of suspicious regions because it eliminates the need for the repositioning of the breast between the full-field and spot images, and the "spot" location is determined by a computer analysis of the digital full-breast image rather than estimated by eye from a radiograph. Since the method is automated and does not require the on-line review of a radiologist to determine the locations of the suspicious regions, it could potentially be used in screening and could potentially eliminate the need for diagnostic callback studies in many cases.

Another approach would be to perform full-field stereomammography instead of stereo spot mammography. This could be carried out either as a replacement for the conventional full-field mammogram or as a supplement to that mammogram. The advantages of limiting the stereo imaging to a spot region as opposed to full-field are that spot mode limits the amount of breast tissue exposed to additional radiation, and it decreases the volume of tissue that scatters radiation thereby improving image contrast. The anticipated dose for stereo spot mammography would be considerably less than that of a full-field mammogram because of the smaller field size. The exact dose for the spot technique would depend upon the size of the spot collimated region, the x-ray technique factors (kVp, target, filter, mAs), the x-ray beam half-value-layer, the tissue thickness, the breast composition (amount of glandular tissue) in the path of the spot-collimated x-ray beam, and the amount of glandular tissue that is exposed to scattered x-rays as well as the amount of the x-ray scatter. Stereo spot mammography does involve the acquisition of 2 spot images (the left- and right-eye images), instead of 1, so one might think that it would therefore require twice the dose of a single spot image. However, the eye-brain system integrates the noise from both images when they are viewed as a stereo pair, so theoretically, the same signal-to-noise ratio could be achieved by using only about half the dose for each image of the stereo pair. This is confirmed by the results of an experimental study recently published by Maidment *et al.*⁹ on the effects of quantum noise and binocular summation on dose requirements for full-field stereoradiography. Maidment *et al.* found "the total dose needed to produce a stereoradiographic image pair is approximately 1.1 times the dose needed for a single projection in standard radiography."

Automated detection of suspicious dense regions is a crucial element of the stereo spot mammography method. In this paper, we describe an observer study that was performed to compare the suspicious regions selected by radiologists with those selected by a CAD program developed in our laboratory at the University of Michigan. We also compared these to true regions of interest (TROIs) containing the masses. At the time of our study, a substantial collection of full-field digital mammography (FFDM) images with biopsy results

was not available, and our CAD algorithms had not yet been adapted to FFDM images. Therefore, we employed digitized film images and our film-based CAD algorithms. Although the images and the CAD system are not completely representative of the images and analysis that will be employed in the eventual implementation of the automated stereo spot method, they provide valuable preliminary data about a new application of CAD, namely automated detection of ROIs to be worked-up with spot imaging.

II. MATERIALS AND METHODS

IRB approval was obtained to review 200 digitized mammograms for this study.

Five MQSA qualified radiologists participated as the readers. Their experience in reading mammograms ranged from 5.5 to 25 years (mean=13.7 years, standard deviation=7.6 years). The radiologists viewed digitized film mammograms that have been employed previously in the development of our CAD system. The film set included craniocaudal (CC) view and the mediolateral oblique (MLO) view mammograms of both breasts of patients at our clinics. The mammograms were digitized with a LUMISYS 85 laser film scanner (Lumisys, Inc., Sunnyvale, CA) at a pixel size of $50\ \mu\text{m} \times 50\ \mu\text{m}$. This digitizer has a gray level resolution of 12 bits and a nominal optical density (O.D.) range of 0 to 4. To keep the reading time reasonable (i.e., about 3 hours), we had the observers perform the study on 200 mammograms. All of the observers also repeated the study 3 to 13 months later for an evaluation of their reproducibility.

Each observer was given the task of outlining in each mammogram, 0 to 3 suspicious regions which in an ideal world they would have spot-imaged. A computer program, TRACEIMAGE, was developed for this study to allow the radiologists to trace the boundaries of the suspicious regions using a computer mouse. The TRACEIMAGE program incorporates a graphical user interface (GUI) that displays the digitized mammogram within a window on the computer monitor. The GUI includes slide bars for adjusting the contrast and brightness of the image, a display of a histogram of the grayscale values within the breast region in the image, a pull-down menu for selecting the pixel size of the displayed image (the choices are 200, 400 and 800 microns), slide bars for panning the image when the selected pixel size is 200 microns, a trace routine that permits the operator to outline the suspicious regions within the displayed mammogram via a series of computer mouse motions and mouse clicks, a button for erasing an individual trace within the image, and another button for erasing all traces within the image. The GUI was designed to be very user friendly. It automatically finishes a particular trace when a mouse click is within a preset number of pixels from the beginning point of the trace, and it then automatically proceeds to the next trace. Also, the user can re-adjust any of the traces in an image. Examples of the GUI display are shown in Figs. 1 and 2. The TRACEIMAGE program with its GUI was implemented on a PC computer with a high quality Hitachi (Hitachi, Ltd., Japan) Superscan 753 19 in. color monitor operating at a reso-

lution of 1024×768 in a noninterlaced true color mode with an 85 Hz refresh rate. The monitor was adjusted to meet DICOM standards, and the room lights were dimmed to a very low level during the observer studies. During the study, each observer analyzed each image separately without referring to previous or subsequent images. They were not told which MLO views corresponded with which CC views and *vice versa*. Furthermore, all patient identification information was removed from the images.

For the computer-selected regions, we employed a mass detection CAD program developed at the University of Michigan.¹⁰ This program consists of 4 steps. First, the digitized mammogram is processed with a density weighted contrast enhancement (DWCE) filter that adaptively enhances local area contrast in order to emphasize mammographic structures. Second, an edge detection algorithm is employed to define the borders of the enhanced structures, resulting in a set of detected structures. Third, a local refinement algorithm, which includes erosion and K-means clustering, is applied to the detected structures to improve the accuracy of the borders and to split large connected regions. Fourth, the refined detected objects are classified as masses or normal breast structures based on the input of extracted morphological and textural features into a linear discriminant classifier. Potential masses are identified using decision thresholds that are based on the linear discriminant classifier score and the maximum number of marks allowed per image. For the present study, we adjusted the detection threshold of the CAD program¹⁰ to mark between 0 and 3 regions in each image, with 3 the most likely number. In addition, rectangular bounding boxes that enclosed each known true mass in the set of mammograms were also obtained. These were regions identified by a radiologist from analyses of the mammograms along with associated pathology biopsy data and biopsy images. Of the 200 images that were evaluated by the radiologists in our observer study, 98 images contained TROIs. There were 83 images with a single true mass, 13 with two true masses, and 2 with three true masses. Thus, there was a total of 115 ($=83+13 \times 2+2 \times 3$) TROIs. Out of the 98 images with TROIs, 51 (52%) contained malignant masses. 75% of the exams with TROIs were worked-up with spot compression (57% of these were malignant and 43% were benign). It should be noted that our data set of 200 images was for 37 patients, and there were images from 2 or more years for 11 of the patients. Some of these patients had "normal" mammograms in earlier years and suspicious regions worked-up with spot mammography and/or lateral views in later years.

A second computer program, DENSECOMP, was developed to display sets of traced regions in each image. The traces for up to 3 readers can be displayed at once using different colors. For example, the traces for one radiologist would be filled-in as red, those for a second radiologist would be filled-in as green and those for a third radiologist would be filled-in as blue. Anywhere in the image where all three of the readers' traces intersected would be displayed as white. For the study described in this paper, we were interested in the intersections between the radiologist-selected regions of

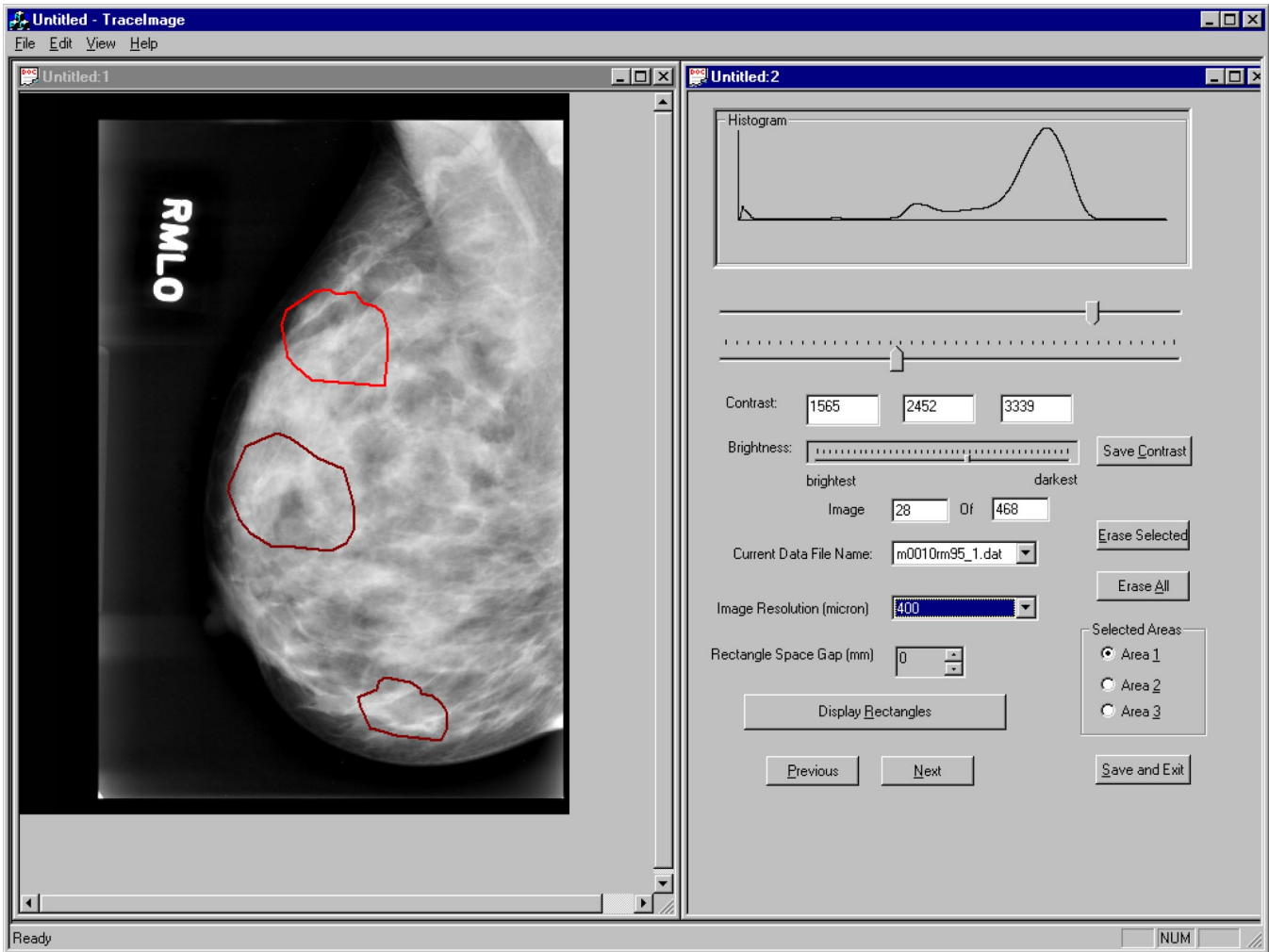


FIG. 1. Graphical user interface display showing one of the digitized mammograms in this study, 3 regions for spot imaging that were traced by a radiologist, a histogram of the pixel values within the breast region (top right), sliders (beneath histogram) for adjusting the range of pixel values that are mapped into an 8-bit output for display, and a pull-down menu for selecting the image resolution (the resolution displayed is 400 microns).

interest (ROIs) and those selected by the CAD program (for simplicity, the ROIs selected by the computer program will be referred to as CAD ROIs or CAD-selected ROIs in the following discussion) and also the intersections of both of these with the true mass regions. We found that the large number of possible intersections of the regions for 3 readers (e.g., 27 possible single intersections) can result in a complicated display that is difficult to interpret. We therefore decided to display the results for 2 readers at a time where a “reader” is either a radiologist, the CAD program or the true mass region. In addition to filling-in the ROIs with color, the program can also display the traces themselves superimposed on the mammogram in different colors. This allows one to see the suspicious lesion within the trace.

The DENSECOMP computer program computes overlap indices between the radiologist-selected ROIs and the CAD-selected ROIs. Furthermore, it computes overlap indices between the radiologists-selected ROIs and the TROIs and overlap indices between the CAD-selected ROIs and the TROIs. The overlap index for the radiologist vs CAD program comparison is defined as the area of the intersection

between the radiologist-selected and CAD-selected ROIs divided by the area of the CAD ROI. This definition was chosen so that a value of 100% would be obtained if the CAD ROI was completely contained within the radiologist’s ROI, indicating that the CAD ROI would be completely imaged with a spot region determined by the radiologist. Examples of 100% overlap are shown in Figs. 3(a) and 3(b). The overlap index between the radiologist’s or CAD ROI and the TROI was defined as the area of intersection divided by the smaller area. This definition yields an overlap index of 100% if the radiologist’s ROI or CAD ROI is completely contained within the TROI or if the TROI is completely contained within the CAD or radiologist’s ROI. It is an indication of the degree to which there is a “hit” between the radiologist’s or CAD ROI and the TROI. In addition, we computed the number of “hits” by determining the total number of times that the overlap indices with the TROIs were greater than or equal to a threshold value of 25%. The 25% value was considered to be a reasonable threshold for indicating agreement in ROIs. For the computation of the overlap indices between radiologist’s and CAD ROIs, all possible pairings of the se-

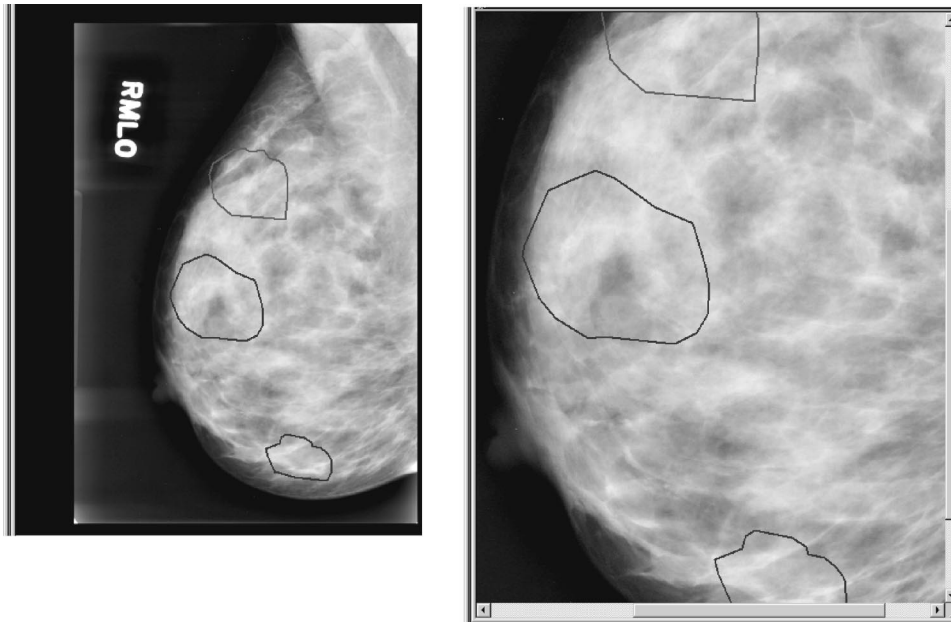


FIG. 2. An example of the GUI display of the same image at 400 micron (left) and 200 micron (right) resolution. Radiologists could use either display resolution for each image in the study, and could switch between the resolutions during their analysis. Many of the 200-micron resolution images were larger than the display window. The GUI included a panning feature to enable shifting of the viewed portion of the 200-micron image within the display window so the entire image could be viewed and analyzed.

lected regions were considered, and the indices were ordered from the largest to the smallest. For example, if the radiologist selected 3 regions and the CAD program selected 3 regions, there would be 9 possible intersections for the indices and the results would be listed in 9 columns with column 1 corresponding to the largest index. The radiologist's ROI-vs-TROI and CAD ROI-vs-TROI comparisons were special cases because we knew the true mass regions. For these particular comparisons, we computed the largest overlap indices individually for each TROI. For images with two or three TROIs, the largest overlap index for all of the TROIs was assigned to TROI #1, the largest overlap index for the remaining one or two TROIs was assigned to TROI #2, and if there was a third TROI, the largest overlap index for that TROI was assigned to TROI #3.

III. RESULTS

Comparisons of the ROIs selected by radiologists and the CAD program are shown in Fig. 3. In each image, the ROIs determined by the radiologist are filled-in in red, and those of

the CAD program are filled-in in green. The intersections are displayed as yellow (red + green = yellow.) A comparison of the ROIs selected by each of the 5 radiologist readers [(a)–(e)] and the TROIs for an image containing 3 TROIs is shown in Fig. 4. The CAD selected ROIs are also compared with the TROIs in part (f) of this figure. Note that instead of filling-in the ROIs with color as in Fig. 3, the actual borders are displayed in black for the radiologist and CAD ROIs and in white for the TROIs in Fig. 4. The frequencies that each radiologist selected 0, 1, 2, and 3 ROIs in the 200 images at each reading session are listed in Table I, along with the frequencies for the CAD program.

The average values of the largest overlap indices between the radiologist-selected ROIs and the CAD-selected ROIs are listed for each radiologist and each reading session in Table II. A histogram summarizing the individual results for all of the radiologist vs CAD ROI pairings having the largest overlap index is shown in Fig. 5. The overall average overlap index is 69.6% with a standard deviation of 44.3%.

The average overlap indices between the TROIs and the

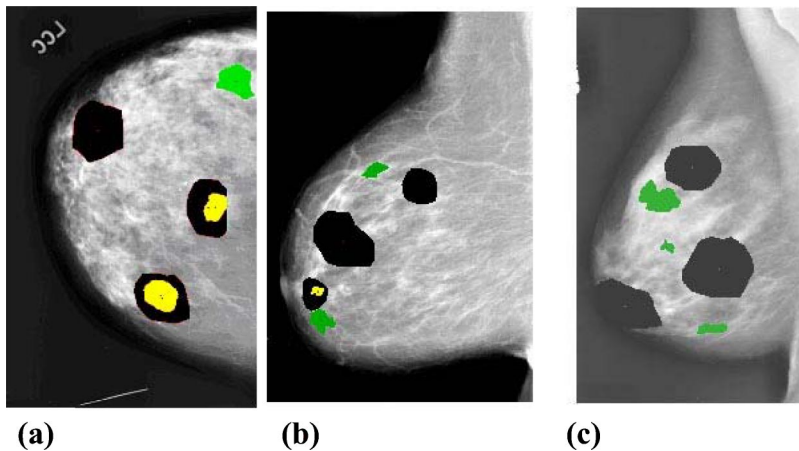


FIG. 3. Examples of the radiologist and computer selected ROIs. The radiologist ROIs are filled-in in red (black in figure), the CAD selected ROIs are filled-in in green (grey in figure), and the intersection areas are displayed in yellow (white in figure). In (a) the radiologist and computer agreed on 2 of the 3 ROIs; in (b) they agreed on one ROI; and in (c) they disagreed on all 3 ROIs.

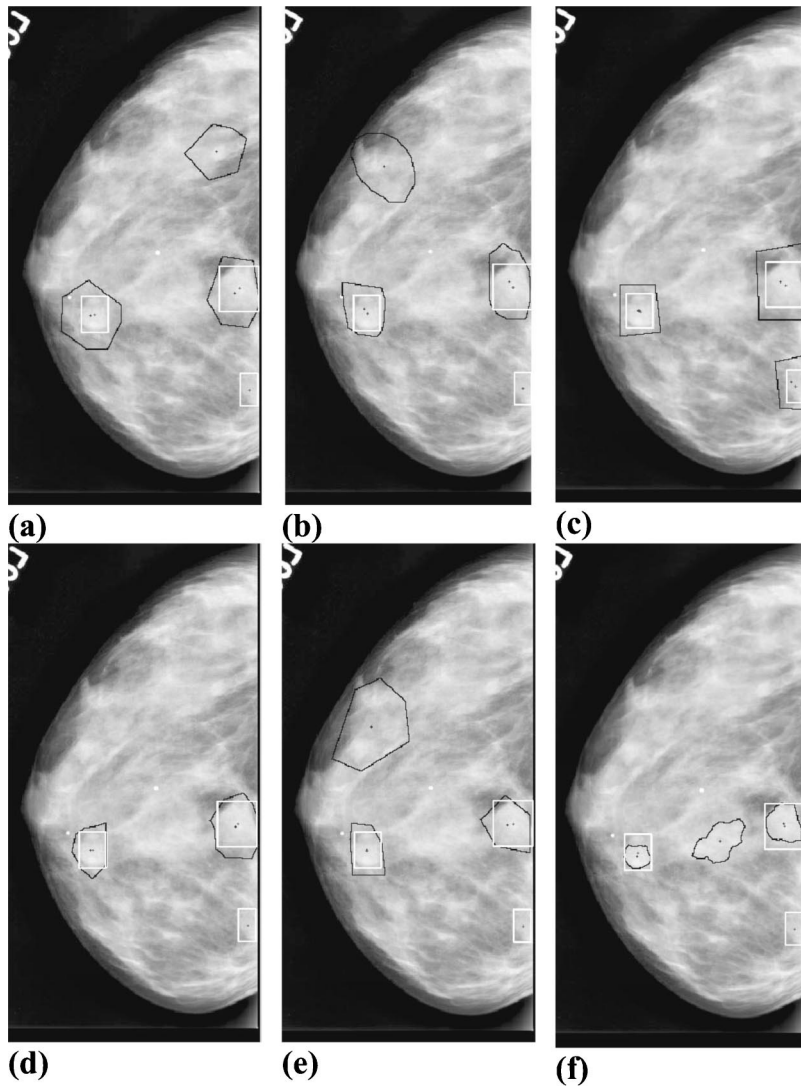


FIG. 4. Comparisons of the ROIs selected by the radiologists with the true mass regions (a)–(e) and the ROIs selected by the CAD program with true mass regions (f). The boundaries of the radiologist and CAD determined regions are traced in black, and the boundaries of the true regions are traced in white. Note that for this particular mammogram, radiologist (d) chose to select 2 regions instead of 3. Also, radiologist (c) was the only one whose selected regions intersected all 3 TROIs.

ROIs selected by the radiologists and the CAD program are listed in Table III. The average percentages of hits between each reader's selected ROIs and the TROIs are listed in Table IV.

IV. DISCUSSION

As shown in Table I, the radiologists executed the task of selecting up to 3 spot regions in each image differently, with some radiologists selecting almost 3 ROIs in each image, and others selecting far fewer with average numbers of ROIs per image even less than 1. The CAD program, as designed, selected 3 ROIs in nearly all (93%) of the images.

For the 5 radiologist-vs-CAD comparisons, the average agreement between the radiologist-selected ROIs for spot imaging and at least one of the CAD-selected ROIs for the entire set of images ranges from 43% to 98% (Table II). This unusually wide range of agreement can be explained if one compares the overlap indices listed in Table II with the average number of ROIs per image listed in Table I. A linear least squares fit between these data yields a correlation coefficient of 0.99 with a positive slope. That is, the greater the

average number of ROIs per image selected by a radiologist, the greater the agreement with the computer. In general, the agreement or average overlap index between the radiologist-selected ROIs and at least one of the CAD-selected ROIs in each image was only about 50% for radiologists who traced an average of about 1 ROI per image. On the other hand, this agreement was much improved to 84% for radiologists who traced an average of about 2 ROIs per image, and it was about 96% for those who traced about 3 ROIs per image. Thus, it is anticipated that had we not given the radiologists full freedom to trace between 0 and 3 ROIs in each digitized mammogram for spot imaging, as in our original study design, but had we instead instructed them to trace 3 ROIs per mammogram similar to the number chosen for the computer, all of the radiologists would have had excellent agreement with the computer. Yet another factor that was not controlled in this observer experiment that could influence the agreement is the sizes of the radiologist selected ROIs, with greater sizes having greater probabilities of overlap with the CAD ROIs. The reproducibility of the overlap indices between the radiologist and CAD ROIs (Table II) for the two

TABLE I. Percentages of the 200 images in which each radiologist selected 0, 1, 2, and 3 ROIs during each reading session. Also, the corresponding percentages for the CAD program.

Reader	Reading session	% with 0 ROIs	% with 1 ROI	% with 2 ROIs	% with 3 ROIs	Average number of ROIs per image	Standard deviation
Radiologist a	1	1	0	16.5	87.5	2.86	0.43
Radiologist a	2	0	0.5	5.5	94	2.94	0.27
Radiologist b	1	1.5	12.5	38	48	2.32	0.75
Radiologist b	2	0.5	4.5	37	58	2.50	0.61
Radiologist c	1	31.5	53.5	13.5	1.5	0.85	0.70
Radiologist c	2	41	47.5	10	1.5	0.72	0.70
Radiologist d	1	39.5	50.5	10	0	0.70	0.64
Radiologist d	2	46.5	43	9	1.5	0.66	0.71
Radiologist e	1	2.5	23	40	34.5	2.06	0.82
Radiologist e	2	16.5	56	18	9.5	1.20	0.83
CAD program		0	2	5	93	2.91	0.35

TABLE II. Overlap indices between the ROIs selected by the radiologists and by the CAD computer program. The averages of the largest overlap indices for each image are listed for each radiologist in each reading session. These are averages for 200 images and up to 3 ROIs per image. The overlap index is defined to be the area of the intersection between the radiologist-selected and CAD-selected ROIs divided by the area of the CAD-selected ROI in percent.

	Radiologist a	Radiologist b	Radiologist c	Radiologist d	Radiologist e
Reading #1	94%	84%	51%	45%	82%
Reading #2	98%	86%	47%	43%	65%

TABLE III. Overlap indices between the ROIs selected by the readers and the true regions of interest (TROIs). The averages of the largest overlap indices for each TROI in each image are listed, as well as the averages for all TROIs. The overlap index in this case is defined to be the area of the intersection divided by the smaller of the TROI and reader ROI areas, in percent. There were 83 images with one TROI, 13 with two TROIs and 2 with three TROIs. Wilcoxon Signed Rank Test *p*-values relative to CAD for the All TROI results of each radiologist at each reading session are listed in the final column.

	TROI #1	TROI #2	TROI #3	All TROI	<i>p</i> -value
Radiologist a					
Reading #1	88%	58%	50%	84% ^a	0.003
Reading #2	94%	74%	46%	90% ^a	<0.0001
Radiologist b					
Reading #1	88%	70%	49%	85% ^a	0.008
Reading #2	88%	66%	46%	84% ^a	0.010
Radiologist c					
Reading #1	70%	42%	47%	66%	0.399
Reading #2	71%	44%	44%	67%	0.680
Radiologist d					
Reading #1	69%	45%	0%	65%	0.272
Reading #2	66%	48%	50%	64%	0.179
Radiologist e					
Reading #1	87%	50%	37%	81%	0.118
Reading #2	85%	67%	47%	82% ^a	0.018
Computer (CAD)	78%	53%	0%	73%	

^a=significantly different from the CAD program at a 0.05 level or less.

TABLE IV. Average percentages of “hits.” A hit is defined to occur whenever the overlap index between the reader ROI and the TROI is greater than or equal to 25%. Values relative to the total number of masses (TROIs) in the data set in percent are listed for each reading session. Wilcoxon Signed Rank Test *p*-values relative to CAD for each radiologist and reading session are also listed.

	Reading #1	<i>p</i> -value	Reading #2	<i>p</i> -value
Radiologist a	89%	0.074	93% ^a	0.006
Radiologist b	90% ^a	0.036	89%	0.062
Radiologist c	70%	0.081	70%	0.081
Radiologist d	69%	0.055	67% ^a	0.027
Radiologist e	87%	0.153	87%	0.135
CAD	80%			

^a=significantly different from the CAD program at a 0.05 level or less.

reading sessions is very good (within 4% for 3 radiologists and 17% for one) indicating each radiologist was very consistent in identifying suspicious ROIs for spot imaging at each session.

The agreement between the radiologist-selected ROIs and the true mass regions (TROIs) is very similar on average to that between the CAD-selected ROIs and the TROIs. For the entire set of 115 TROIs (Table III, column 4), the overall average overlap indices for the radiologists' ROI-vs-TROI comparisons ranged from 64% to 90% with a mean of 76.8% \pm 10.0%, and the overall average overlap index for the CAD ROI-vs-TROI comparisons was 73%. As listed in Table III, 5 of the 10 radiologists' ROI-vs-TROI overlap indices were found to be statistically significantly different from the corresponding CAD ROI-vs-TROI overlap indices. Nonparametric statistics were employed for this comparison because, as can be deduced from Fig. 5, the data were not from a normal distribution. The Wilcoxon signed rank test was utilized. It is the nonparametric equivalent of the paired *t*-test. We found a very similar statistical result when we compared the ROI-vs-TROI overlap indices between pairs of radiologists. For example, for the 10 possible pairings of radiologists' results for the first reading sessions (the combination of 5 radiologists taken 2 at a time), we found 6 pairings were statistically significantly different. Thus, in terms of both the mean overlap indices of ROIs with TROIs and the statistical significance of the overlap indices, the CAD program performed comparably with an average radiologist.

The percentages of “hits” for which the overlap indices were greater than or equal to 25% for all TROIs ranged from 67% to 93% with a mean of 80.9% \pm 10.6% for the radiologists' ROI-vs-TROI comparisons and was 80% for the CAD ROI-vs-TROI comparisons (Table IV). As shown in Table IV, 3 of the 10 (5 radiologists \times 2 reading sessions) numbers of hits with the TROIs for the radiologists were statistically significantly different from the number of hits for CAD. This is less than the 5 of 10 statistically significant differences for the radiologist-vs-CAD overlap indices discussed above, due to the thresholding effect associated with determining the number of hits. Overall the results for the number of hits confirm that the performance of the CAD method is very close to the average of the experienced radi-

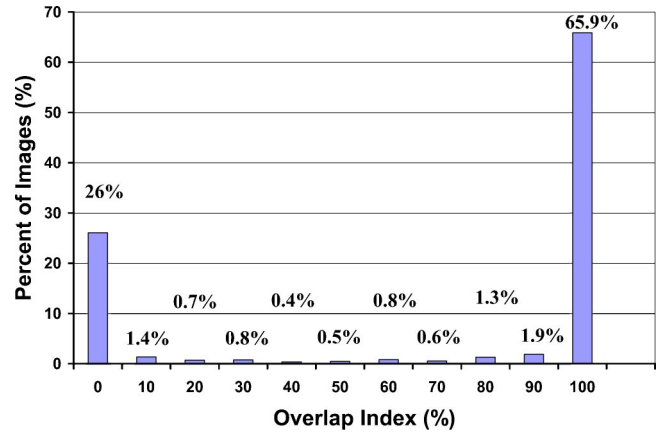


FIG. 5. A histogram summarizing the distribution of largest overlap indices between all radiologist and CAD-selected ROIs in 200 images including data for both reading sessions.

ologists. Finally, just as the agreement between the radiologist- and CAD-selected ROIs was highly correlated with the average number of radiologist-selected ROIs per image, so too was the number of “hits” highly correlated with this average number ($r=0.91$). That is, radiologists who selected more ROIs per image had a higher probability of hits with the TROIs.

The values of the radiologists' ROI-vs-TROI overlap indices were nearly identical for the two reading sessions [reproducibility within 1–7% for all TROIs (Table III, column 4) as were the percentages of “hits” (within 4%, Table IV)]. This indicates the radiologists were very consistent in their selections of suspicious regions for spot imaging relative to the TROIs.

It should be acknowledged that the Hitachi monitor we employed for image display might have influenced the observer study results. The grayscale contrast, brightness and resolution of this monitor are inferior to those of a 2000 line physician's read monitor. Also, we did not employ full (50 μ)-resolution mammograms in this study. Nevertheless, the radiologists all felt the display of the 200 μ and 400 μ resolution images on the Hitachi monitor was adequate for the task of identifying suspicious density and mass regions.

The good agreement between the CAD-selected ROIs and the TROIs indicate that the CAD mass detection program has promise in an implementation of automated stereo spot mammographic imaging of dense areas. We have also developed a microcalcification detection program for CAD. The evaluation of the CAD microcalcification program for automated stereo spot mammographic imaging will be pursued in future studies.

Several practical issues would have to be addressed for the eventual implementation of the automated stereo spot technique including: the number of acceptable false positives for the CAD program, the minimum and maximum sizes of the spot areas, and criteria for determining whether to combine 2 or more ROIs into one. A secondary collimator would have to be designed and built to restrict the x-ray beam to the ROI region at the two projections involved in stereo spot

imaging. In practice, the positions of the collimator blades would be determined from ROI (s) obtained from CAD evaluation of the full-field mammogram, taking into account calibration and geometric factors. The blades would automatically collimate the beam as the mammography technologist shifted the x-ray tube first to the left (or right) and then to the right (or left) to take the stereo spot pair of images. Or, ideally, the x-ray tube or the focal spot should be shifted to the left and the right positions automatically to minimize the time between the images in the stereo pair and thus the breast compression time. The full-field mammograms along with any stereo spot images would be sent to a physician's workstation for interpretation by the radiologist. That workstation would incorporate stereo image display capability. This can be accomplished by adding a stereo graphics board and a transmitter for synchronizing liquid crystal glasses such that the radiologist's left eye would only see the left-eye spot image and the radiologist's right eye would only see the right-eye spot image. We have employed such systems in our stereo mammography research.¹¹⁻¹⁵

A potential limitation of the automated stereo spot mammography method is that radiologists' abilities to fuse images and see them in stereo can be variable. That is, a certain percentage of radiologists may have poor binocular or stereo acuity, which unlike monocular visual acuity cannot be corrected with glasses. However, stereo acuity may be improved through training and use of depth cues provided by three-dimensional 3-D pointers or 3-D cursors. We have developed such cursors¹¹⁻¹³ and may investigate this aspect of stereo imaging in the future.

An alternative method that is not limited by the stereo acuity of the observer and yet should solve the tissue superposition problem would be to perform digital tomosynthesis¹⁶⁻¹⁸ of the spot regions. The equipment required for such an automated spot tomosynthesis method is considerably more expensive than that required for stereo spot imaging. A potential advantage of spot tomosynthesis over full-field tomosynthesis is that the analysis could be concentrated on the suspicious areas. However, CAD techniques could also be employed on the full-field tomosynthesis images to highlight suspicious areas and speed up image analysis, obviating the need for spot tomosynthesis. Whether spot tomosynthesis has utility or not will be determined as we gain more experience with this imaging method.

V. SUMMARY AND CONCLUSIONS

Our observer study showed the regions identified by radiologists as being suspicious and warranting workup spot imaging are in good agreement with those identified by our CAD program. Also, both the radiologist- and the CAD-selected ROIs are in good agreement with the true mass regions as determined from analysis of mammograms and biopsy pathology results. Thus CAD could be used to identify suspicious regions for automatic stereo spot imaging. The stereo spot images are expected to improve perception of lesions that are camouflaged by overlying and underlying tissues in conventional mammograms. Stereo spot imaging

may permit the discrimination of pseudo masses produced by the superposition of overlapping tissues in conventional mammograms, and it may enable better appreciation of the 3-D characteristics of lesions and microcalcifications. Thus, stereo spot imaging has the potential to improve the sensitivity and specificity of mammography. Since the method can be automated, and does not require the presence of on-site radiologists, it can be used in screening. A potential pitfall is the variable ability amongst radiologists to view images stereoscopically. This might be solved through stereo vision training. Another possibility would be to employ spot tomosynthesis of the suspicious regions.

ACKNOWLEDGMENTS

This work is supported by U. S. Army Medical Research and Materiel Command Grants No. DAMD17-99-1-9294 and No. DAMD17-98-1-8210. 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.

^aCorresponding author: Mitchell M. Goodsitt, Ph.D., Department of Radiology, University of Michigan Hospitals, Room B1 F510C, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109-0030. Office: 734-936-7474; fax: 734-936-7948; electronic mail: goodsitt@umich.edu

^bAlso at CDRH, FDA, HFZ-142, 12720 Twinbrook Parkway, Rockville, Maryland 20857.

¹S. Shapiro, "The status of breast cancer screening: a quarter century of research," *World J. Surg.* **13**, 9-18 (1989).

²M. S. O'Malley, S. W. Fletcher, and B. Morrison, "Does screening for breast cancer save lives?" in *Preventing Disease*, edited by R. S. Lawrence and R. Goldbloom (Springer-Verlag, New York, 1989), pp. 251-364.

³R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**, 613-617 (1992).

⁴L. D'Angincourt, "Technique is everything when breast is dense," *Diagn. Imaging* **10**, 57-61 (Sept. 1993).

⁵V. P. Jackson, R. E. Hendrick, S. A. Feig, and D. B. Kopans, "Imaging of the radiographically dense breast," *Radiology* **188**, 297-301 (1993).

⁶R. Hayes, M. Michell, and H. B. Nunnerly, "Evaluation of magnification and paddle compression techniques in the assessment of mammographic screening detected abnormalities," *Clin. Radiol.* **44**, 158-160 (1991).

⁷R. M. Faulk and E. A. Sickles, "Efficacy of spot compression magnification and tangential views in mammographic evaluation of palpable breast masses," *Radiology* **185**, 87-90 (1992).

⁸M. M. Goodsitt, H. P. Chan, H. Huang, and C. Zhou, "Automated spot mammography for improved imaging of dense breasts," *Radiology* **217**, 346 (2000).

⁹A. 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).

¹⁰N. Petrick, B. Sahiner, H. P. Chan, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evaluation of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients," *Radiology* **224**, 217-224 (2002).

¹¹M. M. Goodsitt, H. P. Chan, and L. Hadjiiski, "Stereomammography: Evaluation of depth perception using a virtual 3D cursor," *Med. Phys.* **27**, 1305-1310 (2000).

¹²M. 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).

¹³M. 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." *Proceedings of the 5th International Workshop on Digital Mammography*, Toronto, Canada, 11-14,

- June 2000; IWDM 2000, *5th International Workshop on Digital Mammography*, edited by M. Yaffe (Medical Physics Publishing, Madison, WI, 2001), pp. 45–50.
- ¹⁴H. 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,” in Ref. 13, pp. 13–18.
- ¹⁵H. 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).
- ¹⁶L. T. Niklason *et al.*, “Digital tomosynthesis in breast imaging,” *Radiology* **205**, 399–406 (1997).
- ¹⁷R. 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).
- ¹⁸S. 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).