

AD _____

GRANT NUMBER DAMD17-94-J-4345

TITLE: Evaluation of Digital Mammography Display

PRINCIPAL INVESTIGATOR: Etta Pisano, M.D.

CONTRACTING ORGANIZATION: University of North Carolina
at Chapel Hill
Chapel Hill, NC 27599-4100

REPORT DATE: September 1996

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

19970226 119

DTIC QUALITY INSPECTED 1

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE September 1996	3. REPORT TYPE AND DATES COVERED Annual (1 Sep 95 - 31 Aug 96)	
4. TITLE AND SUBTITLE Evaluation of Digital Mammography Display		5. FUNDING NUMBERS DAMD17-94-J-4345	
6. AUTHOR(S) Etta Pisano, M.D.			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of North Carolina at Chapel Hill Chapel Hill, NC 27599-4100		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick Frederick, Maryland 21702-5012		10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited		12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200) <p>During the first two years we have carried out experiments to evaluate adaptive histogram equalization and intensity windowing applied to mammograms. We found statistically significant improvement in detection of spiculations with contrast limited adaptive histogram equalization processing, and found statistically significant improvement in detection of both calcifications and masses with intensity windowing.[Pisano]. We have also looked at intensity window selection methods based on types of tissue, namely dense breast. We are presently developing a method that will lead to automatic image enhancement.</p> <p>We have identified the appropriate components (Sun Ultrasparc workstation, Dome Md5/Sun 2048x2560 graphics card, Orwin Systems 2048x2560 monitors) for the soft copy mammographic workstation and have them on order. We have prototyped the mammography viewing application, and are awaiting the arrival of the components to assemble the clinical workstation in first part of year three. We expect to have completed the workstation and implement user tools mid way through year 3.</p>			
14. SUBJECT TERMS Breast Cancer		15. NUMBER OF PAGES 2.4	
		16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

DAMD17-94-J-4345 Progress Report year 2

Evaluation of Digital Mammography Display

Abstract

The purpose of this research is to experimentally determine the diagnostic accuracy and interpretation speed of digitally acquired mammograms displayed on the best available display methods.

We propose to conduct an ROC study comparing a film based display to the best available state-of-the-art electronic workstation.

During the first two years we have carried out experiments to evaluate adaptive histogram equalization and intensity windowing applied to mammograms. We found statistically significant improvement in detection of spiculations with contrast limited adaptive histogram equalization processing, and found statistically significant improvement in detection of both calcifications and masses with intensity windowing.[Pisano]. We have also looked at intensity window selection methods based on types of tissue, namely dense breast. We are presently developing a method that will lead to automatic image enhancement.

We have identified the appropriate components (Sun Ultrasparc workstation, Dome Md5/Sun 2048x2560 graphics card, Orwin Systems 2048x2560 monitors) for the soft copy mammographic workstation and have them on order. We have prototyped the mammography viewing application, and are awaiting the arrival of the components to assemble the clinical workstation in first part of year three. We expect to have completed the workstation and implement user tools mid way through year 3.

Introduction

a) Nature of the problem (from original proposal)

A new type of digital mammography device has been developed at the University of Toronto. This scanning slot digital mammography system provides 50um, 12-bit pixels with inherently better contrast than that of conventional mammogram. The advent of digitally acquired mammograms offers the possibility of further improvements in early breast cancer detection. Specifically, digital acquisition systems decouple the process of x-ray photon detection from image display by using a primary detector that directly quantifies transmitted photons. This allows digital systems to be more efficient in utilization of radiation dose. Digital systems also allow a wide dynamic range so that a wider range of tissue contrast can be appreciated. Subtle contrast differences can be amplified and the distinction between benign and malignant might be increased. The new Toronto scanning slot digital mammography system has the further advantage of reduced scatter compared with both conventional and phosphor plate technologies. Furthermore, digital systems have the capacity to bring revolutionary advantages to breast cancer detection and management: 1) image processing for increased lesion conspicuity; 2) computer-aided diagnosis for enhanced radiologic interpretation; 3) teleradiology, or image transmission, as a means of bringing world-class expertise to community hospitals and remote areas; 4) improved image access and communication through digital image archiving and transmission; and 5) dynamic, or "real time" imaging for use during biopsy and localization procedures.

However, there are limitations to both laser-printed film and electronic displays, the two possible display methods for digital mammography. The best quality film printers can only display 87um pixels in an 8"X10" printing of the digital data. This would not provide sufficient spatial bandwidth for the available data. These printers may also lack sufficient greyscale bandwidth. The best possible 2500x2000 pixel monitors can generate over 170-680 nits luminance without pixel bloom. To gain access to the full grey scale bandwidth, monitor display would require intensity windowing, and to view the image at the full 50

mm spatial resolution, roaming and zooming would be necessary. Clearly, any display modality requires compromises that will effect diagnostic accuracy and interpretation speed.

b) Background of previous work (from original proposal)

For a number of years, the Medical Image Presentation research group at UNC-CH has been exploring various issues concerning the display of medical images. Early on we addressed the issues of standardization of display devices to assure legitimate comparison of various display methods under investigation. The display is perceptually linearized so that each intensity step in the acquired image is displayed as an equally perceptible step in the grey levels of the display [Pizer 1981, 1987, 1989, Johnston 1985, Rogers 1987]. In addition, our group, under another grant, (R01 CA44060) has developed and experimentally evaluated the ergonomic and cognitive aspects of electronic workstations. We constructed a prototype workstation called FilmStrip using a single 2048x2560 pixel high-brightness monitor, a very simple interaction, and an extremely fast image display time (0.1 sec). A controlled subject experiment was used to evaluate FilmStrip relative to film and alternator [Beard 1993]. All reports were of clinically acceptable accuracy. Based on our experimental results, we are 95% confident that FilmStrip is no more than 1.5 minutes faster and no more than 30 seconds slower than film. This is the first time a radiology workstation has been shown to be as fast as film for interpretation of medical images under clinically realistic conditions. We have conducted a subsequent experiment showing that a lower cost version of FilmStrip called FilmStriplet can also be clinically viable with sufficient training [Beard 1993].

Under a medical image presentation program project grant, (P01-CA47982), we have been exploring different image processing methods, specifically various versions of the Contrast Limited Adaptive Histogram Equalization algorithm, and have developed an experimental method to optimize the parameters for a given enhancement algorithm that takes into account the deleterious effects of image noise and that does not require the performance of a full clinical trial [Puff, 1992]. This work has involved the conduct of a number of image quality assessment experiments.

Under the previously described interactive Digital Mammography Development Group grant, Gray Scale Image Processing For Digital Mammography, (R01 CA 60193), we are conducting preliminary experiments to determine the effect of the variable amount of radiographically dense breast tissue, the mammographic characteristics of various lesion types, and the location of lesions within the breast on the choice of appropriate intensity windows and other image processing algorithms selected for electronic viewing of mammograms. The results of this investigation will also give us some indication of the number of intensity windows that might be useful, or needed, for display of the recorded digital information.

c) Purpose of present work

The purpose of this study is to determine experimentally the diagnostic accuracy and interpretation speed of the available display methods.

d) Methods of approach

We propose to conduct an ROC study involving the best available display methods, one representative of a film based display, and one using the best available state-of-the-art electronic workstation.

Research Methods and Results to date

1. To achieve the goals of this research, we propose using digitally acquired mammograms. Availability of the clinical digital units have been continuously delayed

because of detector upgrades and manufacturing problems. However, the first unit has been delivered at the end of August to Brooke Army Medical Center. The second and third units are expected to be delivered to Thomas Jefferson Hospital and UNC sometime between October and December '96. We are presently in the process of site preparation for the unit to be delivered to UNC. We expect to begin collection of clinical images the first quarter of '97. During the 03 year, we will install the clinical unit, test, calibrate and begin clinical use. The actual ROC observer studies will not begin until sometime in the 04 year.

2. Since the inception of this grant, a number of technical advances have been made that directly modify the experimental procedures to be carried out under this proposal. A major change is that there are now laser printers that can meet the requirements for display of mammograms on an 8"x10" format with 12 bits of gray levels.

We will have a laser printer (Kodak) obtained along with the Fischer Digital Mammography unit to be located at UNC Hospitals. We will have access to our own digital mammograms as well as those from Thomas Jefferson hospital and from Brooke Army Medical Center. Thus, the delay in obtaining the digital units is offset by the eventual increased availability of clinical images.

3. During the first two years of this grant, a number of changes in the state-of-the-art of monitor technology has occurred, a) High brightness/resolution monitors, although commercially available, have not been as readily available as once promised. There are manufacturing problems in quality assurance and meeting performance specifications. We have evaluated a number of different brands in our laboratory and with collaboration of Dr. Hans Rhoerig at Univ. of Arizona and Dr. Harwig Blume at Philips Medical. As a result of these extensive evaluations, we have purchased a DataRay and two Orwin monitors. Unfortunately, the interface electronics to drive 2k x 2.5k monitors from conventional host computers at greater than 8 bits grey levels has lagged behind and only now are becoming available on a very limited basis. We have purchased the electronics from Dome (10 bits grey level) and will carry out installation over the next few months.

4. We have completed experiments to determine the parameter values to be used in conducting observer experiments to evaluate the use of intensity windowing and contrast limited adaptive histogram equalization (CLAHE) applied to mammograms. As reported last year, observer studies using CLAHE showed significant improvement in the detection of spiculations, but not for masses [ref]. We also completed observer studies with preset intensity windows selected for masses and calcifications. Our results showed statistically significant improvement in detection of both features with specified values for the intensity windows [ref]. We have begun the development of improved intensity windowing methods that automatically determine the appropriate intensity windowing range individually for each mammogram. This research is also partially supported by NIH R01-CA60193.

5. We have designed an experiment to determine the effect of display luminance range on the detection of mammographic features. This observer study uses film displayed with maximum luminance at 30, 100, 200, and 800 FL to simulate the luminance range of typical and high brightness monitors compared to the lightbox. This observer experiment is presently underway and should be completed the first quarter of year 03.

6. We have developed our softcopy mammography display system, MammoView, for viewing digital mammograms on video monitors. The design is based upon concepts developed in our previously successful softcopy experiments in chest CT and chest Xray, but adapted according to the results of our mammography eyetracking experiments [Beard] and our GOMS (human computer interaction) modeling of radiologists reading mammograms. This software is being ported to the new display system hardware that has been ordered for the clinical workstation.

7. We are pursuing research in the area of display function standardization, to allow as similar as possible presentations on the softcopy images as on the film images. To this end, Mr. Hemminger is helping authoring the ACR/NEMA display function standard, and is pursuing research aimed at quantifying how closely matched a display system is to the proposed display function standard.

Conclusions

Although the acquisition of digital mammograms has been delayed due to delivery problems of the new Digital Mammogram systems, we have made significant progress in:

1. evaluation of the intensity windowing as an image enhancement method, and development of improved automatic intensity windowing methods.
2. evaluating available high brightness monitors and associated driving electronics, and ordering of clinical workstation components.
3. developing the software tools for the electronic mammographic workstation.

Proposed research for the 03 year period

1. Implement the UNC electronic mammography workstation.
2. To install the Fischer digital mammographic unit and Kodak laser printer into UNC Hospitals. Calibration of the unit and coupling to the UNC mammographic workstation.
3. To redesign the experimental protocol for improved and more efficient data collection to meet the goals of this grant. The redesign in no way alters the ultimate goal of this research. Primarily, it accomodates the advances in technology that has occurred since the original experiments were proposed and should result in improved ROC studies.
- 4 As a result of the delay in delivery of a Digital Mammographic acquisition system, the slow development of the state-of-the-art high brightness monitors, and lack of access to clinical digital mammograms, we operated under a reduced budget during the 02 year. Our intention is to ramp up the activity level of the personnel and the budget upon delivery of the Digital Mammography unit in order to begin the acquisition of clinical studies. We expect, that because of the delay of availability of digital images, to request an extension into an 05 year with no new funding. This is the purpose of reducing our present budget expenditures.

References

- Beard DV, Hemminger BM, Perry JR, Mauro M, Muller K, Warshauer, Zito A, and Smith M Single-Screen Workstation vs. Film Alternator for fast CT Interpretation, *Radiology*, 1993; 187(2):1-6.
- Johnston RE, Zimmerman JB, & Pizer SM: "Perceptual standardization", Proc. SPIE 536, 44-49, 1985.
- Pizer SM Intensity mappings to linearize display devices, Comp. Graph. Image Proc. 17, 262-268, 1981.
- Pizer SM, Rogers D, Johnston RE, & Beard, DV "Effective Presentation of Medical Images on an Electronic Display Station", *RadioGraphics*, Vol. 7, No. 6, 1267-1274, Nov. 1987.
- Pizer SM, & Beard DV. "Medical Image Workstation: State of Science & Technology". *Journal of Digital Imaging*, Nov. 1989 2(4) 185-193.

Rogers C, Johnson RE, Hemminger BM, Pizer SM, (1987) Effect of Ambient Light on Electronically Displayed Medical Images as Measured by Luminance Discrimination Thresholds. *J. Optical Society Am* 4(5) 976-983.

Publications:

Bradley M. Hemminger, Hans Roehrig, Hartwig Blume, David Channin, Fred Prior, R. Eugene Johnston, "Demonstration of image quality and display system standardization for high quality softcopy video display systems", InfoRad presentation at RSNA95.

Hemminger BM, Blume HR, "Are Medical Image Display Systems Perceptually Optimal: Measurements before and after Perceptual Linearization", SPIE Medical Imaging Vol 2707-58, Feb 1996.

Blume HR, Hemminger BM, "The ACR/NEMA Proposal for a Gray-Scale Display Function Standard", SPIE Medical Imaging Vol 2707-35, Feb 1996.

Blume HR, Hemminger BM, "Image Presentation in Digital Radiology", submitted to Radiographics 1996.

Beard DV, Bream P, Pisano ED, Conroy P, Johnston RE, Braeuning P, McLelland R, Clark R. "Eye Movement During Mammography Interpretation: Eyetracker Results and Workstation Design Implications. Accepted for publication by Journal of Digital Imaging.

Manuscripts in preparation:

The following papers have been presented at the AUR mtg in Birmingham, Ala., April 19, 1996, and are in preparation for journal submission.

Pisano ED, Heminger BM, Garrett W, Johnston E, Chandromouli J, Glueck D, Muller K, Braeuning MP, Puff D, Pizer S. Does CLAHE Image Processing Improve the Detection of Simulated Masses in Dense Breasts in a Laboratory Setting?

Pisano ED, Heminger BM, Garrett W, Johnston E, Zong S, Glueck D, Muller K, Braeuning MP, Puff D, Pizer S. Does CLAHE Image Processing Improve the Detection of Simulated Spiculations in Dense Breasts in a Laboratory Setting?

Appendix:

Pisano ED, Chandramouli J, Hemminger BM, Johnston RE, Muller KE, Pizer SM. "The Effect of Intensity Windowing in Detection of Masses Embedded in Dense Mammograms In A Laboratory Setting". Submitted to Academic Radiology.

APPENDIX

The Effect of Intensity Windowing on the Detection of Masses Embedded in Dense Mammograms in a Laboratory Setting.

Etta D. Pisano, MD*#
Jayanthi Chandramouli
Bradley M. Hemminger, MS*
Deb Glueck, MS^
R. Eugene Johnston, PhD*
Keith Muller, PhD^
M. Patricia Braeuning, MD*#
Derek Puff, PhD`°°
William Garrett, MS@
Steve Pizer, PhD**#@

From the Departments of Radiology*, Computer Science@, Biomedical Engineering`, and Biostatistics^.

The University of North Carolina-Chapel Hill.

School of Medicine, School of Public Health and College of Arts and Sciences.

°°Current affiliation, Department of Radiation Oncology, University of Michigan

Member, UNC-Lineberger Comprehensive Cancer Center #

Supported by NIH PO1-CA 47982, NIH RO1-65583 and DOD DAMD17-94-J-4345

Abstract

Purpose:

To determine whether intensity windowing (IW) improves detection of simulated masses in dense mammograms.

Materials and Methods:

Simulated masses were embedded in dense mammograms digitized at 50 micron pixels, 12 bits deep. Images were printed with no windowing applied and with nine window widths and levels applied. A simulated mass was embedded in a realistic background of dense breast tissue, with the position of the mass (against the background) being variable. The key variables involved in each trial included the position of the mass, the contrast levels and the IW setting applied to the image. Combining the 10 enhancement conditions, 4 contrast levels and 4 quadrant positions gave 160 combinations. The trials were constructed by pairing 160 combinations of key variables with 160 backgrounds. The entire experiment consisted of 800 trials. Twenty observers were asked to detect the quadrant of the image into which the mass was located.

Results:

There was a statistically significant improvement in detection performance for masses when the window width was set at 1024 with a level of 3328.

Conclusion:

IW should be tested in the clinic to determine whether mass detection performance in real digital mammograms is improved.

Background and Significance

Effective image display allows for an improvement in the clarity of structural details. Mammography, especially in patients with dense breasts, is a low contrast examination that might benefit from increased contrast between malignant tissue and normal dense tissue. Image processing may allow for improved visualization of details within medical images [1]. Our overall aim is to improve the accuracy of mammography with image processing since as many as 10% of palpable breast cancers are not visible with standard mammographic techniques[2].

Contrast enhancement methods accentuate or emphasize particular objects or structures in an image by manipulating the gray levels in the display. This is done by imposing a predetermined transformation that amplifies the contrast between structures and effectively "resamples" the recorded intensities to enhance the properties of the displayed image [3]. These methods are not designed to increase or supplement the inherent structural information in the image, but simply improve the contrast and theoretically enhance particular characteristics [4]. Intensity Windowing (IW) is an image processing technique that involves the determination of new pixel intensities by a linear transformation which maps a selected band of pixel values onto the available gray level range [4].

Many investigators have studied the application of digital image processing techniques to mammography. McSweeney tried to enhance the visibility of calcifications by using edge detection for small objects, but never reported any clinical results [5]. Smathers showed that intensity band-filtering could increase the visibility of small objects compared to images without such filtering [6]. Chan used unsharp masking (an edge-sharpening technique used in photography for many years) to remove image noise for computerized detection of calcification clusters [7]. Chan noted that while these techniques improved detection, the improvements may have been greater if the observers had been trained to make diagnoses from the processed mammograms rather than the unprocessed (normal) mammograms [8]. Hale et al. have applied non-specific contrast and brightness adjustment through Adobe Photoshop™ to digitized mammograms and have found improved performance by radiologists in determining the likelihood of malignancy of mammographically apparent lesions [9]. Yin et al. showed that nonlinear bilateral subtraction is useful in the computer-detection of mammographic masses [10,11].

Previous work at UNC has explored the use of Intensity Windowing (IW) and the Adaptive Histogram Equalization (AHE) family of algorithms in mammography and computed tomography [12-14]. We have previously described a laboratory-based method for testing the efficacy of an image processing algorithm in improving the detection of masses in dense mammographic backgrounds [15]. With that method, upon which our current work is based, radiologists and non-radiologists exhibit similar trends in detection performance. While non-radiologists did not perform as well as radiologists overall, the two populations displayed parallel increases and decreases in performance due to image processing.

The experiments described in this paper were performed to determine whether IW could improve the detection of simulated masses in dense mammograms in a laboratory setting.

Materials and Methods

The experimental paradigm reported here is based on the model we have previously described and allows for the laboratory testing of a range of parameter values (in this case, window width and level) [15]. The experimental subject is shown a series of test images that consist of an area of a dense mammogram with a simulated mass embedded in the image in one of its four quadrants. The observer's task is to determine in which quadrant the mass is located. The test images are displayed in both the processed and unprocessed format, and the contrast of the object is varied from quite easy to detect to impossible to detect.

A computer program randomly selected one of 40 background images and rotated that background to one of four orientations. The 40 background images of 256x256 pixels each were extracted from actual clinical mammograms digitized using a Lumiscan digitizer (Lumisys, Inc. Sunnyvale, CA) with a 50 micron sample size with 12 bits (4096 values) of intensity data per sample. The images contained relatively dense breast parenchyma. They were known to be normal by virtue of at least three years of normal clinical and mammographic follow-up. They were selected by a breast imaging radiologist from digitized film screen craniocaudal or mediolateral oblique mammograms. Figure 1 shows one of the backgrounds.

The grey scale values for the mammographic backgrounds are assigned the values recorded by the Lumisys digitizer. The digitizer assigns digital values in the range [495, 4095] to the optical density range [3.43, 0.08]. The relationship between OD values and digitized values is constant (i.e. the same optical density produces the same digital value).

These 40 images and four orientations provided 160 different dense backgrounds. Next, the program added a phantom feature (a mass) into the background. The image was processed with IW to yield the final stimulus.

Mammographic masses were simulated by blurring (via convolution with a Gaussian kernel with a standard deviation of 2.0 pixels) a disk that is approximately 5mm in diameter (1.51 degree visual angle at a 38 cm viewing distance). The intensity difference of the mass from background, then, is defined to be the maximum gray level at the center of the mass before addition to the background. The masses were then embedded at specific differences in intensity level relative to background so as to be equally spaced in perceived brightness by a pixel-wise addition of the structure and background images. Although the simulated structures were not entirely realistic, they did, however, possess the same scale and spatial characteristics of actual masses typically found at mammography. Figure 2 shows an example of a simulated mass. Figures 3a and 3b show a typical background image with the mass added to it. We used simulated features instead of real features so that we could have precise control over the location, orientation, and figure to background contrast of the masses.

A three by three (3x3) grid of window and level parameters was designed, based on the results of pilot preference studies done with two radiologists who specialize in breast imaging. In the pilot studies, the two radiologists reviewed dense mammograms with real clinical lesions that were judged to be difficult to visualize using standard film screen mammography. There were 7 cases of this type reviewed with 70 combinations of window width and level applied. They scored each combination of values as showing no change over the standard image, improving the visibility of the lesion, or worsening its visibility.

For experiment 1, the grid spanned all the likely optimal settings (windows of 512, 768, 1024 and levels of 3072, 3328, 3584). Thus, there were a total of 10 IW settings (including the default unprocessed image, with Window width= 4096, Level = 2048) that were applied throughout experiment 1.

To confirm the results of the first experiment and to examine other IW settings, experiment 2 was performed. Experiment 2 also included the unprocessed (wide open window width) condition and 9 other IW conditions. The combinations of parameters evaluated in Experiment 2 were as follows: window width of 640 with levels of 3456, 3584 and 3840; window width of 1024 with levels of 3200, 3328 and 3584; and window width of 1536 with levels of 2944, 3072, and 3328).

The digital images were printed in nonmagnified fashion at 50 microns per pixel onto standard 14X17 inch single-emulsion film (3M HNC Laser Film, 3M, St. Paul, MN) using a Lumisys Lumicam film printer (Lumisys Inc., Sunnyvale, CA). Forty images were printed per sheet of film. Each image measured 3.3 cm². The images were randomly ordered into an 8 by 5 grid on each sheet of film. Both film digitizer and film printer were calibrated, and measurements of the relationship between optical density on film and digital units on the computer were made to generate transfer functions describing the digitizer and film printer.

In order to maintain a linear relationship between the optical densities on the original analogue film and the digitally printed film, we calculated a standardization function that provided a linear matching between the digital and printer transfer functions. The film printer produces films with a constant relationship between an optical density range of 3.35 OD to 0.13 OD, corresponding to a digital input range of 0 to 4095, respectively. The relationship between optical density values and digital values was different between the film digitizer (which we used to acquire the mammographic backgrounds) and the film printer (which we used to produce the observer films). So that images would be as similar as possible to the original mammograms when printed to film with the film printer, we corrected the film printer output via a standardization lookup table function applied to the digital values on the film printer. This function is simply the inverse of the uncorrected transfer function of the entire Digitizer-Printer Process, as described in Hemminger.¹⁶

There were 20 observers for each experiment. These were graduate students from the medical school, biomedical engineering department, and computer science department. Performance bonus pay was provided. Observers selected the quadrant of the image that they thought contained the mass. All images contained a mass.

Observers were told to make their best guess.

Films were displayed in a darkened room on a standard mammography lightbox that was masked so that only the grid of images on the film was illuminated. Observers could move closer to the image and could use a standard mammography magnifying glass, as desired. The observers were trained for the task through the use of two sets of stimulus image films with instructive feedback before actually starting the experiment.

Both experiments had the same basic design. The order of the presentation of the stimuli was counterbalanced so as to eliminate any systematic effect of non-important variables. All 160 possible combinations of processing condition (10 IW levels), contrast level (4 contrasts) and location of the masses (4 quadrants) were used in the experiment. The experiment was designed to have 5 self-contained blocks, in which all 160 combinations appeared. The intent was to have the observer see all the combinations in each block, in case the observer was unable to complete the experiment. In fact, all observers did complete the experiment. There were 40 backgrounds and 4 possible rotations of each background, for 160 possible background patterns. For each block, a different background pattern was assigned uniquely to each of the 160 possible combinations. The assignment was different for each block. Each observer looked at a total of 800 images, which were the 160 possible combinations, each superimposed on 5 backgrounds.

Observers were instructed to take breaks after each block of stimuli, and more often if necessary. No time limit was imposed on the observers viewing duration of the test images. Overall, the experiment took 2 hours for each observer, divided into two sessions of approximately 60 minutes each. The two sessions were always scheduled on two different days within a week of each other.

Data Analysis Overview

Classical sensory discrimination theory predicts that since contrast values were varied from virtually imperceptible to highly apparent, a typical S-shaped curve will describe the data[2]. At values where the contrast was very low, observers will on average guess randomly and get approximately 25% right, since there are four choices. Where the contrast is very high, they will almost always get the correct answer. This relationship between \log_{10} of the intensity offset of the object relative to the background intensity and the percent correct can be described with a probit model. This model is typically used to describe the relationship between a continuous predictor (log intensity offset) and a discrete variable (percent correct), and assumes that the curve between them is described by the cumulative Gaussian distribution.

Probit models were fit for each subject and enhancement condition using density above background as the predictor. The probability that a subject gets a correct answer is given by the following equation:

$$Pr\{correct\} = \frac{1}{4} + (1 - \frac{1}{4}) \Phi \left(\frac{x - \mu_{ij}}{\sigma_i} \right)$$

where i indexes subjects, and j indexes enhancements. For each subject, this gave a separate location parameter estimate for each enhancement, and a common spread parameter estimate. Our assumption is that there is a common spread parameter makes sense biologically, since it corresponds to linearity of the perceptual mapping. It is advantageous to an organism to have the same amount of change in stimulus produce a constant perceptual response, and that is precisely how the human visual system works.

The location parameter, μ , is the mean of the corresponding Gaussian distribution and the inflection point of the sigmoidal probit curve. Processing conditions that improve detection will cause this parameter to be smaller, and the curve will shift to the left. This occurs because lower contrast levels are required to spot the object. When the processing of the image makes detection harder, higher contrast levels are needed to locate the mass, and the curve shifts to the right. The values of σ , the spread parameter, correspond to the slope of the line. Large values of σ correspond to steep slopes.

The probit analysis summarized the relationship between contrast and proportion correct for each subject and processing condition. To compare the processing conditions and to examine the effect of window width and level, further analysis was needed. To include both the mean and the location parameter from the probit analysis, we defined an overall measure to be $\theta_{ij} = \mu_{ij} + \sigma_i$, which corresponds to 88% correct. Because we were interested in the improvement offered by IW, we measured the "success" of a processing condition by calculating the difference between its θ score and the θ score for the unprocessed image for each subject. A large positive difference-of- θ score reflects improved performance, because it indicates better detection with processed images than with unprocessed images.

For each experiment, two analyses were performed using this outcome measure. To keep an overall experiment-wide type I error rate of .05, a repeated measures analysis of variance was done at the .04 level, with a set of T-tests at the overall .01 level.

Repeated measures analysis of variance (ANOVA) is a technique used to analyze data in which many measurements were made on different subjects. It allows one to examine the effect of processing conditions and their interactions, while allowing for the dependence of measurements taken on the same observers. With the difference in θ scores as the outcome, and window width and level as the predictors, the repeated measures ANOVA model was fitted.

The model can be thought of as a response surface in three dimensions with performance plotted against window width and level. A flat surface would mean that window width and level had no effect on the outcome. The major hypothesis tested in the ANOVA is equivalent to asking the question "Is the response surface flat?". If it is not flat, the step-down hypotheses allow one to ask what shape the surface is, whether it is curved in both directions (quadratic by quadratic trends), curved in one direction and sloped in the other (quadratic by linear trends), or sloped in both directions (linear by linear trends). A peak in the surface means that there is one image processing technique that is better than any other. Conversely, if the difference score is equal to zero for any intensity windowing setting, it would correspond to no difference between the processed image and the unprocessed image. That is what the T statistics test.

Results: Experiment 1

The repeated measures ANOVA revealed that there was a significant interaction between window width and level ($p=.0001$, Geiser-Greenhouse estimate of $\epsilon = .8347$). To examine the nature of this interaction, a series of step-down tests was planned. There was a significant interaction between a quadratic trend in window width and a quadratic trend in level ($F=31.08$, $p=.0001$). Because the quadratic by quadratic interaction was significant, no further tests were examined. A quadratic by quadratic trend means that the surface was curved with respect to both window width and level, and that the shape of the curve differed for fixed levels of window width and level. (Figures 4 and 5).

At the overall .01 level, the differences between the enhancement conditions and the unenhanced were examined. The null hypothesis is that there will be no difference between the mean θ for the unenhanced and an enhancement condition. There are nine such hypotheses, corresponding to the nine enhancements. A Bonferroni correction to control the overall error rate made each individual nominal type I level .0011. Four settings of intensity windowing made finding the masses significantly harder, three made the task significantly easier and two made no significant difference. The settings that made the task easier are window width 1024 with level 3328, window width 768 with level 3584 and window width 1024 with level of 3584. (Table 1)

Results: Experiment 2

Again the repeated measures ANOVA showed that there was significant interaction between window width and level ($p<0.0001$, $F=60.9$, $\hat{\epsilon}=.3369$). (Figures 6 and 7) As in experiment 1, a quadratic by quadratic interaction was significant ($p<0.0001$, $F=32.61$). Table 2 shows the results of nine two-sided t-tests. Only one image processing setting resulted in significantly better performance than the unprocessed, namely window width of 1024 with a window level of 3328 ($p<0.0001$). Seven of the settings were not significantly different from the unprocessed image. One

setting was significantly worse. (Table 2)

The probit model predicts that IW will increase detection of masses by as much as 17% in cases near the threshold of detection. (Figures 5 and 7).

Discussion

These results are encouraging. This is the first experiment in mammography that demonstrates that an algorithm can improve the conspicuity of a mass placed in a dense mammogram. At the same time, it is obviously important to choose the window width and level with care since performance can be significantly degraded if inappropriate parameters are chosen.

What do these results mean for clinical mammographers? Will we be using this technology in the clinic in detecting lesions in dense mammograms? The use of graduate student observers and the use of simulated masses in this study might incorrectly predict the performance of radiologists in detecting real masses in real patients. However, we have demonstrated previously that graduate student performance at this task parallels the performance of experienced mammographers [15]. Evaluation by radiologists on real patients will determine the ultimate utility of this algorithm in the clinical setting. Because we have used real clinical images and we have simulated masses using relatively realistic stimuli, we are optimistic that these methods will improve clinical performance. If so, radiologists will be using IW to help them determine whether mammograms of women with dense breasts really do contain masses.

Digital mammography is coming to the clinic very soon. It is obvious that image processing will be used to optimize the visibility of lesions in digital mammograms. (17) In the simplest approach, any image processing algorithm that might be useful would be tested on real patients in that setting. That would be an expensive and time consuming process that would involve real patients making clinically important decisions about their own breast health, including the adviseability of biopsy, lumpectomy and mastectomy. It would be preferable, in our opinion, before this technology arrives in the clinic, for radiologists to have some idea of which category of algorithms to test in that setting. This work is intended to help radiologists narrow the choices before expensive clinical tests are undertaken. This kind of work is necessary to test both the available image processing algorithms and the parameter settings of the algorithms that most improve conspicuity of lesions in some objectively verifiable manner.

One could take the approach that the IW dials should be spun until a clinically pleasing image is displayed. This approach might be acceptable and even convincing to some radiologists. It is, however, possible that what pleases radiologists in viewing an image might not improve the detection performance. This project was intended to be more rigorous in exploring the window widths and levels that might be useful in the

most challenging areas of the breast, namely the dense parts. We have performed similar experiments on the AHE class of algorithms also. (18, 19)

This experiment does not address how IW would effect the appearance of fatty areas of the breast, and the conspicuity of lesions in those parts. We would not want to apply an algorithm that degrades performance in areas of the breast where sensitivity is quite high with current technology. There are two possible technical responses to that concern. First, IW could be applied selectively to only the dense areas as an adjunct to the more standard appearing mammogram with the radiologist pointing and clicking to the areas where windowing would be desirable. Alternatively, the IW could be individualized to the patient's unique intensity histogram so that the areas to be processed of the image could be selected by the computer itself. In fact, ideally the computer could be programmed to choose an individual IW setting for each portion of the mammogram so that contrast was preserved in all portions of the image. Ongoing experiments in our laboratory are currently exploring the latter possibility.

Of course, our results to date cannot estimate the exact frequency of false positive diagnoses when intensity windowing is used. Many alternate forced choice tests (in our case, 4-AFC) yield proportion correct as the primary outcome. Macmillan and Creelman discussed methods for converting proportion correct in this setting to a value of d' , the sensitivity parameter of an ROC analysis.²⁰ The particular choice of conversion depends on side conditions concerning the nature of any rater bias. Given the characteristics of the study design, subjects and training, we believe that superior proportion correct will translate into superior d' . If this is true, the practical value of intensity windowing must be tested in a clinical setting. Then ROC analysis will allow separate analysis of a reader's sensitivity and pay off function on the performance of the technique as part of a diagnostic system.

The testing of these methods on patients with palpable and mammographically detected lesions has been funded by the National Cancer Institute and the Department of Defense, and will be ongoing over the next few years at UNC and Thomas Jefferson University Hospital. We expect to evaluate both Intensity Windowing and Contrast Limited Adaptive Histogram Equalization (CLAHE) in the clinical setting to determine whether or not these algorithms improve the performance of radiologists in detecting and characterizing breast lesions.

References

1. Rosenman J, Roe CA, Cromartie R, et al. Portal film enhancement: Technique and clinical utility. *Int J Radiat Oncol Biol Physics* **1993**;25(2):333-338.
2. Homer MJ. *Mammographic interpretation: a practical approach*. New York, NY: McGraw Hill, **1992**:4-5.
3. Pizer SM. Psychovisual issues in the display of medical images. In: K.H.Hoehne, ed. *Pictorial Information Systems In Medicine*. Berlin: Springer-Verlag, **1985**: 211-234.
4. Jain AK. *Fundamentals of Digital Image Processing*. Englewood Cliff, NJ: Prentice Hall, **1989**.
5. McSweeney MB, Sprawls P, Egan RL. Enhanced image mammography. *AJR* **1983**;140:9-14.
6. Smathers RL, Bush E, Drace J, et al. Mammographic microcalcifications: Detection with xerography, screen-film, and digitized film display. *Radiology* **1986**;159:673-677.
7. Chan HP, Doi K, Galhorta S, et al. Image feature analysis and computer-aided diagnosis in digital radiography: I. automated detection of microcalcifications in mammography. *Med. Phys.* **1987**;14(4):538-547.
8. Chan HP, Vyborny CJ, MacMahon H, et al. Digital mammography ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications. *Investigative Radiology* **1987**;22:581-589.
9. Hale DA, Cook JF, Baniqued Z, et al. Selective Digital Enhancement of Conventional Film Mammography. *J Surg Onc* **1994**;55:42-46.
10. Yin F, Giger ML, Vyborny CJ, et al. Comparison of Bilateral-Subtraction and Single-Image Processing Techniques in the Computerized Detection of Mammographic Masses. *Inv Rad* **1993**;28(6):473-481.
11. Yin F, Giger M, Doi K, et al. Computerized detection of masses in digital mammograms: Analysis of Bilateral Subtraction Images. *Med Phys* **1991**; 18(5):955-963.
12. Hemminger BM, Johnston RE, Muller KE, et al. Comparison of clinical findings between intensity-windowed versus CLAHE presentation of chest CT images. *SPIE Medical Imaging VI: Image Capture, Format and Display* **1992**;1653:164-175.

13. Pizer SM, Zimmerman JB, Staab EV. Adaptive grey level assignment in CT scan display. *Journal of Computer Assisted Tomography* **1984**;8(2):300-305.
14. Puff DT, Cromartie R, Pisano ED. Evaluation and optimization of contrast enhancement methods for medical images. *Proceedings of the SPIE Visualization in Biomedical Computing Conference* **1992**;1808:336-346.
15. Puff DT, Pisano ED, Muller KE, et al. A method for determination of optimal image enhancement for the detection of mammographic abnormalities. *J Dig Imaging* **1994**;7:161-171.
16. Hemminger BM, Johnston RE, Rolland JR, Muller KE. Introduction to perceptual linearization for video display systems for medical image presentation. *J Digital Imaging* **1995**;8(1):21-34.
17. Shtern F. Digital mammography and related technologies: a perspective from the National Cancer Institute. *Radiology* **1992**;183(3):629-30.
18. Pisano ED, Hemminger BM, Garrett W, Johnston RE, Chandramouli J, Gleuck D, Muller K, Braeuning MP, Puff D, Pizer S. Does CLAHE image processing improve the detection of simulated masses in dense breasts in a laboratory setting? *Association of University Radiologists Meeting*. Birmingham, Alabama, April 19, **1996**.
19. Pisano ED, Hemminger BM, Garrett W, Johnston RE, Zong S, Gleuck D, Muller K, Braeuning MP, Puff D, Pizer S. Does CLAHE image processing improve the detection of simulated spiculations in dense breasts in a laboratory setting? *Association of University Radiologists meeting*. Birmingham, AL, April 19, **1996**.
20. MacMillan NA, Creelman CD. *Detection theory: a user guide*. Cambridge, England: Cambridge University Press, **1991**:135-136.

CAPTIONS:

Figure 1: An example of one of the dense normal backgrounds taken from a patient's mammogram and used in the reported experiments.

Figure 2: An example of a simulated mass. The actual size of the masses used in the experiments was only 5 mm.

Figures 3a and 3b: A dense background with a simulated mass embedded in it in the right upper quadrant (arrows). Figure 3a is the default unprocessed image with window width 4096 and level 2048. Figure 3b is the same image with window width 1024 and level 3328.

Figure 4: Interpolated predicted values from repeated measures ANOVA for Study 1: difference in θ value versus window width and window level.

Figure 5: Estimated detection probability from Study 1 for window width of 1024 and window level of 3328 versus unprocessed condition. The shift in the curve to the left reflects improved detection.

Figure 6: Interpolated predicted values from repeated measures ANOVA for Study 2: difference in θ value versus window width and window level.

Figure 7: Estimated detection probability from Study 2 for window width of 1024 and window level of 3328 versus unprocessed condition. The shift in the curve to the left reflects improved detection.

Table 1: Summary of differences between unenhanced and enhanced theta for Study 1.

Table 2: Summary of differences between unenhanced and enhanced theta for Study 2.

Table 1, Study 1: Differences between unenhanced and enhanced θ . Positive values in mean difference in θ column correspond to improved detection of simulated masses.

Window Level	Window Width	Mean Diff in θ	Std Dev	p-value
3072	512	-.50	.108	.0001
3072	768	-.32	.093	.0001
3072	1024	-.34	.089	.0001
3328	512	-.11	.074	.0001
3328	768	.04	.087	.0706
3328	1024	.18	.104	.0001
3584	512	-.03	.097	.1716
3584	768	.14	.082	.0001
3584	1024	.12	.121	.0004

Table 2, Study 2: Differences between unenhanced and enhanced θ . Positive values in mean difference in θ column correspond to improved detection of simulated masses.

Window Level	Window Width	Mean Diff. In θ	Std Dev	p-value
3456	640	0.04	0.08	0.0239
3584	640	-0.05	0.09	0.0215
3840	640	-0.31	0.09	0.0001
3200	1024	0.04	0.07	0.0142
3328	1024	0.14	0.08	0.0001
3584	1024	0.01	0.09	0.6155
2944	1536	-0.02	0.07	0.1255
3072	1536	0.06	0.08	0.0045
3328	1536	0.06	0.07	0.0013