

AD _____

GRANT NUMBER: DAMD17-93-J-3015

TITLE: Computer Assisted Quality Control and Telemammography

PRINCIPAL INVESTIGATOR: Matthew Freedman, M.D.

CONTRACTING ORGANIZATION: Georgetown University
Washington, DC 20057

REPORT DATE: August 1996

TYPE OF REPORT: Final

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

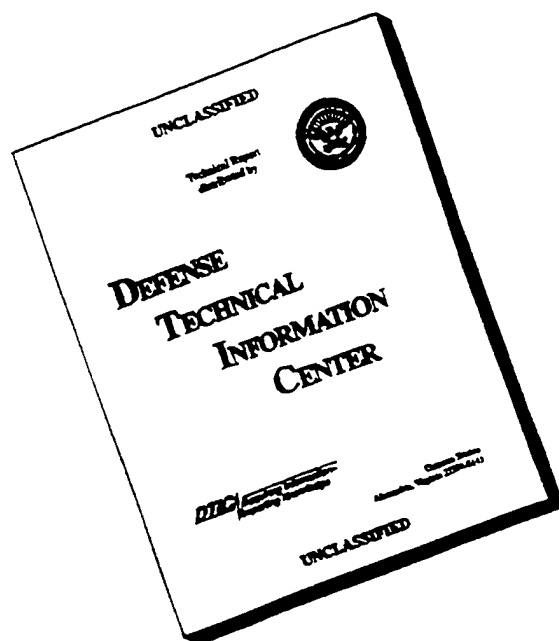
DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

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.

19970327 011

1 DTIC QUALITY INSPECTED 1

DISCLAIMER NOTICE



**THIS DOCUMENT IS BEST
QUALITY AVAILABLE. THE
COPY FURNISHED TO DTIC
CONTAINED A SIGNIFICANT
NUMBER OF PAGES WHICH DO
NOT REPRODUCE LEGIBLY.**

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 August 1996		3. REPORT TYPE AND DATES COVERED Final (15 Dec 92 - 14 Jul 96)	
4. TITLE AND SUBTITLE Computer Assisted Quality Control and Telemammography				5. FUNDING NUMBERS DAMD17-93-J-3015	
6. AUTHOR(S) Matthew Freedman, M.D.					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Georgetown University Washington, DC 20057				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Commander U.S. Army Medical Research and Materiel Command Fort Detrick, MD 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) This purpose of this project, DAMD17-93-J-3015 was to develop systems to enable the integration of mammography into MDIS for digital storage and soft copy display, to allow the acquisition of mammograms in facilities remote from the radiologist who will interpret the images, to develop automated methods of quality control of image quality and to develop methods for the teletransmission of digital mammograms. We have demonstrated the effectiveness of digital mammography using a direct digital method, but in our tests, digitized film was inadequate for diagnosis. We have developed a system for the teletransmission of mammograms that results in hard copy display of adequate diagnostic quality. We have developed a computerized system for the detection of under and over-exposure of mammograms for both digitized screen film mammograms and digitally captured mammograms. We were unable to devise a system for the soft copy display of mammography that provided adequate quality for soft copy interpretation. Implementation with a Fuji FCR 9000 with image processing parameters that we have optimized and printed as hard copy on a 3M 969 laser printer is feasible. Deprad, which is a product of the Analogics Corporation that we helped to design, is a suitable method for teledigital mammography.					
14. SUBJECT TERMS Breast Cancer				15. NUMBER OF PAGES 186	
				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		

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

 10/24

PI - Signature Date

Table of Contents

Front Cover	1
SF298 Report Documentation Page	2
Foreword	3
Introduction	5
Collaborative Relationship with Army Medical Centers	5
Overview	6
Details of Research Performed	8
For Project A. The evaluation of the feasibility of interpreting mammograms on a soft copy viewing station.....	8
For Project B. Investigation of methods for digital storage and retrieval of mammograms	8
Direct Digital Images: Viewer Acceptance of Hard Copy Display	8
ROC Study of Direct Digital Images	9
For Project C. Identification of mammograms that are of substandard quality.....	9
For Project D. Identification of mammograms that require additional views prior to the release of the patient.....	11
For Project E. Tele-Digital Mammography	12
Summary	14
Bibliography	15
List of Appendices	17
Personnel.....	18

Computer Assisted Quality Control and Telemammography Army Report 1996

Introduction

This project, DAMD17-93-J-3015, started in December, 1992 and with a one year no cost extension is to end in December 1996. The purpose of this project was to develop systems to enable the integration of mammography into MDIS, to allow the acquisition of mammograms in facilities remote from the radiologist who will interpret the images, to develop automated methods of quality control of image quality and to develop methods for the teletransmission of digital mammograms.

This project contains five components.

- A. The evaluation of the feasibility of interpreting mammograms on a soft copy viewing station.
- B. Investigation of methods for digital storage and retrieval of mammograms.
- C. The identification of mammograms that are of substandard quality.
- D. The identification of mammograms that require additional views prior to the release of the patient.
- E. The demonstration of the effectiveness of telemammography.

As we complete this project, we have learned that digitized film mammograms displayed soft copy are not of sufficient diagnostic quality for primary interpretation of microcalcifications, that digital storage and retrieval of both digitized film and digitally acquired mammograms is feasible and that those acquired digitally are of probable diagnostic quality. We have developed methods of identifying mammograms that are of substandard quality and have demonstrated the effectiveness of telemammography when used with direct digital capture of breast images. We were unable to create a system that, at this time, is capable of automatically identifying women who need additional images.

We will discuss the work done on each of these sub-projects and indicate the portions that succeeded and those that did not succeed. After these discussions, we will describe a working system for telemammography and quality control suitable for obtaining mammograms at remote sites, their transmission and the remote interpretation of mammograms.

Collaborative Relationship with Army Medical Centers

Our original intention of working with Madigan Army Medical Center (MAMC) and Brooke Army Medical Center (BAMC) has been modified so that we are now working with Colonel Robert Shah, MD at BAMC and Colonel Ted Raia's designee, (Major Morgan Williamson, MD) at Walter Reed Army Medical Center (WRAMC). We also consult with and obtain advice from Major Donald Smith, MD at MAMC. Two WRAMC radiologists are participating in the receiver operating characteristic (ROC) reading study of digitally stored and transmitted mammograms. We have had discussions with radiologists at WRAMC, Bethesda Naval Medical Center, Major Mike Freckleton, MD at Wilford Hall AFMC, and Colonel Anna Chacko, MD at BAMC regarding our findings on digital mammography and teledigital mammography. We have also discussed at special meetings with General Russ Zajtchuk, MD and his assistant Major Paul Zimnick, MD the appropriate

utilization of digital telemammography for incorporation into the mammography van that they are currently working to develop. We have had discussions of the implementation of a clinical trial of digital telemammography at WRAMC and Ft Belvoir.

Overview

In summary, the following tasks have been accomplished:

A. The evaluation of the feasibility of interpreting mammograms on a soft copy viewing station. An evaluation of digitized film mammograms displayed on a workstation showed that radiologists had great difficulty in identifying microcalcifications and also misidentified film defects as microcalcifications.

B. Investigation of methods for digital storage and retrieval of mammograms. We have digitized and stored several thousand digitized mammograms and approximately 250 digitally acquired breast images. The retrieval of these images has been accomplished. The digitally acquired and retrieved images are of high quality and are probably of diagnostic quality. We are currently performing an ROC study to prove this. The digitized screen film mammograms are of lesser quality and do not appear adequate for primary diagnosis. They are clearly not acceptable when displayed soft copy.

C. The identification of mammograms that are of substandard quality.

1. A set of training cases has been assembled.

A. Set A is a set of images of geometric test objects that are available in direct digital form as well as digitized film. The accuracy of diagnosis of microcalcifications on these varies with the signal to noise ratio which is related to exposure. From this data, the exposure level at which information loss occurs has been calculated allowing one to determine whether there are regions within mammograms in which full information has not been captured. Data from this project was reported to SPIE Medical Imaging 1995. (Freedman 1995) A report of this work is attached in Appendix 1.

B. A program has been written to determine cases that are under- and overexposed. A set of 181 selected clinical mammograms were obtained from clinical files and digitized. While all of these were considered clinically acceptable originally, the radiologist interpreting them considered some of them somewhat overexposed or underexposed and listed them in a record book. Thirty images were underexposed, 83 normally exposed and 58 overexposed based on clinical criteria. In this set, the automatic program provided clear classification. Of the 83 with normal exposure, 76 were so classified. Of the 30 underexposed images all were classified as underexposed. Of the 58 overexposed images all were classified as overexposed. The report of this work is attached in Appendix 2.

2. We have developed segmentation algorithms for digitized screen film mammography that divide the image into regions of different pixel values. This will enable us to determine the average pixel value for under- and overexposed regions of the image.

3. Quality control procedures for the acquisition of digital images including mammograms using Fuji equipment have been established. Manuals for the Fuji AC-1, AC-3 and Fuji 9000 have been prepared. The manual for the AC-3 is attached (Appendix 3). These manuals provide methods for quality control of data acquisition.

D. The identification of mammograms that require additional views prior to the release of the patient

1. Our computer aided diagnosis program for microcalcifications has undergone further development. It can now identify 97% of microcalcification clusters in our test set. (Appendix 4) We are currently developing a user friendly interface so that this data can be easily displayed. We are currently performing a preclinical test of this system with the goal of determining the cause of false positive detections. Currently the system still has too many false positive detections to be of clinical use as an aid to the technologist for determining whether or not additional views will be necessary. The detection program, however, should be of use in aiding the radiologist in detection suspicious clusters of microcalcifications. Once we have completed our local preclinical trial, we expect to work with Major Morgan Williamson, MD at WRAMC and with Col Robert Shah, MD at BAMC in a larger clinical test of this algorithm.

2. We are concurrently working on a system for microcalcification classification--into calcification clusters that are associated with benign disease and those associated with malignancy. This work which was started in 1994 was substantially delayed because of instability with our high resolution film digitizer. We have replaced that system. Automated classification combined with computer aided diagnosis of microcalcifications would aid the technologist not only to know where there were clusters of microcalcifications that might benefit from magnification views, but would also allow her to eliminate obtaining the magnification views in some of those patients where the computer could clearly indicate that the microcalcifications were benign. (Appendices 5 & 6) At this stage in the development of these algorithms, it is still too early to implement them for use by technologists.

E. The demonstration of the effectiveness of telemammography.

In June, 1994, we first demonstrated teledigital mammography using digitized film and sending the image from Washington, DC to St. Paul, Minnesota to be printed. The quality was not adequate for diagnosis. In 1995, we were co-developers of a system that could transmit Fuji direct digital mammograms with what we believe to be adequate quality. Digital mammograms obtained at Georgetown were stored on a hard disk, transferred to a WORM optical disk, recalled to the hard disk, transferred to another hard disk, transferred via push over T-1 to another hard disk and then sent over the Internet to be stored on another hard disk and then laser printed. We believe that the quality is now appropriate, and will be spending the next few months optimizing the final image quality. We are currently performing an ROC study comparing digitally stored and locally transmitted mammographic images to the original screen film image.

In November 1995, at the Radiologic Society of North America (RSNA) annual meeting, we provided Internet access to our digital mammograms through the Inforad booth of our industrial collaborator, Analogic. At the recent presentation for Congress on Breast Cancer Imaging Improvements, we continuously transmitted images from Georgetown to the Hubert H. Humphrey Building of the Department of Health and Human Services (HHS), printing them on-line using a Polaroid Helios Printer. The quality is quite high and we believe it is suitable for diagnosis. An ROC study based on the Polaroid dry printing technology is planned for the near future. The ROC study using the 3M 969 wet laser is currently in process. This method has been demonstrated to General Russ Zajtchuk and Susan Blumenthal, MD, Assistant Secretary of Health for Women's Health, HHS.

Details of Research Performed

For Project A. The evaluation of the feasibility of interpreting mammograms on a soft copy viewing station.

An evaluation of digitized film mammograms displayed on a workstation showed that radiologists had great difficulty in identifying microcalcifications and also misidentified film defects as microcalcifications. For this project we collected 50 cases of cancerous and non-cancerous mammograms with screen film, direct digital and digitized film images. Digitization was performed at both 50 and 100 microns. Detailed testing of the 100 micron digitized screen film images showed that the four radiologists had a high degree of preference for the hard copy images. A detailed report is attached in Appendix 7.

For Project B. Investigation of methods for digital storage and retrieval of mammograms.

We have digitized and stored several thousand digitized mammograms and approximately 250 digitally acquired breast images. The retrieval of these images has been accomplished. The digitally acquired and retrieved images are of high quality and are probably of diagnostic quality.

We have been collecting a set of cases in which we have in each case the original screen film mammogram, the storage phosphor 100 micron direct digital mammogram obtained on the latest update of equipment and software, and the biopsy specimen radiograph. We currently have over 130 cases containing more than 30 proven cancers. The data for the direct digital mammograms is stored electronically, and each of the original screen film mammograms has been digitized and is available in 50 micron and 100 micron formats. Because of an important update in software that resulted in improved digital mammography acquisition, an earlier dataset was not used; the current dataset has been collected since March 1995. We delayed starting our ROC study because the laser imagers available to us were not considered adequate. We received our new 3M 969 laser and created with the aid of Analogic Corporation an adequate interface during the Summer of 1996 and are currently performing our ROC analysis.

Images are currently acquired using Fuji HR-V imaging plates and a Fuji 9000 Computed Radiography system. Based on problems we identified during beta testing of this system, it was modified and appears adequate for digital mammography. Images are automatically processed using Fuji's standard parameters and the unprocessed data is then stored on the Fuji 954 workstation. We reprocess the image data sets to meet our optimized image processing standards. The image data sets are then transferred for permanent storage to a Fuji optical disk drive to provide for long term optical disk storage. This image data can be transferred through an Analogic Data Acquisition System Manager (DASM) to an Analogic experimental workstation which serves both as a print server and a method of transferring the images over our internal network. The Analogic print server can (as recently developed) mimic the Fuji print parameters (except for Dynamic Range Control (DRC)) printing to a 3M laser printer. It can also send processed images for printing as TIFF files to other printers.

Direct Digital Images: Viewer Acceptance of Hard Copy Display

We have displayed comparison images from the Fuji computed radiography digital system and conventional screen film images in several meetings. The acceptability of these images has been considered high by many of those viewing the images at meetings. At the RSNA Annual Meeting, 1994, we demonstrated with a backlit display the original screen film images and CR digital images of the standard breast geometric test objects, three cancers

(5, 6, and 15 mm) manifested by microcalcifications, one 8mm cancer manifested by spiculated mass, and two microcalcification cases that on biopsy showed benign findings. We provided magnifying glasses and asked those who wished to indicate which images they preferred. Ninety-four percent of those who responded preferred the digital images of the test objects, 83% preferred the clinical microcalcification cases, and 57% preferred the one mass case we showed. We have shown a similar exhibit at the American Roentgen Ray Meeting in May 1995, where the exhibit received a bronze medal. The direct digital images appear to be acceptable to many of those viewing them.

ROC Study of Direct Digital Images

We have selected from our collected set of digitally acquired mammograms 24 biopsy proven cancer cases, 24 benign biopsy cases and 50 cases selected by radiologists as being normals for incorporation in our ROC study. In each case we have both conventional and digital mammograms. These cases have been randomized and radiologists are now providing interpretations of these comparing (in separate reading sessions one month apart or greater) interpretations of screen film and digital mammograms. The first readers results are available and showed Az screen film = 0.7373; Az digital = 0.7646 when tested for detection of cancer. When tested as a method of discrimination between cancer and benign lesions that were detected, the Az of screen film = 0.5743. The Az of the digital system = 0.7412. We believe that when this ROC study is completed in November 1996 that it will show that digitally acquired, stored, and retrieved mammograms are equivalent to screen film mammograms.

For Project C: Identification of mammograms that are of substandard quality.

1. A set of training cases has been assembled.

A. Set A is a set of images of geometric test objects that are available in direct digital form as well as digitized film. The accuracy of diagnosis of microcalcifications on these varies with the signal to noise ratio which is related to exposure. From this data, the signal to noise ratio (SNR) at which information loss occurs can be calculated allowing one to determine whether there are regions within mammograms in which full information has not been captured. Data from this project was reported to SPIE 1995 Medical Imaging. (Appendix 1)

This report demonstrated that as one decreased the exposure, that the ability to detect the smallest microcalcifications was decreased. This is summarized in the following chart:

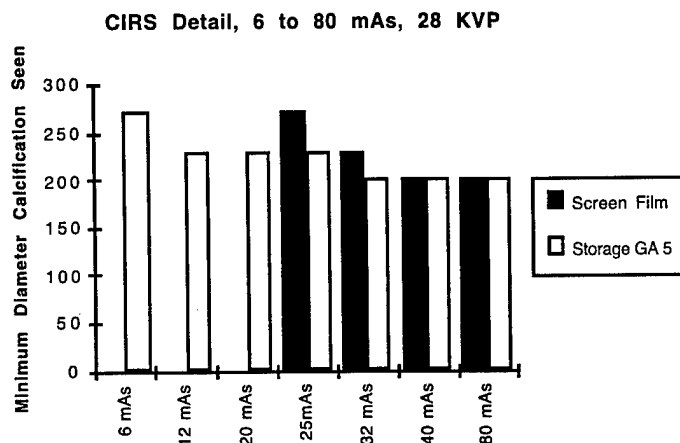


Chart 1: This chart demonstrates the smallest sized test details visible on the CIRS Detail Phantom as exposure is varied, comparing two systems: conventional screen film mammography and Fuji Digital Storage Phosphor radiography, processed in Sensitivity Mode. At low exposure levels, the digital system allows the detection of simulated microcalcifications that cannot be seen in the conventional system. This is analogous to detecting microcalcifications in radiodense regions of the breast. (Appendix 1)

The major factors affecting the detectability of small details in a geometric test object are signal to noise ratio and contrast. Image processing can be used to change the contrast scale, but increasing the contrast also increases the contrast and visibility of the image noise. By calculating the signal to noise ratio in these images, one can determine whether or not full information about small microcalcifications has been obtained. Because SNR is an easily calculated quantity, one will then be able to use a computer to look at regions of the mammogram, segmented as discussed below, to determine whether or not exposure parameters are appropriate.

B. A program has been written to determine cases that are underexposed and overexposed. A set of 181 selected clinical images was obtained from clinical files and these mammograms were digitized. While all of these were considered clinically acceptable originally, the radiologist interpreting them considered some of them somewhat overexposed or underexposed and listed them in a record book. Thirty were under exposed, 83 normally exposed and 58 over exposed based on clinical criteria. In this set, the automatic program provided clear classification. Of the 83 with normal exposure, 76 were so classified. Of the 30 underexposed images all were classified as underexposed. Of the 58 overexposed images all were classified as overexposed. The report of this work is attached in Appendix 2. In this study, the program detected all under- and overexposed images correctly, but did assign 7 of the 83 normal images to the overexposed category. Since all of these images were considered within the acceptable range at the time of initial reading, though noted to be a little dark or a little light, the system is both sensitive and relatively specific. It should be able to serve as a good guide for the automated assessment of mammograms obtained at remote sites.

2. We have developed segmentation algorithms for digitized screen film mammography that divide the image into regions of different pixel values. This will enable us to determine the average pixel value for different regions of the image. The segmentation

system is adjustable and can use different thresholds. It works by looking at regional pixel values to determine margins within an image. In Appendix 8 , I have attached images from this segmentation program and a brief description. This program is used to identify regions of underexposure.

3. As our work with digitized film progressed, it became clear that teletransmission of digitized screen film obtained mammograms would be of limited utility because of the introduction of artifacts simulating microcalcifications and the obscuration of microcalcifications that can occur. Our report on this is in Appendix 7. Successful teletransmission of mammograms would require the use of digitally acquired mammograms. The system we chose to use for development is a storage phosphor based system -- the Fuji 9000. We discovered that if this machine were not maintained at a high level of quality, that artifacts could result resembling calcifications. A very high level of quality control was therefore necessary. We sent one of the faculty scientists to the Fuji factory training course to learn everything that he could about the machine and had him come back to write a quality control manual for use with this system.

Quality control procedures for the acquisition of digital images including mammograms using Fuji equipment have been established. Manuals for the Fuji AC-1, AC-3 and Fuji 9000 have been prepared. The manual for the AC-3 is attached. These manuals provide methods for quality control of data acquisition. If one is to produce high quality digital mammograms, one must have methods for assuring that the quality of image acquisition is high and that the machine used for this acquisition is properly maintained. Our experience in using storage phosphor mammography is that it requires very careful attention to machine cleanliness and that the functions of the machine must be checked on a regular basis. Attached are two articles related to QC of storage phosphor devices that we have recently published. (Appendices 9,10) The more complete manual for machine QC is also attached. (Appendix 2)

For Project D. Identification of mammograms that require additional views prior to the release of the patient.

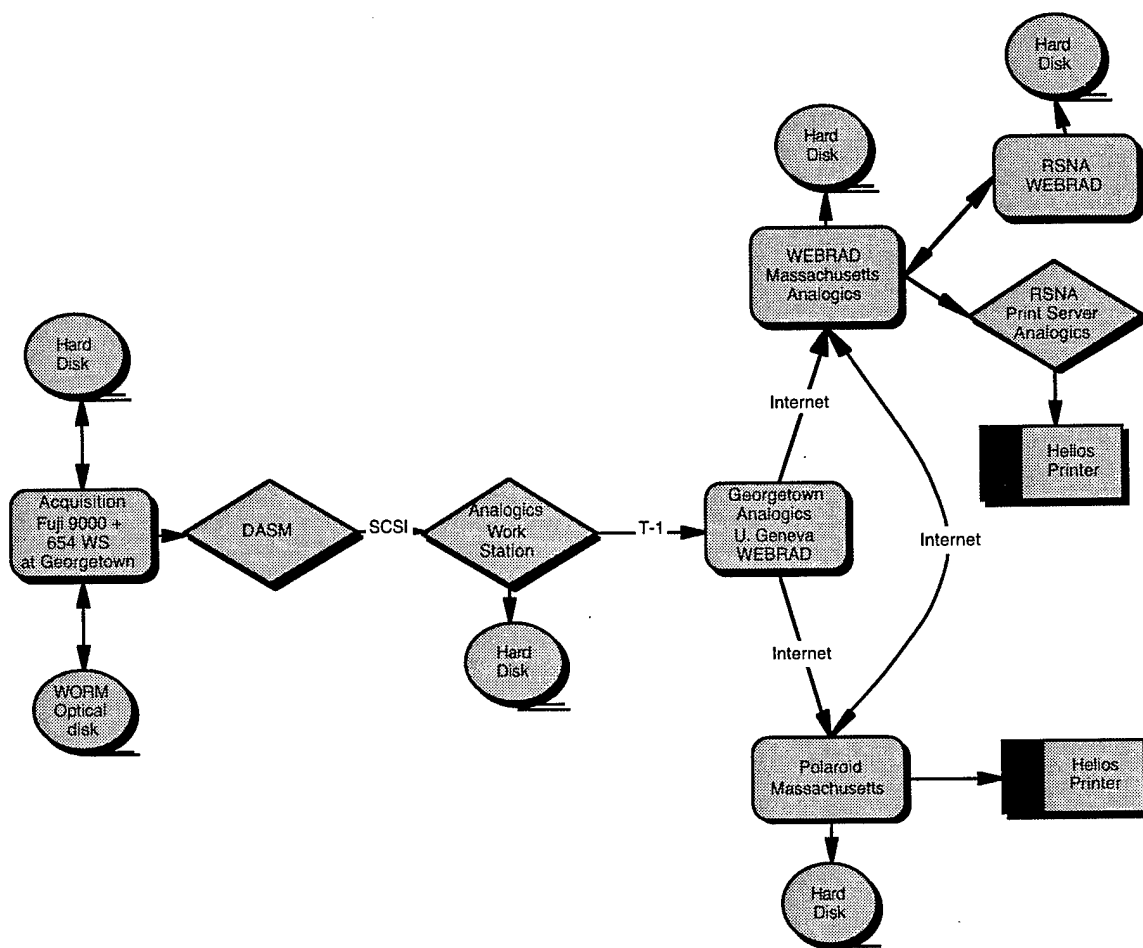
1. Our computer aided diagnosis program for microcalcifications has undergone further development. It can now identify 97% of microcalcification clusters in our test set. We are currently developing a user friendly interface so that this data can be easily displayed. We are preparing to test this system in a preclinical setting working with Col Ted Raia, MD at Walter Reed Army Medical Center and with Col Robert Shah, MD at Brooke Army Medical Center. Our original goal was to develop a program that would enable the technologist to know when a magnification view was needed. Currently, the computer aided diagnosis system has been tuned for screen film mammograms digitized at 100 micron pixel size. While our current program has a high true positive detection rate, it still has an average of 1 false positive detection per image. We are currently doing a careful analysis of the types of findings that are causing the false positives and have submitted an abstract to present a paper at SPIE Medical Imaging 97. A copy of this abstract is attached. (Appendix 2) What we are finding in our review, so far, of more than 200 images, is that real findings are causing the false detections. These include groups of several film defects, benign microcalcifications, grid artifacts, film scratches, and what we believe are dilated terminal ductal units in the breast. These dilated terminal ductal units are a common site of benign calcification formation and closely resemble the appearance of calcification clusters. We believe that improving our knowledge of the causes of false positives will enable us to develop improved algorithms and improved quality of our screen film mammograms that will help us avoid these problems. At present, the system would result in too many extra films if technologists were to use it as a guide.

We have written several articles describing our method of microcalcification detection. These are included in Appendices 4,1 1.

2. It is clear that a system that identifies calcifications that radiologists can easily classify as being benign would, in the end, not be useful in telemammography. We therefore have been working on systems to aid in microcalcification classification--into calcification clusters that are associated with benign disease and those associated with malignancy. This work which was started in 1994 was substantially delayed because of instability with our high resolution film digitizer. We have replaced that system. Automated classification combined with computer aided diagnosis of microcalcifications would aid the technologist not only to know where there were clusters of microcalcifications that might benefit from magnification views, but would also allow her to eliminate obtaining the magnification views in some of those patients where the computer could clearly indicate that the microcalcifications were benign. We have run two different datasets through our developing program. In the first dataset of biopsy specimen radiographs, the system performed moderately well in distinguishing benign and malignant cases. This has been previously reported. (Appendix 5) We have just completed testing a second, more difficult dataset of digitized screen film mammograms. In this set the radiologists were not able to distinguish benign from malignant (all had gone to biopsy as suspicious lesions after diagnostic workup). The computer program was not able to distinguish benign from malignant in these cases either. This was presented in 1996. (Appendix 6) We are now working to test a set of cases that radiologists performing screening mammography considered suspicious, but which were then classified into benign or malignant categories by diagnostic mammograms. These results have been submitted for presentation in February 1997, but we do not know yet whether the paper has been accepted.

For Project E: Teledigital Mammography

In June 1994, we first demonstrated teledigital mammography using digitized film and sending the image from Washington, DC to St. Paul, Minnesota to be printed. The quality was not adequate for diagnosis. In 1995, we were co-developers of a system that could transmit Fuji direct digital mammograms with what we believe to be adequate quality. The schematic diagram of this project follows:



Digital mammograms obtained at Georgetown were stored on a hard disk, transferred to a WORM optical disk, recalled to the hard disk, transferred to another hard disk, transferred via push over T-1 to another hard disk and then sent for over Internet to be stored on another hard disk and then laser printed. During the annual meeting of the Radiological Society of North American, November 1995, we demonstrated teledigital mammography over the Internet transmitting images from Boston to the Radiologic Society of North America and printing the images in the Inforad booth of our industrial collaborator, Analogic. We believe the quality is now appropriate and will be spending the next few months optimizing the final image quality. Once image quality is optimized, we will perform an ROC analysis comparing the original mammogram obtained with screen film to the direct digital mammogram teletransmitted images in a set of proven cases. The final images resulted from a combination of technology developed by Fuji, Georgetown, Analogic Corp. and Polaroid.

In October 1996, with further developments, we were able to demonstrate on-line digital teledigital mammography from Georgetown to the Hubert H. Humphrey Building of HHS in Washington, DC, printing images pulled over the Internet. The image quality appears to be of diagnostic quality.

The system that we designed and helped to build now transmits 100 micron pixel digital mammograms, but could be easily converted to a 50 micron system should that become the standard for digital mammography and should the problems of image display be solved.

Summary

The goal of this project was to develop methods to improve the ability of mammograms to be obtained at remote sites and to develop methods for their teletransmission to a central site of excellence in the interpretation of mammography. We were unable to accomplish this using digitized screen film mammograms, but have succeeded in doing this with direct digital mammograms. The ROC study that will prove this is currently underway with the results of the initial reader being favorable with Az screen film = 0.7373. Az digital = 0.7646 when tested for detection of cancer. When tested as a method of discrimination between cancer and benign lesions that were detected, the Az of screen film = 0.5743. The Az of the digital system = 0.7412. We have demonstrated the effectiveness of digital storage and retrieval for which the same ROC study will provide the proof. We have demonstrated the feasibility of teletransmission of images transmitting them over the Internet from Georgetown to Boston to the Radiologic Society of North America annual meeting in Chicago and separately from Georgetown over the Internet to the Hubert H. Humphrey Building of the HHS in Washington, DC. We have defined quality control procedures for this system. We have been able to build a system that identifies under- and overexposed mammograms based on digitization of the mammographic films. We have been unable to develop a clinically useful method of directing technologists as to which patients need additional images during their mammogram based on the detection of microcalcifications. This still remains a human rather than a computer based task.

Bibliography

Relevant Publications

1. Dawkins T, Freedman M, Lo S-C B, Mun S K: 35 Micron CCD Based Film Digitizer for Mammography. SPIE Poster. SPIE Medical Imaging paper 1897-53 (February 1993).
2. Freedman M, Pe E, Nelson M, Lo S-C B, Mun S K: Digital Mammography: Demonstration of a Method of Achieving Adequate Resolution on a Storage Phosphor System. CAR 93, 7th International Symposium Berlin (June 24-26, 1993); 783pp.
3. Fields F, Freedman M, Lo S-C B, Zuurbier R, Nelson M, Mun S K: Comparison of Conventional Film Screen Magnification Mammography and Electronic Magnification. Proc. SPIE: 2167, 682-689. Medical Imaging 1994: Image Processing Murray H. Loew, Ed.
4. Freedman M, Pe E, Zuurbier R, Katial R, Jafroudi H, Nelson M, Lo S-C B, Mun S K: Image Processing in Digital Mammography. SPIE: Medical Imaging, Vol. 2164 (1994); 537-554pp.
5. Freedman M, Steller D, Jafroudi H, Lo S-C B, Zuurbier R, Katial R, Hayes W, Wu Y C, Lin J-S J, Mun S K: Digital Mammography: Effects of Decreased Exposure. SPIE: Medical Imaging (1995). Paper 2432-49.
6. Freedman M, Steller D, Jafroudi H, Lo S-C B, Zuurbier R, Katial R, Hayes W, Wu Y C, Lin J-S J, Steinman R, Tohme W G, Mun S K: Digital Mammography: Tradeoffs Between 50- and 100-micron Pixel Size. Proc. SPIE: Medical Imaging 1995. Paper 2432-09. pp 114-125/ Physics of Medical Imaging, Richard L Van Metter; Jacob Beutel; eds.
7. Y. Chris Wu, Richard Patt, Matthew Freedman, Seong Ki Mun . Three dimensional image visualization and analysis in breast MR imaging for diagnosis of breast cancer. SPIE Medical Imaging 1996. In press.
8. S.-C. Benedict Lo, Huai Li, Matthew Freedman, BH Krasner, SK Mun Optimization of wavelet decomposition and feature-guided image compression for mammography.. SPIE Medical Imaging 1996. In press
9. Robert Jennings, FDA , Hamid Jafroudi, Georgetown; FDA: R Gagne; NIH: James Vucich. Georgetown: D. Artz, M. Freedman, SK Mun Storage phosphor-based digital mammography using a low dose x-ray system optimized for screen-film mammography system. SPIE Medical Imaging 1996. In press
10. Hamid Jafroudi, SCB Lo, MT Freedman, SK Mun. Dual Energy in mammography: feasibility study. SPIE Medical Imaging 1996. In press
11. Matthew Freedman, Dot Artz, H Jafroudi, J. Hogge, RA Zuurbier, R. Katial, SK Mun. Digital mammography in the radiodense and complex pattern breast. Proc. SPIE 2701, 783-793. Medical Imaging 1996. Image Processing, Murray H. Loew, Kenneth M Hanson eds.
12. O.Tsujii, YC Wu, MT Freedman, A Hasegawa, RA Zuurbier, SK Mun. Classification of microcalcifications in digital mammograms for the diagnosis of breast cancer. SPIE Medical Imaging 1996. In press
13. Matthew Freedman, Dot Artz, H Jafroudi, J. Hogge, RA Zuurbier, R. Katial, SK Mun Digital Mammography: an evaluation of the shape of microcalcifications. SPIE 2708, 626-632. Medical Imaging 1996. Physics of Medical Imaging. Richard L Van Metter and Jacob Beutel, eds.

14. Freedman MT, Artz DS, Mun SK: Image processing in digital mammography: the optimum image for each woman's breasts. pp. 1/1-1/3. IEEE Colloquium on Digital Mammography. 1996. The Institution of Electrical Engineers, London.

Relevant Exhibits

1. Dawkins T, Freedman M, Lo S-C B, Mun S K: 35 Micron CCD Based Film Digiter for Mammography. SPIE Poster. SPIE Medical Imaging Poster 1897-53 (February 1993).
2. Freedman M, Pe E, Nelson M, Lo S-C B, Mun S K: Digital Mammography: Demonstration of a Method of Achieving Adequate Resolution on a Storage Phosphor System. CAR 93, 7th International Symposium, Berlin, Germany (June 24-26, 1993); 783pp.
3. Fields F, Freedman M, Lo S-C B, Zuurbier R, Nelson M, Mun S K: Comparison of Conventional Film Screen Magnification Mammography and Electronic Magnification. SPIE: Medical Imaging (February 1994). Paper 2167-67.
4. Freedman M, Pe E, Zuurbier R, Katial R, Jafroudi H, Nelson M, Lo S-C B, Mun S K: Image Processing in Digital Mammography. SPIE: Medical Imaging, Vol. 2164 (1994); 537-554pp.

Relevant Abstracts

1. Freedman M, Zuurbier R, Pe E, Jafroudi H, Mun S K, Lo S-C B: Image Processing in Digital Mammography (abstract). Radiology (1993); 189P:408.
2. Freedman M T, Steller D E, Hasegawa A, Zuurbier R A, Wu Y C, Smith D V, et al: image Processing in Digital Mammography (abstract). Radiology (1994); 193(P):474.
3. Freedman M T, Steller D E, Zuurbier R A, Jafroudi H, Mun S K: Experimental Digital Mammography in the Detection of Microcalcifications 300 μ m and smaller (abstract). Radiology (1994); 193(P):422.

Relevant Presentations

1. Dawkins T, Freedman M, Lo S-C B, Mun S K: 35 Micron CCD Based Film Digiter for Mammography. Poster. SPIE: Medical Imaging, Paper 1897-53 (February 1993).
2. Freedman M, Pe E, Zuurbier R, Katial R, Jafroudi H, Nelson M, Lo S-C B, Mun S K: Image Processing in Digital Mammography. SPIE: Medical Imaging, Vol. 2164 (1994); 537-554pp.
3. Freedman M T, Steller D E, Hasegawa A, Zuurbier R A, Wu Y C, Smith D V, et al: Image Processing in Digital Mammography (abstract). Radiology (1994); 193(P):474.
4. Freedman M T, Steller D E, Zuurbier R A, Jafroudi H, Mun S K: Experimental Digital Mammography in the Detection of Microcalcifications 300 μ m and smaller (abstract). Radiology (1994); 193(P):422.

Appendices

Digital Mammography: The Effects of Decreased Exposure

Quality Control in Digital Mammography: Automatic Detection of Under and Overexposed Mammograms

Quality Control for Fuji AC-3 Computed Radiography

Gain of Using Irreversible over Error-Free Data Compression in Digital Radiography

Classification of Microcalcifications in Radiographs of Pathological Specimen for the Diagnosis of Breast Cancer

Classification of Microcalcifications in Digital Mammograms for the Diagnosis of Breast Cancer

Implementation and Testing of Digital Mammography for MDIS Environment

Breast Segmentation Images

Quality Control of Storage Phosphor Digital Radiography Systems

Quality Control for Storage Phosphor Radiography Machines

Detection of Clustered Microcalcifications Using Fuzzy Classification Modeling and Convolutional Neural Network

Staff Supported on Contract # DAMD17-93-J-3015DAR

Matthew Freedman, M.D.
Seong Ki Mun, PhD
Betty Ann Levine
Jyh-Shyan Lin, PhD
Wendelyn Hayes, D.O.
Richard Steinman, MD
Carole MacDonald, RT
Brian H Krasner, Ph.D.
Gary Norton, R.T.
Susan Kirby
Dorothy Stellar Artz, R.T.
Jeff Collmann, Ph.D.
Akira Hasegawa, Ph.D.
Kevin Legendre, M.D.
Yue Wang, PhD
Rebecca Zuurbier, M.D.
Hamid Jafroudi, Ph.D.
Kevin Cleary, Ph.D.
Andrew Mun
Lydia Lantang

Digital mammography: The effects of decreased exposure

Matthew Freedman MD MBA, Dot Steller RT(R), Hamid Jafroudi PhD, S.-C. Benedict Lo PhD, Rebecca A. Zuurbier MD,
Raj Katial RT(R), Wendelin Hayes DO, Y. Chris Wu PhD, Jyh Shien Lin MS, Seong Ki Mun PhD

Division of Imaging Science and Information Systems and Division of Breast Imaging, Georgetown University School of
Medicine, 3800 Reservoir Road NW, Washington, DC USA 20007

Keywords: Digital Mammography Exposure

ABSTRACT

It has been stated that digital mammography will reduce the exposure required for mammography. This poster explores the effects of decreased exposure on the information present in digital mammography. In general, the digital system performed better than screen film mammography with lower exposures. With the usual exposures used for screen film mammography, performance was equal. With high exposures sufficient to result in a dark film (OD 1.5), the digital system performed better than screen film with very small test objects.

Proposals have been made to decrease the tube loading required for slot scanning devices by increasing KVP. This would result in their being less object contrast due to the decreases in the absorption coefficient of calcium compared to water at higher KVP. This poster looks at the potential for correcting the loss in object contrast that would result from the use of high contrast look up tables. It was found that in the tested system, one could restore the information in one of the two test objects used (but not the other), but that the image processing methods used would result in an image that radiologists would probably find inadequate for interpretation.

METHODS

Tested System

Images for this series of experiments were obtained using a GE-CGR (Milwaukee, Wisconsin) 600 T dedicated mammography machine. The machine is in clinical use in an American College of Radiology approved facility and meets the quality standards for this approval. The images obtained at 28 KVP at varying mAs were obtained with a molybdenum target and filter. Those obtained at various KVPs were obtained with a molybdenum target and an aluminum filter. The mammographic screen film system used is Fuji (Tokyo) Kyokko UM Fine Screen with Fuji UM-MA-HC film. The digital system tested is the Fuji 9000 using high resolution imaging plates (HR-V).

Two test objects were used: the CIRS Detail (square) test object (Norfolk Virginia) and the CDMAM test object (Nuclear Associates, Carle Place, New York). The CDMAM test object was placed on top of four cm of acrylic.

Image Processing Optimization

Image processing methods are those included in the Fuji 9000 system. The optimized image processing settings for the detection of small objects were calculated through the application of the mathematical technique of multiple iterative response surfaces in geometric test objects.

Exposure reaching the Screen Film Cassette and Storage Phosphor Cassette

CIRS Detail Phantom

CIRS detail

KVP	mAs	mR pre grid	mR post grid	ODSF	KVP	mAs	mR pre grid	mR post grid	ODSF
28	6	15.7	6.57	0.16	22	5.9	2.2	0.83	0.4
28	12	31.6	13.22	0.18	28	3.8	5.9	2.47	1.22
28	20	42.3	17.70	0.21	36	6.9	28.7	13.23	1.5
28	25	53.0	22.18	0.25					
28	32	85.0	35.56	0.29					
28	40	106.6	44.60	0.42					
28	80	213.8	89.46	1.13					

CDMAM test object with 4 cm of added acrylic between test object and receptor

KVP	mAs	mR pre grid	mR post grid	ODSF	KVP	mAs	mR pre grid	mR post grid	ODSF
28	6	2	0.84	0.16	22	67.7	<1.0	<0.38	0.4
28	12	3.9	1.63	0.18	28	8.5	2.5	1.05	1.22
28	20	5.4	2.26	0.21	36	7.1	7.1	3.27	1.5
28	25	6.7	2.80	0.25					
28	32	10.9	4.56	0.29					
28	40	13.6	5.69	0.42					
28	80	27.3	11.42	1.13					

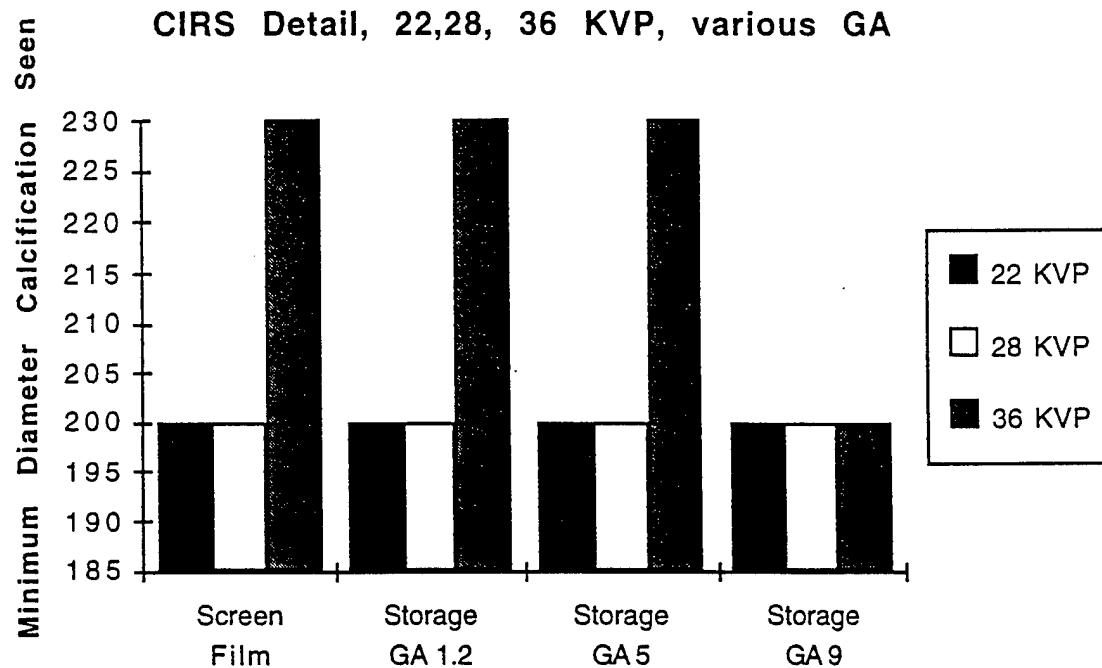
OD SF = optical density on screen film image. Pre-grid = exposure above grid. Post grid = exposure below grid corrected for changes in grid absorption with changes in KVP.

FINDINGS

THE EFFECT OF INCREASES IN KVP

When screen film mammography images and digital images obtained at different KVPs are compared, one can see that the higher KVP results in a decrease in the detection of small objects that cannot be retrieved with the image processing methods available to the authors except by the use of a very high contrast look up table (GA=9). This high contrast look up table results in images that are likely unacceptable for interpretation.

Findings in the CIRS Detail Test Object

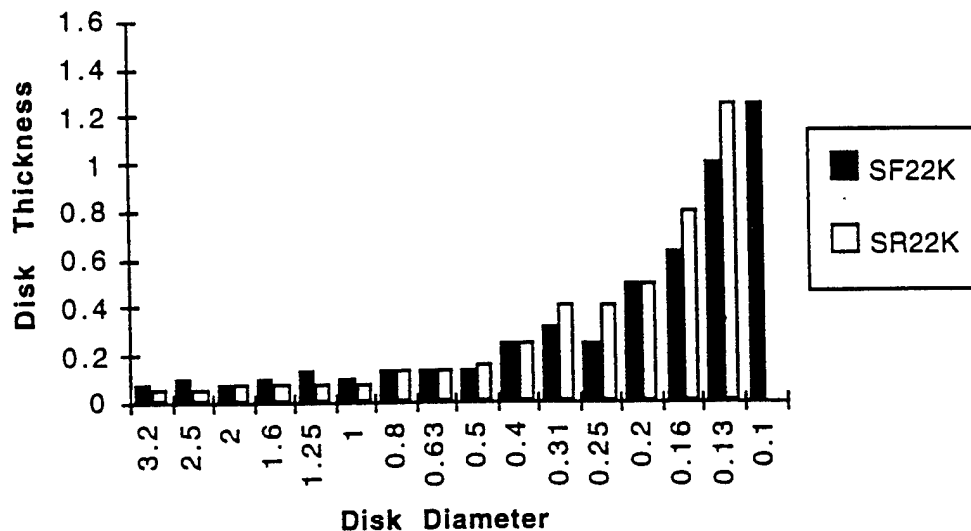


This chart shows the results for the CIRS Detail (square) test object. With this test object, no difference is seen between 22 and 28 KVP, with the smallest object measuring 200 microns. At 36 KVP, screen film, and storage phosphor processed with a GA of 1.2 and 5 did not allow retrieval of the smallest detail. At a GA of 9, at 36 KVP, the 200 micron objects could be seen, but this very high contrast image results in an unreadable appearance to the mammogram. This is because the range of optical densities in the original image exceeds the optimal range for this look up table resulting in images with regions that are maximally black or maximally white, resulting in loss of detail. In addition, the heel effect is exaggerated with these high contrast look up tables so that the entire breast cannot be visualized at the same time. (GA = gradient angle of the look up table. Storage = storage phosphor radiography.)

Findings in the CDMAM Test Object

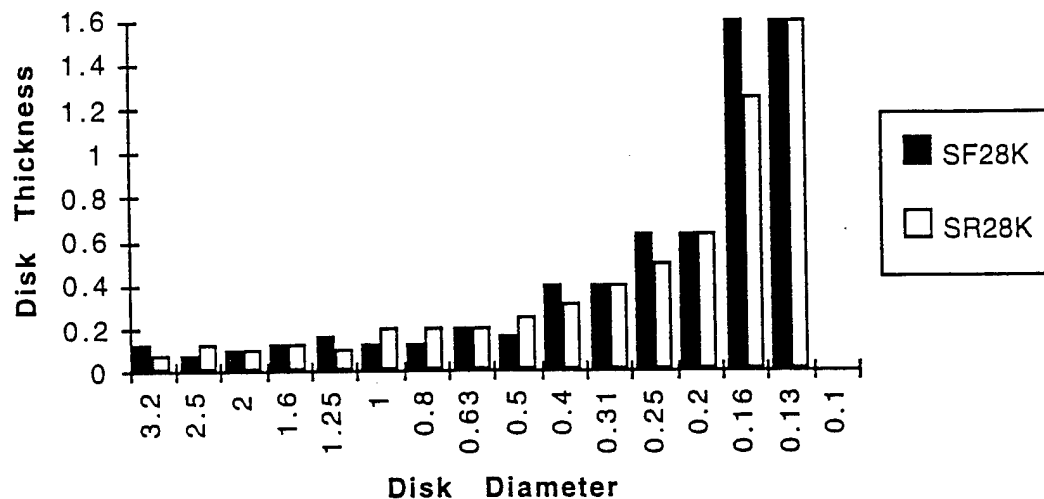
The Nuclear Associates CDMAM test object also was used to compare screen film and storage phosphor radiographs at various KVPs.

Screen Film vs Storage Phosphor 22 KVP, 4 Layers, LFS



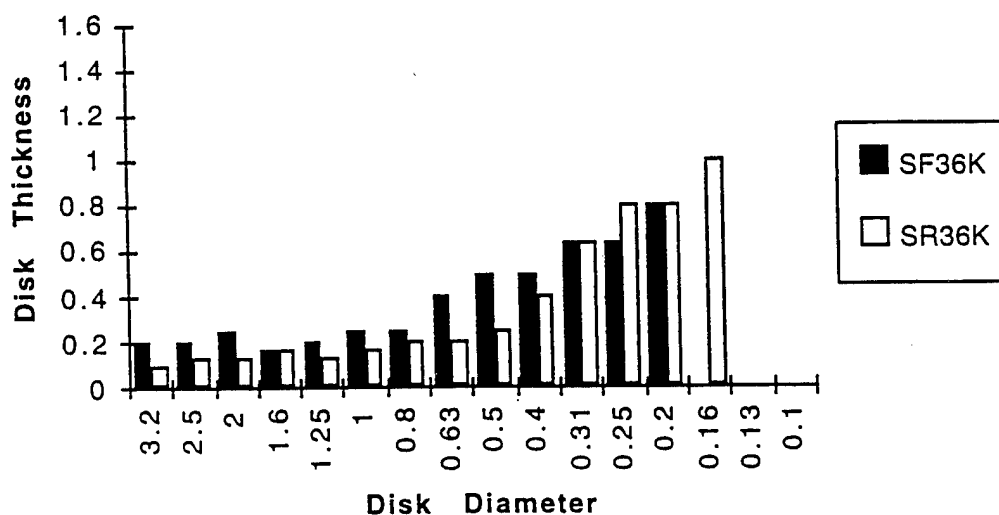
This chart demonstrates that at 22 KVP, the screen film system shows smaller objects and for the smaller objects shows them at lower contrast than the digital system. (SF= Screen Film. SR= Storage Phosphor Radiography. K = KVP)

Screen Film vs Storage Phosphor 28 KVP, 4 Layers, LFS



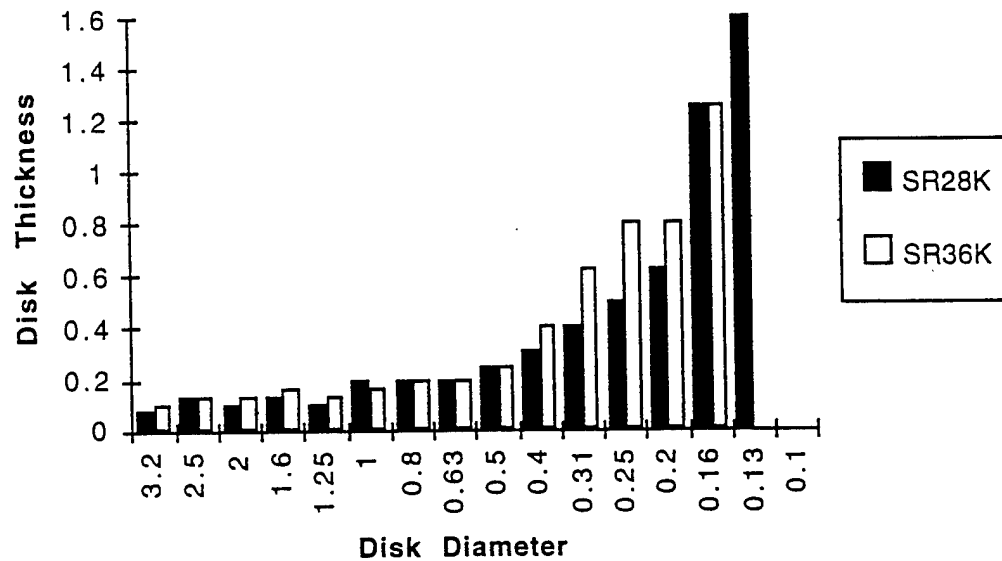
This chart shows that screen film and storage phosphor demonstrate the same size objects at approximately the same contrast levels when obtained at 28 KVP. (SF= Screen Film. SR= Storage Phosphor Radiography. K = KVP)

Screen Film vs Storage Phosphor 36 KVP, 4 Layers, LFS



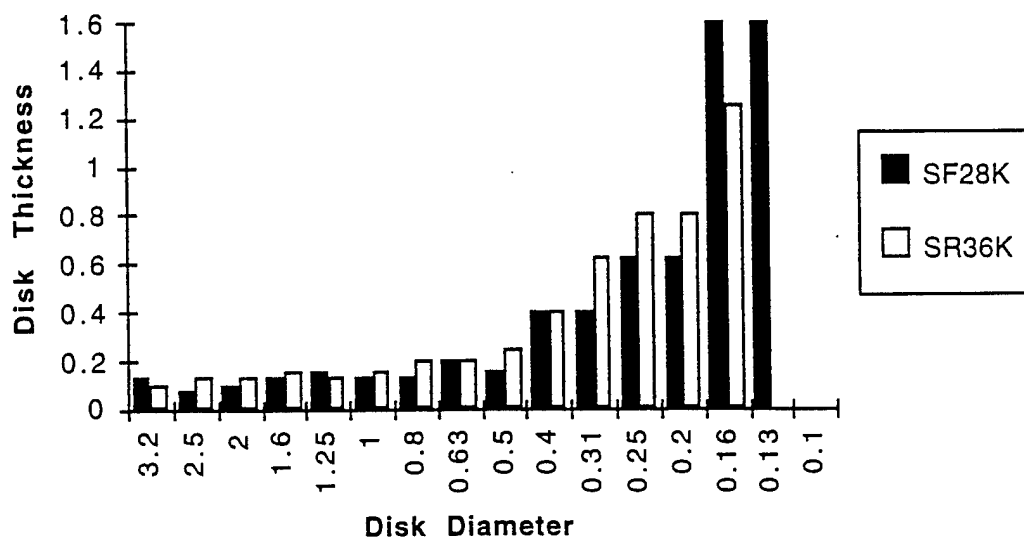
In this chart, the CDMAM phantom demonstrates smaller objects than screen film and shows them at lower contrast (thinner disk). (SF= Screen Film. SR= Storage Phosphor Radiography. K = KVP.)

Storage Phosphor at 28 and 36 KVP



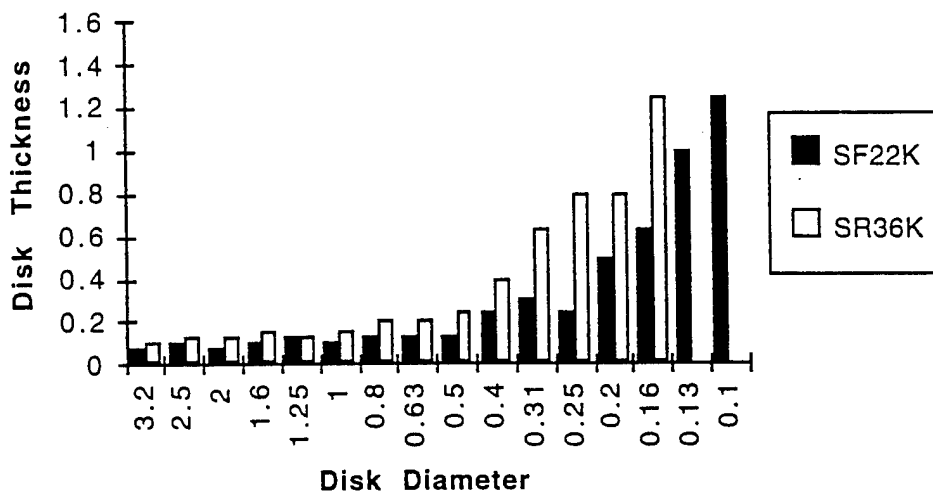
In this chart, one can see that storage phosphor (SR) at 28 KVP performs better than SR at 36 KVP. One can see the smallest (130 micron) object only at 28 KVP and for the other small objects, one sees them at lower contrast at 28 KVP. (SR= Storage Phosphor Radiography. K = KVP.)

Screen Film at 28 KVP vs Storage Phosphor at 36 KVP



This chart demonstrates that one can see the 130 micron object at 28 KVP using screen film, but not with SR at 36 KVP. For most of the object in the lower third of the disk sizes shown, screen film detects them at slightly lower contrast than SR. (SF= Screen Film. SR= Storage Phosphor Radiography. K = KVP.)

Screen Film at 22 KVP vs Storage Phosphor at 36 KVP



This chart demonstrates that screen film images at 22 KVP show the 100 and 130 micron objects which cannot be seen at 36 KVP with SR. Also for all of the other size objects, screen film allows their detection at lower contrast. The best screen film images were obtained at 22 KVP. (SF= Screen Film. SR= Storage Phosphor Radiography. K = KVP.)

Conclusion regarding KVP

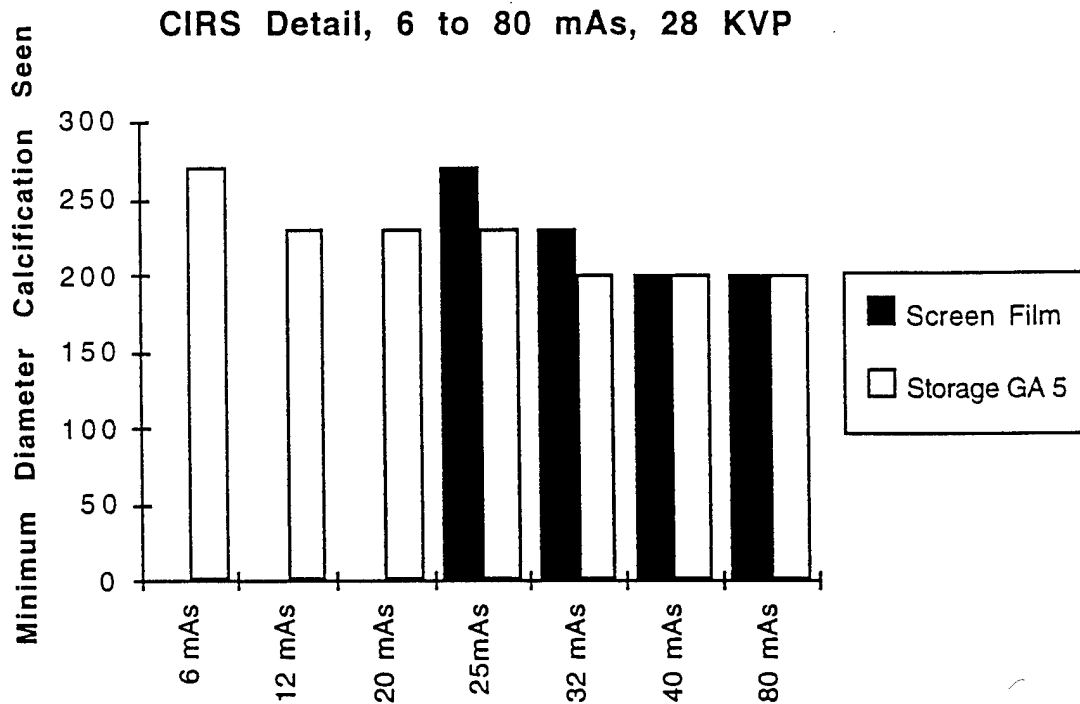
Although SR with image processing allows more to be seen at 36 KVP when compared to screen film, it does not equal digital or screen film at 22 or 28 KVP. Screen film images at 22 KVP show the smallest objects in this test. Within this test system, with both the CIRS and CDMAM test objects, one could not reclaim the information regarding the smallest objects when 36 KVP was used; there was better conspicuity for most objects with both the screen film and digital images obtained at 22 or 28 KVP.

DIFFERENCES IN MAS AT 28 KVP

Tests demonstrated that the digital system performed better at low mAs and at the highest level of mAs. In the intermediate range of exposures, the systems performed equivalently.

Tests with the CIRS Detail Test Object

These charts demonstrate the findings as mAs is varied at 28 KVP.

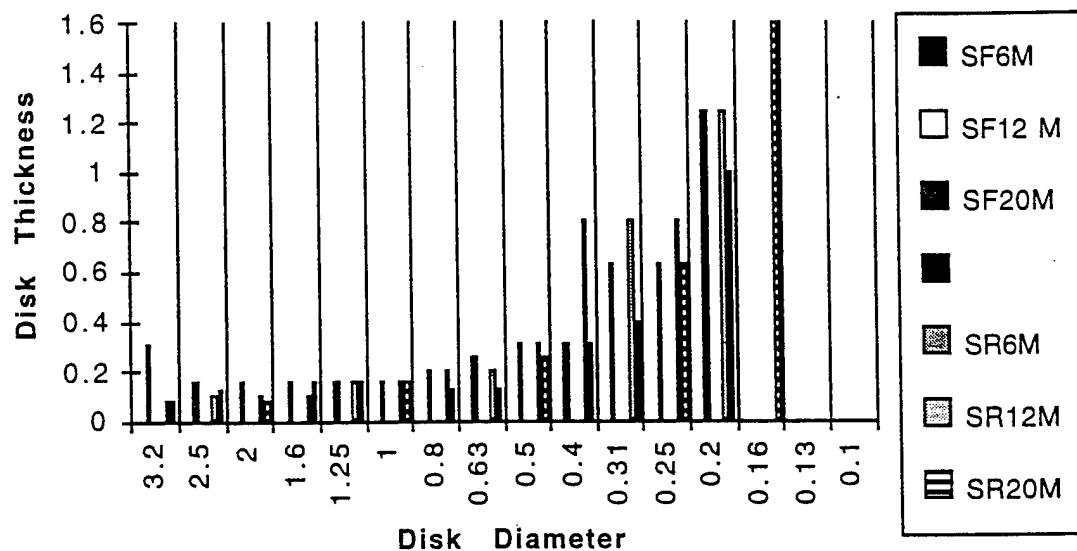


This chart demonstrates that at mAs less than 25, while the storage phosphor system could demonstrate objects, the screen film failed to demonstrate any of the objects. At 25 and 32 mAs, the storage phosphor demonstrated smaller objects than screen film. At 40, and 80 mAs, the two systems performed equivalently. (Storage = Storage Phosphor Radiography. GA = gradient angle of the look up table.)

Tests with the CDMAM Test Object

This test object shows similar results, with the digital system performing better at low mAs and the system performing equivalently at higher mAs.

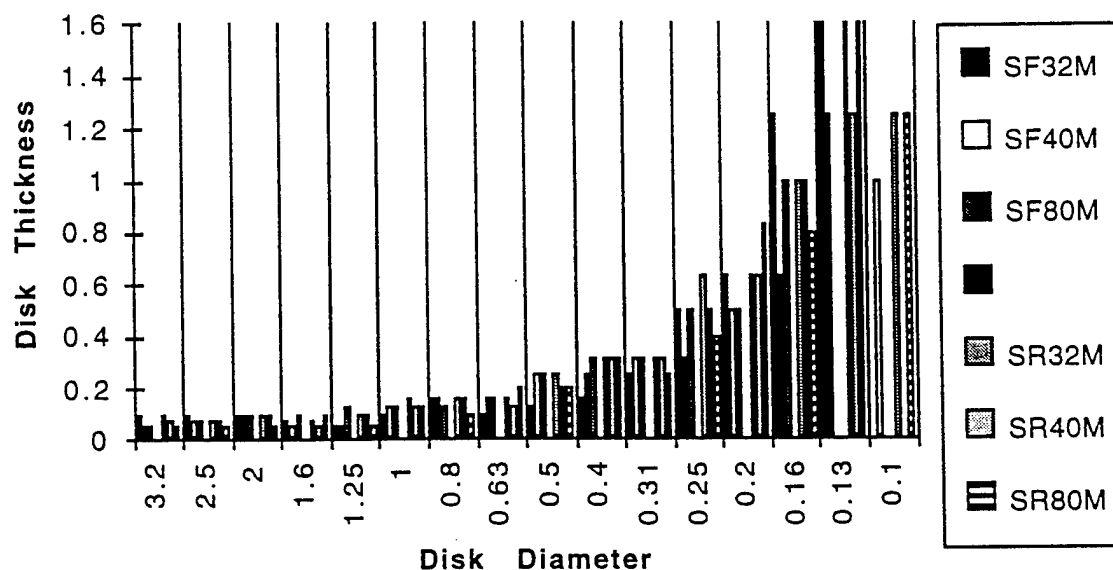
6-20 mAs, Screen Film vs Storage Phosphor, 28 KVP



This chart demonstrates the effect seen as the mAs is increased from 6 to 20 mAs. Within each block, those values to the left are from screen film and those to the right are the digital values. Less is seen with the screen film than with the digital images in this group. (SF= Screen Film. SR= Storage Phosphor Radiography. M = mAs)

With higher mAs, the systems perform similarly:

32 to 80 mAs Screen Film vs Storage Phosphor 28 KVP



This chart demonstrates that the screen film and digital system are performing similarly with the larger test objects. With the smallest diameter objects, the screen film and digital systems allow their detection, but the screen film system allowed it to be seen at slightly lower contrast in this test. (SF= Screen Film. SR= Storage Phosphor Radiography. M = mAs)

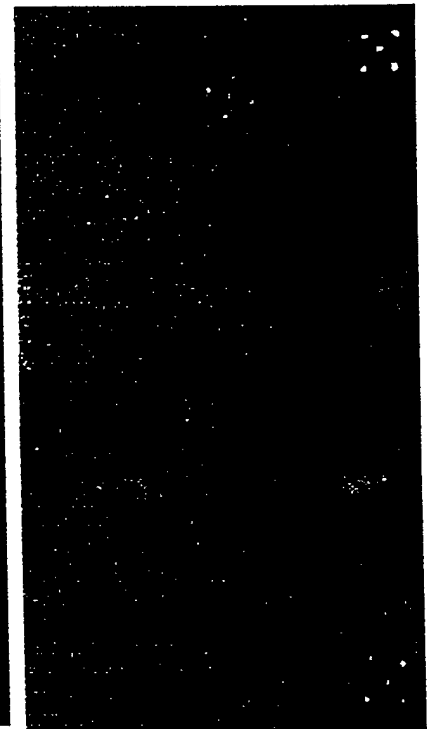
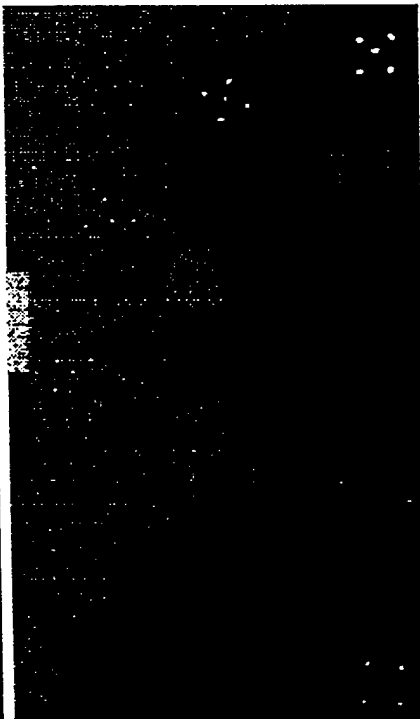
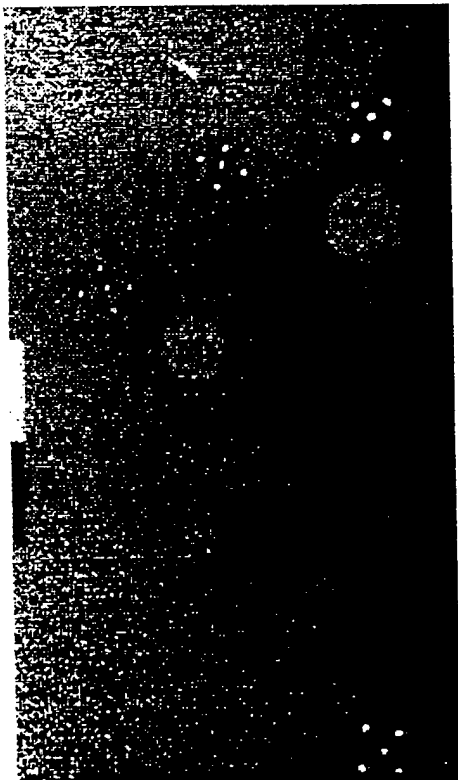
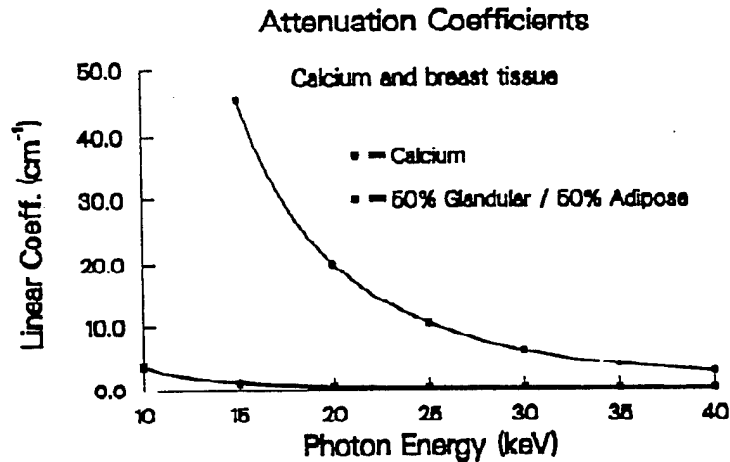
Conclusion of Tests of Changes in mAs

In both test objects, the digital system performed better at lower mAs, and similarly at higher mAs.

CLINICAL IMPLICATIONS FOR DIGITAL MAMMOGRAPHY

Clinical implications of KVP findings:

Some digital mammography machines now under development use a higher KVP so that the heat loading on the x-ray tube will be less. The results of this study suggest that at least with the image processing available to us, one may not be able to recapture the contrast information which is detectable at a lower KVP. This is related to the decreasing absorption coefficient of calcium with a higher KVP.



Storage phosphor 22 KVP Storage Phosphor 28 KV Storage Phosphor 36 KVP

These three views of the CIRS detail test object demonstrate that increasing the KVP lowers contrast with storage phosphor systems and that this contrast information cannot be completely retrieved with image processing.

Using a high contrast look up table to at least partially restore contrast information would also increase the contrast and thereby the visibility of noise. While our tests show that one can use image processing to make the digital image better than the screen film image at 36 KVP, it did not equal the performance of either digital or screen film images at 22 and 28 KVP in the CDMAM test object.

Clinical Implications of the mAs Experiments

The breast shows a wide variation in radiodensity. Denser areas of the breast, if their exposure level falls into the critical range, will have better conspicuity of small objects with the digital system than with screen film. If one uses screen film images, sufficient exposure to demonstrate the smallest objects should be used. While digital systems provide more information at low exposures, a decrease in exposure in the test system we used could be achieved only if one were to accept less than the full information the system could provide.

CONCLUSIONS

The use of storage phosphor digital mammography in its current configuration will not allow the exposures for digital mammography to be decreased if one wishes to maintain full information. In the presence of underexposure (such as in a focally dense area of the breast or in generalized breast increased radiodensity), there is a point below which a digital system does contain more information than a screen film system.

With increases in KVP, one has less object contrast. With the system tested, one cannot recover the contrast information lost at 36 KVP by image processing and still maintain an image of the breast that would be considered readable by a radiologist.

Bibliography

Freedman M, Pe E, Zuurbier R, Katial R, Jafroudi H, Nelson M, Lo S-C B, Mun S K: Image Processing in Digital Mammography. SPIE: Medical Imaging, Vol. 2164 (1994); 537-554pp.

Freedman M, Mun S K, Pe E, Lo S-C B, Nelson M: Image Optimization on the Fuji AC-1. SPIE: Medical Imaging (1993). Vol 1897: 480-502.

SUPPORTED IN PART BY

1. U.S. Army Medical Research Grant DAMD 17-93-J-3008. The content of this report does not necessarily reflect the position or policy of the U.S. Government.

Quality Control in Digital Mammography: Automatic Detection of Under and Over-Exposed Mammograms

Chris Y. Wu, Matthew T. Freedman, Akira Hasegawa, Seong K. Mun
Department of Radiology, Georgetown University Medical Center, Washington, DC

We developed a quality control system (QCS) for digital mammography that can notify technologists in real time of mammograms of poor image quality due to under or over exposure.

Mammograms are digitized by an Lumisys Scanner at 100 micron and 12 bits per pixel. An automatic image segmentation technique is employed to extract area inside the breast in mammogram. Histograms of the segmented areas are then calculated. By analyzing the composition of histograms, the computer program determines whether the original films have properly exposed.

Traditional image segmentation techniques are based on histogram analysis of digitized mammograms. However, such methods often fail with mammograms of low contrast or that are under-exposed because the difference in brightness across the breast skin line is so small that it is difficult to define boundary by thresholding or region growing techniques. We proposed a novel method to detect breast skin line based on statistical changes of gradient.

By analyzing the histogram composition of normal, under and over-exposed films, we defined an image feature that describes the image intensity content of underlying mammograms. The criterion for determining the category of a mammogram were established by studying a training database of normal, under, and over exposed films. We can then classify the mammograms using the image feature, based on the established criterion.

Over 150 real mammograms of different exposure levels were analyzed. The images were classified by the computer system into groups of normal, slightly under-exposed, under-exposed, slightly over-exposed, and over-exposed. We compared the classification results by computer with a radiologist's evaluation. Our QCS system was able to correctly classify over 85% of the cases. Receiver Operating Curve (ROC) analysis will be employed to evaluate the performance of the QCS system in determining the image quality of digital mammograms.

Our QCS program is able to automatically determine whether a mammogram is properly exposed and advise a technologist to re-take additional exposures. The QCS correctly identified 100% of over- and under-exposed mammograms and 92% of mammograms of normal exposure. The QCS can help reduce the cost of recalling patients and improve the overall quality of mammographic service.

Introduction

The objective of this paper is to develop a quality control system (QCS) for digital mammography that can identify mammograms of poor image quality due to under or over exposure in real time and notify the technologists so that the underlying errors can be corrected immediately to obtain high quality mammograms.

High quality of mammograms is essential for proper diagnosis in mammography screening examinations. In population screening mammography studies, approximately 6% of patients are called back for additional imaging examinations.¹ These additional examinations can be repeat views due to inadequate quality of previous mammograms, magnification views, coned compression spot views, and ultrasound examinations. Patients call backs due to poor image quality increase the cost of the service and agonizes patients. A five percent repeat rate for mammography films is considered the limit of acceptable standards for a mammographic service.² Identifying mammograms of inadequate image quality immediately after the examinations before the patients leave the hospital allows a technologist to re-examine a patient in case of a "bad" mammogram resulted for whatever reasons and obtain high quality mammograms for the patient in one hospital visit. This practice can reduce the number of repeat examinations and thus improve the quality of the mammographic service and reduce the cost of providing such service.

Reviewing mammographic images that had been rejected for poor quality demonstrates clearly that the major errors in image quality are reflected in the histograms of density of the images. The images were too light, too dark, lack of contrast due to insufficient compression, blurring of the edges of fibers of breast tissue due to patient motion during the examinations, and poor positioning. In general, mammographic images taken with properly maintained mammographic equipment can be of poor quality for five main reasons: poor positioning, patient motion, inadequate compression, lack of contrast due to too high a KVP, under or over exposure. Our research focuses on the detection of under and over exposed images. Lack of compression and too high KVP can also cause the films to be under exposed.

Technological issues of quality control of mammography examinations have been extensively discussed in the literature.^{2,3} However, there have been few systematic applications of computer assisted "intelligent" systems to solve problems in quality control. Sophisticated computer-aided diagnosis (CADx) algorithms have been successfully developed for diagnosis of breast cancer in recent years.⁴⁻¹⁰ We believe it is feasible to develop an automated quality control system that can identify more than 90% of the poor-quality mammograms that an experienced radiologist would request to be repeated.

The use of computer-assisted quality control and tele-mammography can assure that mammograms of diagnostic quality be obtained by technologists even at the absence of a highly qualified mammographer. It would decrease the number of women recalled for

repeat views, thereby decreasing the costs associated with loss of time from work and the additional cost of reprocessing these patients. It would also allow a more rapid report to be generated for those mammograms obtained at remote sites and to decrease patient waiting time.

Materials and Methods

Mammograms are digitized by a Lumisys scanner with a pixel size of 100 microns and a depth of 12 bits per pixel. An automatic image segmentation technique, gradient vector algorithm, is employed to extract area inside the breast in a mammogram. Histograms of the segmented areas are then calculated. By analyzing the composition of histograms, the computer program determines whether the original films have been properly exposed.

Normalization

A normalization standard is established so that mammograms digitized using different scanners can be analyzed and compared. Each digitized mammogram is re-scaled to 512×380 pixels and 8 bit gray scale depth. The spatial and gray scale resolutions at this setting is sufficient for our purpose of histogram analysis. The reduced resolutions greatly decreases image size and thus the computation time and storage disk space requirement. Therefore, the real time implementation of the algorithm in clinical environment can be facilitated.

Image Segmentation

The level of exposure of a digitized mammogram is to be determined by the analyzing the histogram of the mammogram. Because the area filled by the breast on a mammogram varies from image to image, histograms of mammograms with similar exposure level can have different shapes. In addition, the area outside the breast is of no diagnostic interest in determining the exposure level of a mammogram. What we are really interested is the exposure level inside the breast. For this purpose, we developed an image segmentation technique to extract breast area on a mammogram.

Traditional image segmentation techniques are based on histogram analysis of digitized mammograms. However, such methods often fail with mammograms of low contrast or with mammograms that are under-exposed because the small difference in brightness across breast skin line makes it difficult to define a boundary by thresholding or region growing techniques. However, image intensity ascends statistically from outside of breast to inside, even in low contrast mammograms. Based on this observation, we propose a novel method to detect breast skin line using a normalized differential filter. The normalized differential filter enhances statistical ascent of pixel values and makes the boundary easier to detect.

1. Normalized Differential Filters

Calculate both x- and y-direction differentials at each pixel in an input image using 5×5 kernels indicated in Fig. 1.

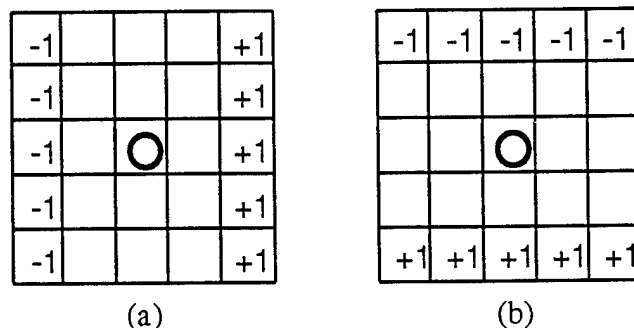


Figure 1 5×5 kernels for normalized differential filters

2. Gradient Vector

Consider each pair of pixel values in output images of the two differential filters as x and y components of a gradient vector, and normalize each vector. The x and y components of the normalized vector construct two output component images of the normalized differential filters. The component images demonstrate different characteristics between area just inside the breast skin line and area outside the skin line or in the center of the breast.

The directions and magnitudes of the gradient vectors just inside the skin line of the breast are observed to be more homogeneous than those of other areas such as the center of the breast or areas outside the skin line. Therefore, gradient vectors in the area just inside the skin line of the breast direct consistently towards the center of the breast. In addition, the gray levels of the gradient vectors' component images are relatively smooth. On the other hand, the underlying tissue of breast parenchyma is usually non-uniform in X-ray images, and the area outside the breast consists of mostly random noise. Therefore, normalized gradient vectors in these areas point randomly to various directions and the gray levels in the vectors' component images show greater variation. The different regions of a component image is demonstrated in Figure 2.

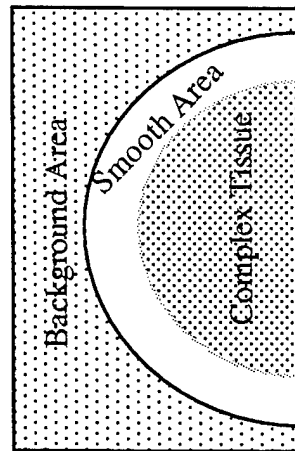


Figure 2 Different Regions of Gradient Vector's Component Image

3. Determination of Breast Area

The skin line of the breast is defined as the outer boundary of the smooth area indicated in Figure 2. To detect the smooth area, calculate variance of each pixel in a 5×5 kernel in a component image, and binarize the image at a certain threshold that has been pre-determined empirically. The pixels in binarized image have values of "1" inside the detected skin line of the breast and values of "0" outside the skin line.

The binarized images of component x and y are merged by logical "OR" operation that is followed by clearing inside of the breast. The final binary image of the breast is thus obtained.

4. CR Mammograms

In the proposed method, it is assumed that random noise exists in the background of mammogram. In screen-film mammography, this assumption always holds because low frequency quantum noises that are random in nature dominate background noises. However, in the FUJI computed radiography (CR) systems, the background area outside the breast is processed and thus is often not dominated by quantum noise in the final displayed image. The gradient vector method can not be applied to such mammograms to detect skin lines of the breast. However, such sophisticated method is often not needed because the contrast in a CR mammograms is so high that the breast area can be easily segmented by simple thresholding.

5. Scanner Sensitivity

Because gradient vector algorithm is designed to be able to detect subtle skin lines in low contrast mammograms, the method is also very sensitive to low frequency noise in background. When mammograms are digitized by a scanner with structured background noises such as observable low-contrast bold lines parallel to scanning direction, the

gradient vector method may fail to detect the true skin line because it can mistaken the low contrast line noise as the skin line.

Histogram Analysis

The pixel values in the normalized mammograms range from 0 to 255. A histogram normally shows two peaks. A major one represents the bulk of fibroglandular tissues in breast parenchyma and a minor represents the fatty tissues, as shown in Figure 3. To measure the level of exposure to the breast, we define Exposure Index (EI) to be the ratio of area under the major peak divided by the area of the valley between the two peaks. The boundaries of the two areas that are used to define EI are determined by analyzing a number of normal films that have been correctly exposed. When a film is correctly exposed, the Exposure Index is in the range of 0.5 ~ 1.5. When a film is under-exposed, the film appears light and area under the major peak increases, resulting in a higher EI. When a film is over-exposed, the film appears dark and the area under the major peak decreases, resulting in a lower EI. Average histograms of normal over- and under-exposed mammograms are shown in Figures 4 ~ 6, respectively.

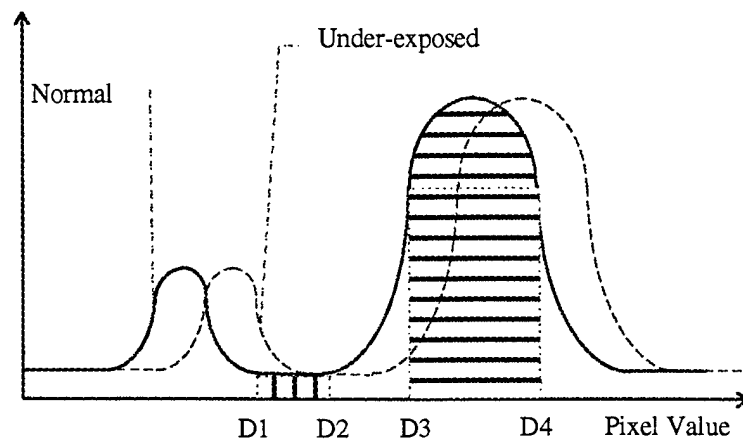


Figure 3 Histogram of a mammogram

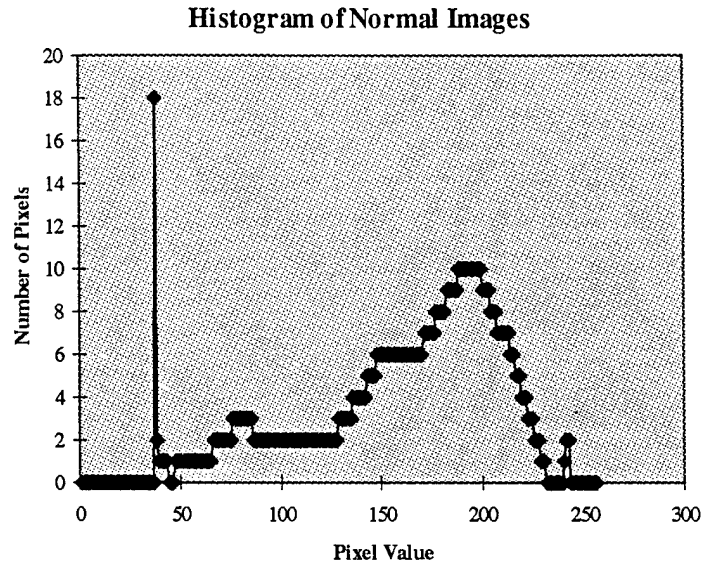


Figure 4 Average Histogram of Normal Mammograms

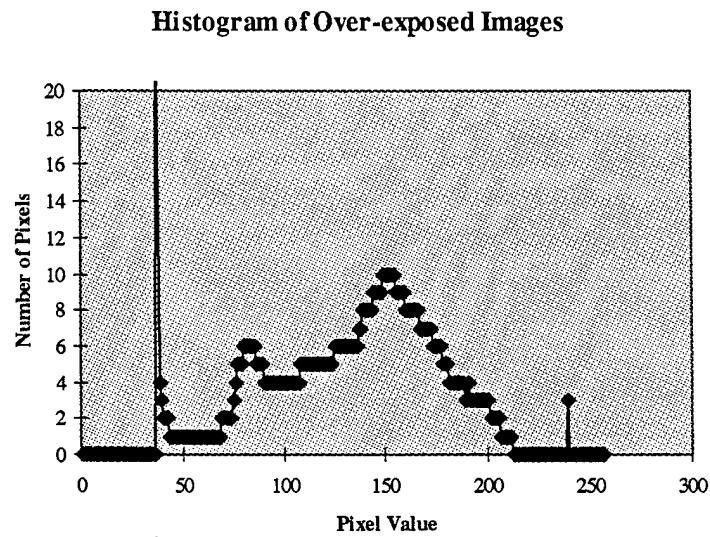


Figure 5 Average Histogram of Over-Exposed Mammograms

Histogram of Under-exposed Images

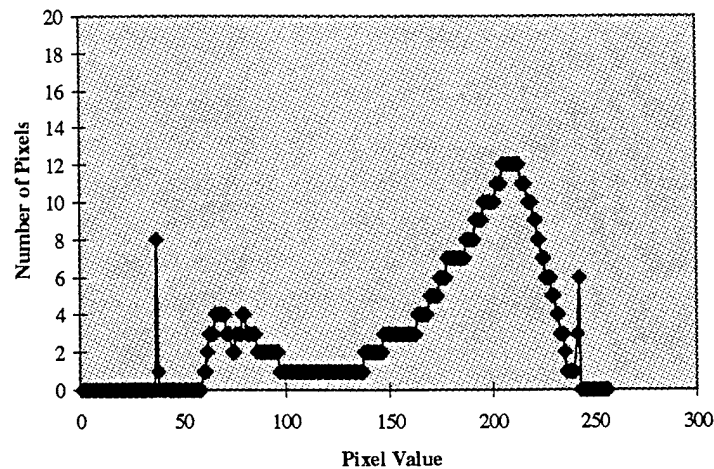


Figure 6 Average Histogram of Under-Exposed Mammograms

Measuring the Exposure Level for CR Images

The mammograms obtained from Fuji CR systems have been processed internally such that the pixel values fall in normal range even if the images are moderately under- or over-exposed. Therefore, the histogram approach used for screen-film system is no longer valid for the CR images. Alternatively, the exposure level can be measured by the quantum noise level in an image. Under-exposed images have higher level of quantum noise because a smaller number of x-ray photons were detected by the storage phosphor. Over-exposed images have lower level quantum noise due to a greater number of x-ray photons detected. We define a different EI measure for CR mammograms as rms noise of a selected region divided by the mean value of the region.

In Fuji CR systems, the exposure level is related to an S-number. The greater the S number, the lower the exposure level in a CR image. To determine whether the S-number is an effective measure of exposure, we imaged a breast phantom with different exposures and obtained S numbers for each exposure. We then calculated RMS noise levels for images of different exposures, as shown in Figure 7. The S-numbers increase, in general, with RMS noise level. However, the S-number is not a reliable measure of exposure level. Within a limited range of exposure, the S-numbers can vary in a much wider range. This observation is demonstrated more obviously by using real CR mammograms. We selected over 30 CR mammograms that are considered by an experienced radiologist to be within normal exposure range. We calculated the RMS noise for each image and compared it with corresponding S-number, as shown in Figure 8. The RMS levels of these mammograms fall within a narrow range because the images were all appropriately exposed. The S-numbers, however, varies greatly from image to image.

RMS of CR Phantom Images and their S-Numbers

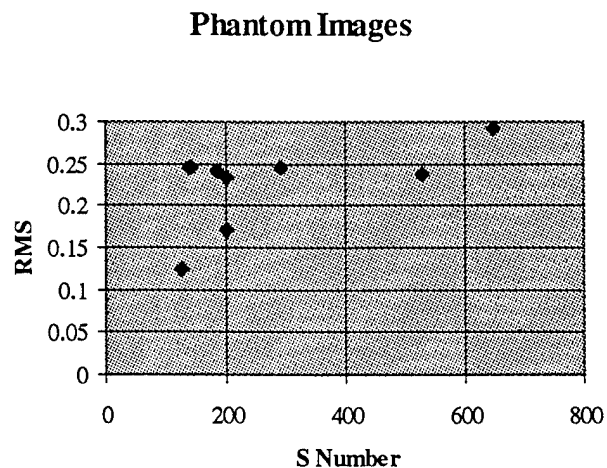


Figure 7 Correlation between S Numbers and RMS Noise for CR Phantom Images

RMS of CR Mammograms and their S-Numbers

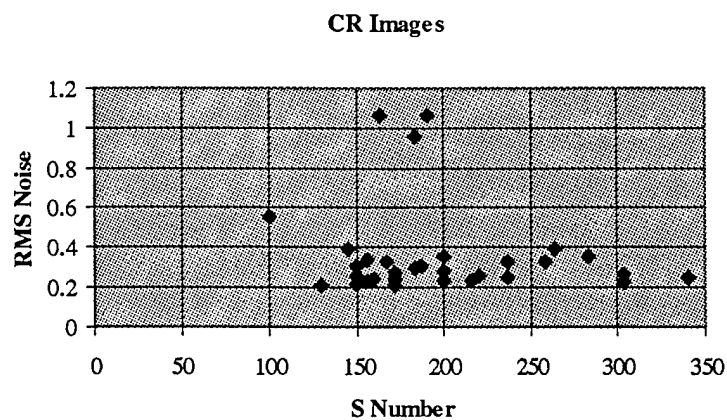


Figure 8 Correlation between S Numbers and RMS Noise for CR Mammograms

Results

1. Screen-Film mammograms

We selected 181 cases from the past mammography examinations. The gradient vector algorithm failed to detect the boundary of the breast in 10 of the 181 originally selected cases. The selected cases were examined by a radiologist to determine the exposure level

of the images. These images were classified by the radiologist into four categories: under-exposed (30), normal (83), and over-exposed (58).

A. Exposure index by the QCS for Classification of mammograms

Table 1 Summary of Output of QCS

	Under-Exposed	Normal	Over-Exposed
Average	4.607	1.072	0.193
Minimum	2.993	0.000	0.099
Maximum	7.720	1.328	0.808
Median	4.747	1.139	0.122
Standard Deviation	1.181	0.269	0.156

Table 1 lists the key statistics of the three groups of images classified by the QCS. The three groups are well separated as shown in Fig. 9. The diagnostic performance of the QCS is summarized in Table 2.

Output of QCS

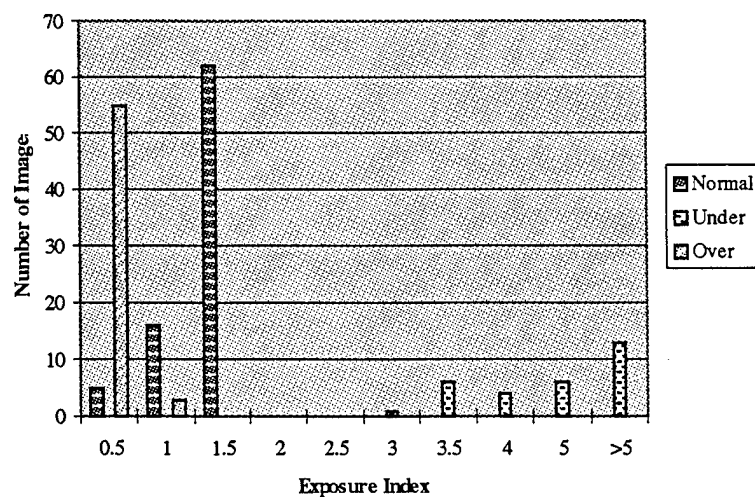


Figure 9 Histogram of Output of QCS

Table 2 Performance of QCS

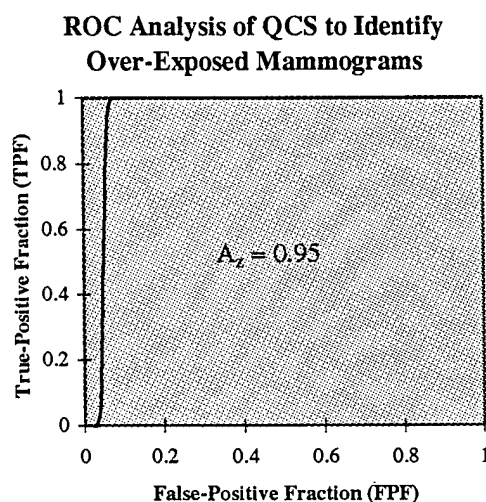
QCS	Normal	Under-Exposed	Over-Exposed
Normal (83)	76	0	7
Under-Exposed (30)	0	30	0
Over-Exposed (58)	0	0	58

The QCS correctly identified 76 (92%) of the 83 normal images and all of the under- and over-exposed images. Notice that none of the under- and over-exposed images are missed by the QCS, assuring correct exposure for all mammograms.

B. ROC Analysis Of the Performance of QCS

To evaluate the ability of the QCS to differentiate between normal and under-exposed or between norm and over-exposed images, we applied ROC analysis to the output of QCS.

(1) Between normal and over-exposed images



(2) Between Normal and Under-Exposed images

The QCS system distinguishes the two types of images perfectly with an A_z value of 1.0.

Discussion

Exposure Index characterizes the exposure level of Screen-Film mammograms accurately. Because the threshold values are empirically determined, the accuracy of the

characterization is subject to variations with different types of images digitized with different film scanners.

The CR mammograms are pre-processed such that the pixel values are reset to a normal range. Histogram based approaches of quality control are no longer applicable to CR images. An alternative method based on the RMS noise level in an image is developed to assess the exposure level in CR images. The RMS noise based method provides a better assessment of exposure than the S-numbers recorded for each image.

Conclusion

QCS is an effective system that can automatically evaluate the exposure level of a Screen-Film mammogram in real time. The QCS identified 100% of the under- and over exposed mammograms while correctly identified 92% of mammograms exposed with normal exposure. Technicians can be notified by such system of any mammograms of inferior image quality due to under- or over-exposure to the films. The use of QCS can improve the overall quality of mammographic services by assuring high image quality of mammograms and reducing or eliminating recalling patients for repeat mammograms. The cost saving in reduction of patient recall can be substantial. The patients will be better served as well.

As digital mammography is gaining wider acceptance across the country and in the rest of the world, the potential of QCS applications looks very promising. In a digital environment, QCS can be integrated into the image acquisition to monitor image quality interactively and in real time.

References

1. Cardenosa G, Eklund G. Rate of compliance with recommendations for additional mammographic views and biopsies. *Radiology* **1991**; 359-361.
2. American College of Radiology. Mammography Quality Control. **1990**.
3. Williams and Wilkins. Workbook for quality mammography. Baltimore; **1992**.
4. Chan H-P, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM. Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography. *Med Phys* **1987**; 14: 538.
5. Chan H-P, Doi K, Vyborny CJ, Lam KL, Schmidt RA. Computer-aided detection of microcalcifications in mammograms: Methodology and preliminary clinical study. *Invest Radiol* **1988**; 23: 664.
6. Chan H-P, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam K-L, Ogura T, Wu Y, MacMahon H. Improvement in radiologists' detection of clustered microcalcifications on mammograms: The potential of computer-aided diagnosis. *Invest Radiol* **1990**; 25: 1102-1110.
7. Giger ML. Computer-Aided Diagnosis. In: Haus AG, Yaffe MJ, eds. Syllabus: A Categorical Course in Physics: Technical Aspects of Breast Imaging. Oak Brook, IL: RSNA; **1993**: 283-298.
8. Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE. Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer. *Radiology* **1993**; 187: 81-87.
9. Wu Y, Doi K, Giger ML, Nishikawa RM. Computerized detection of clustered microcalcifications in digital mammograms: Applications of artificial neural networks. *Med Phys* **1992**; 19: 555-560.
10. Wu CY, Patt RH, Freedman MT, Mun SK. Diagnosis of breast cancer by MRI: A 3D computer visualization and analysis system. SPIE Medical Imaging 1996 —Image Display 1996, 2707: 87-96. Newport Beach, California.

DRAFT

**QUALITY CONTROL
FOR
FUJI AC-3
COMPUTED RADIOGRAPHY**

**Hamid Jafroudi, Ph.D.
Dorothy Steller, RT, (R), (M)
Matthew Freedman, M.D., M.B.A.
Seong Ki Mun, Ph.D.
Gary Norton, RT**

**Georgetown University Medical Center
Department of Radiology
Imaging Science and Information Systems
2115 Wisconsin Avenue, NW, Suite 603
Washington, D.C. 20007**

FOR GOVERNMENT DISTRIBUTION ONLY

Fuji AC-3 Computed Radiography (FCR AC-3)

TABLE OF CONTENTS

1.	Quality Control Testing On Computed Radiography	3
1.1.	Introduction	3
1.2.	Safety Precautions	5
2.	Quality Control Testing On The Image Reader FCR AC-3, CR-IR318	5
2.1.	Preventive Maintenance(PM) On The Image Reader FCR AC-3, CR-IR318	15
2.2.	Corrective Maintenance(CM) On The Image Reader FCR AC-3, CR-IR318	15
3.	Quality Control Testing On The Image Recorder, CR-LP414	15
3.1.	Preventive Maintenance(PM) On The Image Recorder, CR-LP414	21
4.	Fuji Imaging Plates	21
5.	Routine Quality Control Procedures	23
6.	References	24
APPENDIX I Troubleshooting On The Image Reader FCR AC-3, CR-IR318		24
APPENDIX II Troubleshooting On The Image Recorder, CR-LP414		25
APPENDIX III Standard Radiographs For Future Comparison		25

1. Quality Control Testing On Computed Radiography

Computed radiography (CR) based on storage phosphor (SP) detector is a new and complex technology for obtaining medical digital imaging. The CR system was developed such that its image quality in diagnosis (wide latitude, certainty of visual diagnosis), speed (image sensitivity), image processing, and the overall imaging performance exceeds or at least equals those of the conventional screen-film (SF) radiography systems which is considered to be "gold standard".

The unique capability of the CR system places a new responsibility on the medical physicist, radiologist, and radiologic technologist to ensure that the digital images give the same information at least as film, and improve rather than degrade the image. For that reason the quality of the image must be ensured and each component of the system must function properly.

The clinical knowledge about quality control (QC) and the standard procedures for CR devices has not yet been established and is still under development. This manual presents the QC testing of the CR image reader/recorder and film processor. The procedure includes the acceptance of the equipment for its image quality, image sensitivity, and overall imaging performance. The procedures are currently and routinely being utilized in our institution.

1.1. Introduction

A number of digital technologies are challenging the established screen-film (SF) technology such that the separation of image acquisition and image display are no longer applicable. One of these technologies is based on storage phosphor (SP) imaging medium used in computed radiography (CR).¹ The traditional SF radiography systems' components such as image acquisition, image display, and image archiving are combined in the film medium. But in the CR system they are in separate optimizable components. The CR system uses the standard radiographic machine, grids for scatter rejection, and cassette similar to the screen-film radiography system. Positioning of the patient is the same as SF system. The exposure technique (s) (kVp and mAs) are also similar to SF system.

SP digital radiography systems were introduced in the mid 80s to replace SF imaging systems.¹⁻³ The imaging plate for the SP system is a new film-like image receptor⁴⁻⁵; specifically designed phosphors trap and store the radiation energy. The stored energy is stable until scanned with a laser beam, which releases the energy as luminescence. The basic concept of the CR system was to improve the image quality to at least equal to that film.⁶⁻⁷ The system provides good diagnostic information that produces consistent sharp images with a wide

latitude. Because of the image digital format, it can be further modified through image processing, then stored for retrieval and communication in a picture archiving and communication system (PACS).

The CR systems aim to replace the present SF system of analog X-rays. In 1981 Fuji Medical Systems introduced the first CR system based on SP technology, Fuji Computed Radiography (FCR) system (FCR101). Fuji has been the leading manufacturer of CR machines since that time. Fuji has also licensed its system to several other companies including Toshiba (TCR 1984), Philips Medical Systems (PCR 1985), and Siemens AG (1988). A review of the literature reveals that many companies have mounted research efforts using a SP detector for medical imaging: Eastman Kodak Co.; Fuji Photo Film Co., Ltd.; N.A. Philips Corp.; Konica Corp.; E.I. duPont de Nemours & Co.; 3M Co.; Agfa-Gevaert N.V.; Hitachi Ltd.; Siemens AG; Toshiba Corp.; General Electric Corp.; Kasei Optonix, Ltd.; Mitsubishi Chemical Industries, Ltd.; Nichia Corp.; GTE Products Co.; and DigiRad Corp.² Therefore a standard procedure is needed to ensure the system's performance and image quality.

The complete CR system from imaging plate and cassette to photomultiplier and laser for the development of the image, to computer algorithms used for image processing and hard copy output devices must respond properly to the final image. Therefore each section of the CR system should be carefully tested and calibrated. All subsystem components of the CR device such as image acquisition, image processing, and image display are subject to the variations in performance that may cause image degradation. A few protocols for CR machines' QC testing have been identified by several researchers.⁸⁻¹⁰ Currently, The American Association of Physicists in Medicine (AAPM) Task Group #10 is working to define acceptance tests and QC procedures for CR systems. The goal is to set general standards and to develop a protocol for the CR system. The protocols are based on several Fuji CR systems and may not apply to other manufacturers systems such as Agfa's CR machines.

This manual addresses Quality Control on the CR machine from five points:

- Acceptance testing
- Establishing standards of performance
- Continuing checks on performance of a machine that has passed acceptance testing and for which standards have been established.
- Detailed tests for machines that do not pass the routine checks.
- Routine QC testing

The concept that underlies this approach is as follows:

First, make sure that the system is functioning properly. **Second**, obtain standard images on the correctly functioning system to be used as baseline standards for future tests. **Third**, obtain test images at predetermined intervals and compare them to the baseline images. **Fourth**, if the images have changed, do more detailed tests to find the source of the problem.

This manual is divided into several chapters. Chapter 1 of this manual gives an introduction about the Computed Radiography (CR) system as well as the safety precautions necessary for CR devices. Chapter 2 is an important chapter of this manual. It explains step by step procedures on Quality Control testing on the CR image reader. In Chapter 3, the acceptance of the image recorder is discussed. Chapter 4 describes any issues related to imaging plates and cassettes. Chapter 5 describe the routine quality control testing on CR machine. The references are listed in Chapter 6. Appendices I and II are for troubleshooting for the image reader and image recorder, respectively. Appendix III must have standard radiographs that can be used for future comparison.

1.2. Safety Precautions

The CR machine contains a built-in laser source. ALL users should read and understand the safety precautions listed in the Fuji SERVICE MANUAL.¹¹⁻¹²

To prevent any accidents or hazards that may occur during servicing, read the sections of the SERVICE MANUAL that apply. Electrical shock, heat injury, and injury from rotating parts can occur inside the CR machines. The SERVICE MANUAL covers the safety concerns for each task that must be performed on this machine.

2. Quality Control Testing On The Image Reader FCR AC-3, CR-IR318

After the installation check is done, the machine is ready for Quality Control (QC) testing, for **Nonuniformity, Sensitivity, and Density**.

The equipment needed to perform the complete QC test is listed below:

- Imaging plates and cassettes of the following sizes
 - 14 in x 17 in (35 cm x 43 cm)
 - 14 in x 14 in (35 cm x 35 cm)
 - 10 in x 12 in (24 cm x 30 cm)
 - 8 in x 10 in (18 cm x 24 cm)

- Densitometer
- Dosimeter
- 5X, 10X or 20X magnifier glass
- Resolution phantoms (at least 5 lp/mm)
- Anthropomorphic phantoms (chest, hip, shoulder, hand, foot, and breast phantoms)
- Anatomical menu and image processing parameter settings.

The exposure techniques for different anatomy vary for different institutions. The exposure technique, set for the image quality is as follows:

- Distance: 1.8 m
- Voltage: 80 kVp
- Amperage: 50 mA
- Time: 0.013 sec
- Dose: 1 mR

CHECKING SYSTEM PERFORMANCE

After the machine is installed, the system performance must be checked.

To Control Power ON / OFF from machine, check the following processes:

- Verify that the **REMOTE / LOCAL** switch on the power supply unit of the machine is in the **REMOTE** position.
- Check the external units (**LP, IDT**) connected to the machine for the following checkpoints.
 - Verify that the **REMOTE / LOCAL** switch of the **LP** is in the **REMOTE** position, and its breaker in the **ON** position.
 - Verify that the **REMOTE** switch of the **IDT** is in the **ON** position and its power switch in the **ON** position.
- Turn **ON** the power of the machine.
Verify that the machine starts up normally.
- Verify that the external units (**IM, LP, IDT**) starts up normally to be ready for communication.
- Check the **ID** terminal connected to the machine to ensure that:
 - The **REMOTE** switch of the **ID** terminal is in the **OFF** position, while its power switch is in the **ON** position.
- Power **ON** the machine.
The machine does not start up yet.
- Turn **ON** the **REMOTE** switch of the **ID** terminal.
- Verify that the machine starts up normally.
An example of the normal operation is the panel screen appearing when the machine has started up as illustrated below.

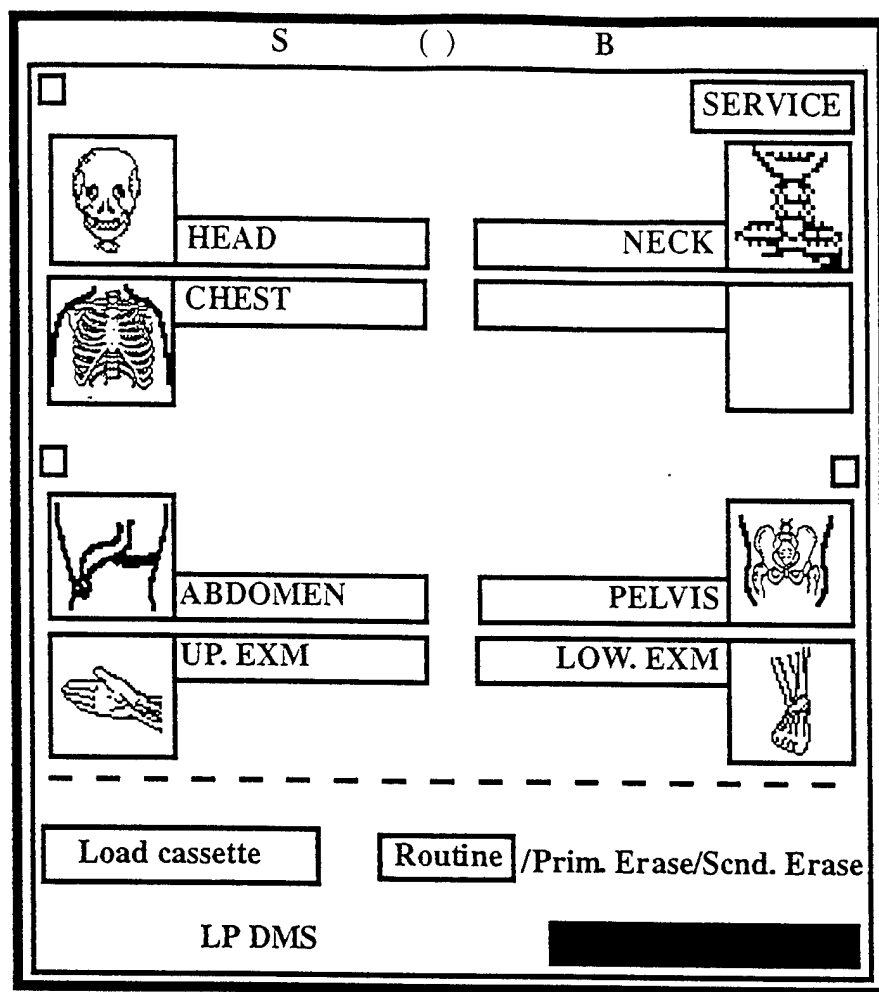


Figure 2-1 Example of Screen Panel after Normal Machine Start Up

Note :

LP Laser Printer
 IM Image Monitor
 IDT Identification Terminal
 "IDT" and "ID terminal" should be understood as "console".

IMAGE QUALITY TESTING

In this section several tests must be performed for acceptance of the machine such as check for **Nonuniformity, Sensitivity, Density, Jitters, Formats**, and check for **Output Characters**.

Before starting the acceptance test, all IPs used in the experiment must be carefully tested for dust, scratches, and cracking around the edges of the plates that cause artifacts.¹⁰ Plates should be cleaned regularly and should at least undergo primary erasure before the start of the test. This also applies to the cassette, the hinges need to be safe when it is running through the image reader. The lead backing used for the larger plates must be checked for nonuniformity that may result in image degradation and artifacts.

In addition, the exam room must be tested for the accuracy of the x-ray exposure and x-ray tube voltage and current. The dosimeter and the kVp divider were used for the measurements of the exposure dose and the voltage. The initial experiments were carried out on two different types of IPs (standard resolution, ST-V and high resolution, HR-V) with different sizes. Fuji GF-1/HR-G (50 speed) film was used for the extremities exams and Fuji GH-1/HR-G (400 speed) film was used for the chest, skull, hip, and shoulder exams. The technique charts for different exams should be available for each institution that the exams are performed.

The procedure for QC of the CR machines are as follow:

a- Checking for Irregularity, Sensitivity, and Density

Turn on the power of both image reader and the laser printer to reset the unit. After all IPs are erased, then use the following tests using SENSITIVITY in the TEST menu. The exposure techniques for these tests are listed in Table 2-1

Table 2-1 Exposure Techniques

No.	Distance	Voltage	Current	Time	mAs	mR
1	1.8 m	80 kVp	50 mA	0.001 s	0.05	0.1
2	1.8 m	80 kVp	50 mA	0.013 s	0.65	1.0
3	1.8 m	80 kVp	50 mA	0.130 s	6.50	10.0

- **Check for Irregularities**

- Uniformly expose flat pattern exposure in each IPs of the following sizes at about **1 mR**.(No.2)

Note that size values in parentheses should be used when the machine is set in metric unit of measure.

14" x 17" (35 cm x 43 cm)

14" x 14" (35 cm x 35 cm)

- 10" x 12" (24 cm x 30 cm)

- 8" x 10" (18 cm x 24 cm)

- Make recording in **SENSITIVITY** from the **TEST** menu of the display panel, and generate output.
- Verify that the image on the output film is free from nonuniformity. The images can be one-to-one or two-to-one. If any nonuniformity is found, take appropriate corrective action. (Please See Chapter F2 of SERVICE MANUAL)

- **Check for Sensitivity and Density**

Make sure that the value of the system sensitivity indicated on the film outputted in "Check for Irregularity" is about 200 and the

film density is about 1.2 ± 0.10 .

Verify that there is no density variation in the main scanning direction.

If anything abnormal is found, take appropriate action.

(Please SEE Chapter F2 of SERVICE MANUAL)

b- Checking for Jitters, Formats, and Output Characters

- **Check for Jitters**

- Place two 15-cm steel scales as illustrated below on **IPs** of the following sizes, and uniformly expose them at about **1 mR**.
Note that size values in parentheses should be used when the machine is set in metric unit of measure.

- 14" x 17" (35 cm x 43 cm)
- 14" x 14" (35 cm x 35 cm)
- 10" x 12" (24 cm x 30 cm)
- 8" x 10" (18 cm x 24 cm)

The IP exposure conditions are the same as specified before.

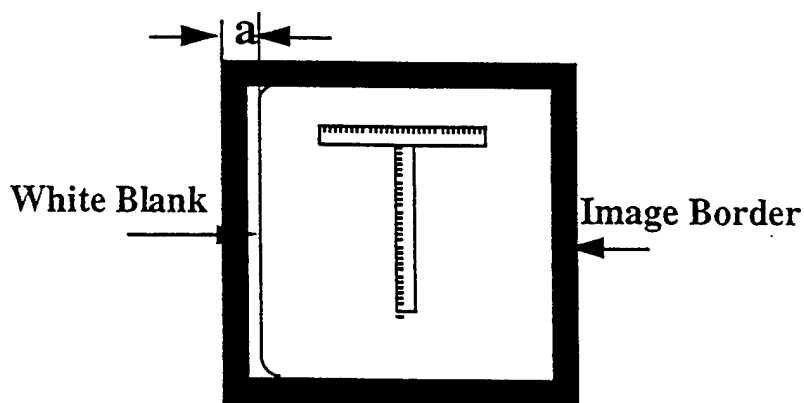


Figure 2-2 Two 15 cm steel scales for one-in-one image output.

- Make recording in **SENSITIVITY** from the **TEST** menu of the display panel, and generate output.
- Make sure that the border and steel-scale image on the output film are free from jitters.
If any jitters is found, take appropriate corrective action.
(Please SEE Chapter F2 of SERVICE MANUAL)

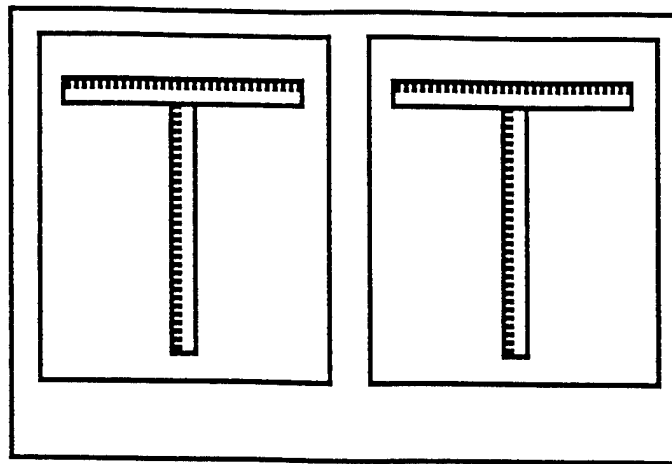


FIGURE 2-3 Output Film Image (Ex. : Two-in-One Image Format)

- **Check for Formats**
Measure the white blank from the outermost corner of the image border of the film generated in "Check for Jitters" and verify that it is less than 2 mm with trimming set to 0. Also check that the image size of the steel scales corresponds to the reduction ratio listed in Table 2-2. This is refer to Figure 2-2.

Table 2-2 Reduction ratio corresponds to image format on the film

Reading Size	Two-in-one image	Full image
14 in x 17 in	1/2	2/3
14 in x 14 in	1/2	2/3
10 in x 12 in	2/3	1/1
8 in x 10 in	6/7	1/1

(a) Reference 11

If the white blank is more than 2 mm in width or if the reduction ratio is improper, take appropriate corrective action.

(Please SEE Chapter F2 of SERVICE MANUAL)

- **Checks for Output Characters**
Make sure that the contents of film character format information set in

Chapter C, Section 8 "Configuration File Setting" are indicated on film outputted in "Check for Jitters".

If anything abnormal is found with the output characters, take appropriate corrective action.

(Please SEE Chapter F2 of SERVICE MANUAL)

- **Checks for Output Characters**

Make sure that the contents of film character format information set in Chapter C, Section 8 "Configuration File Setting" are indicated on film outputted in "Check for Jitters".

If anything abnormal is found with the output characters, take appropriate corrective action.

(Please SEE Chapter F2 of SERVICE MANUAL)

Output Characters

Using the image output taken from previous studies, check the contents of the film character format information which is set in the initial installation set up. This is shown in Figure D-04. The image output should contain the following information:

- O Upper Portion of Image**
 - a: Name of hospital (name of institution)
 - b: IP number
 - c: EDR mode and menu code
 - d: System ID and image number

- O Lower Portion of Image**
 - e: Image processing conditions
 - f: Exposure menu name
 - g: Standardization conditions and correction item
 - h: Engineer code and exposure table information
 - i: Department name
 - j: Patient ID
 - k: Patient name (John)
 - l: Patient name (Smith)
 - m: Date of exposure and time
 - n: Film mark
 - o: Sex
 - p: Age or birth date
 - q: Image reduction ratio
 - r: Set processing information

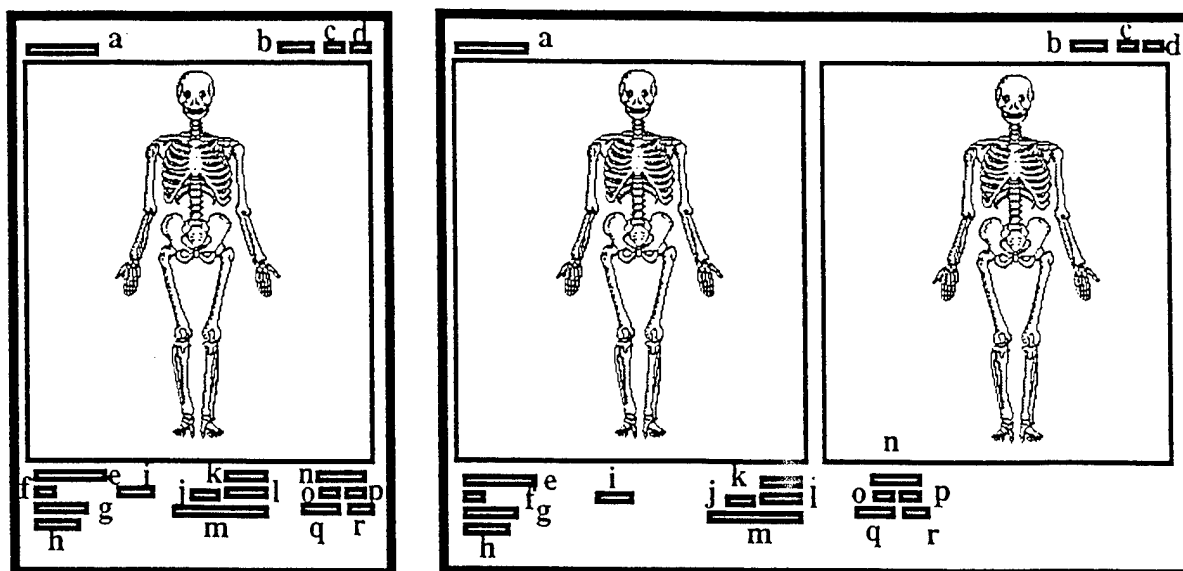


Figure 2-4 Schematic of film characters on one-in-one image or two-in-one-image output

If the image output does not satisfy the output film characters, consult section D2 of the SERVICE MANUAL.¹¹

CHECKING DMS INTERFACE FUNCTIONS

If the Data Management System (DMS) is connected to the machine, verify that images are stored normally in the DMS and that are retrieved normally from the DMS.

DISPLAYING AND CLEARING ERROR LOGS

a- Displaying Error Log

Display an error log in the Maintenance Utility (M-Utility) mode, and verify that there has not occurred any serious error.

- Enter the M-Utility mode.
 - Press the L7 key, then hit the R8 key within 1 sec.)

L7 **R8**

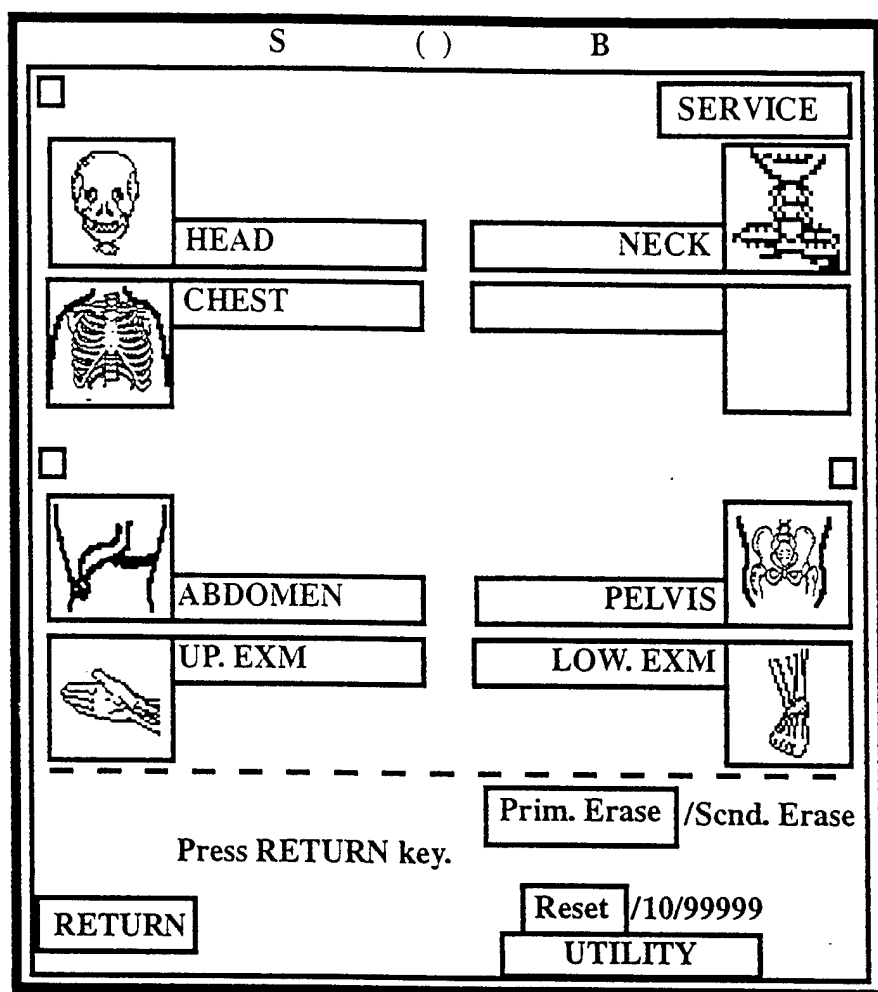


FIGURE 2-5 Shifting to M-UTILITY

- Press the RESET button located at right on the rear of the ID terminal.
- After elapse of about a few seconds, when a message,

GET SHARED MEMORY,

appears, press the [ENT] key.

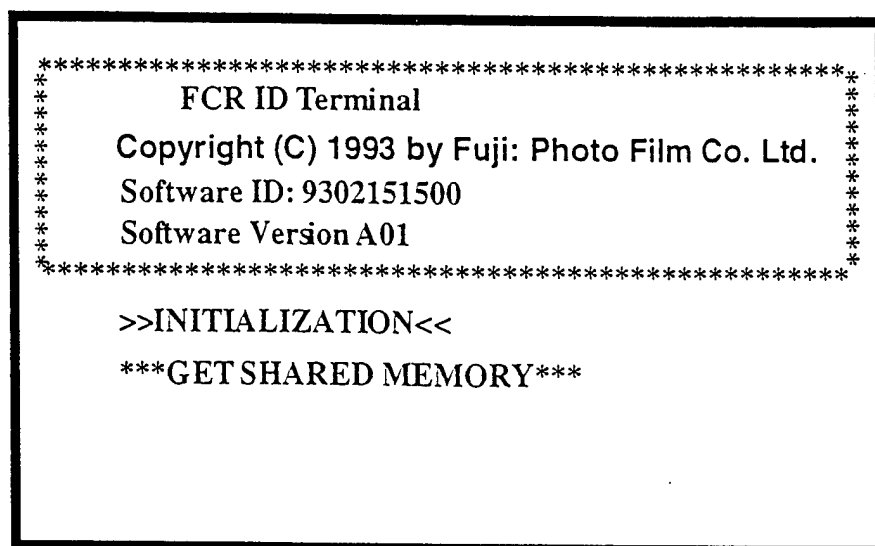


Figure 2-6 GET SHARED MEMORY Screen

Press the **ENT** key before a message, *****LOAD TABLE*****, appears on the screen. If you miss the action, turn OFF the REMOTE switch of the IS terminal and back ON to perform the above described procedure again, starting from the first step. Then, the Service Utility Screen appears as the following:

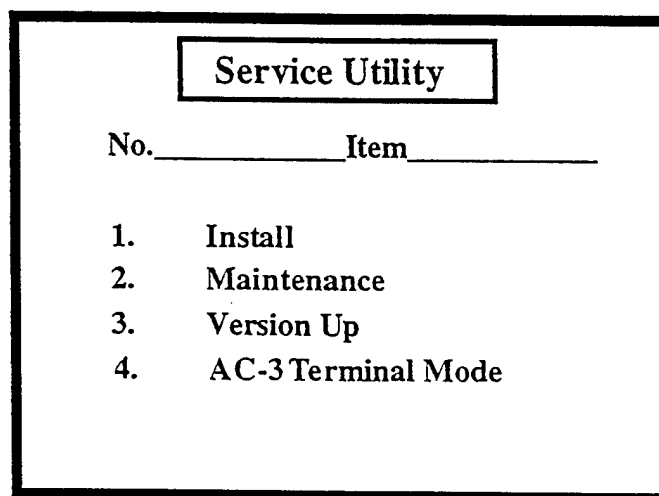


Figure 2-7 Service Utility Screen

- Select AC-3 Terminal Mode
4 **ENT**
- Select ERROR LOG UTILITY, LIST, and ALL in sequence.
1 **ENT** **1** **ENT** **1** **ENT**

Then, an error log list appears

Service Utility			
LISTALL			
NO.	CODE	DATE	ERROR DETAIL
001	FFF	YY. MM. DD. HH	XXXX
002	FFF	YY. MM. DD. HH	YYYYYYY
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
010	FFF	YY. MM. DD. HH	YYYYYYY
EL:LIST:ALL>_			

Figure 2-8 Display Example of Error Log List

To scroll to the next screen, previous screen, or next line, manipulate as follows :

- Next Screen : **ENT**
- Previous Screen : **R7 ENT**
- Next line : **SP ENT**
- Verify that no serious error has occurred.
- Quit the error log display.

. ENT 0 ENT

b- Clearing Error Log

- Select **CLEAR** and **YES** in sequence.

2 ENT 1 ENT

Then, the error log is cleared.

- Change to the Routine Processing mode.

0 ENT 0 ENT L1 ENT

CHECKING SYSTEM REMOTE OFF FUNCTION

Verify that the machine and external unit (s) (IM, LP) connected to power OFF normally, when the power switch of the IDT is turned OFF.

CLEANING COVERS

Clean all the covers of the machine with a **MOISTENED CLOTH**. When the machine is installed for the first time, remove the protective sheets from the operation panels of image reader/recorder and film processor, respectively.

2.1. Preventive Maintenance(PM) On The Image Reader FCR AC-3, CR-IR318

Procedures for maintenance operations must be performed periodically to maintain the CR machine for good image quality and good machine operation. This is discussed extensively in **Chapter E of the SERVICE MANUAL**.¹¹

2.2. Corrective Maintenance(PM) On The Image Reader FCR AC-3, CR-IR318

Procedures for corrective maintenance operations must be performed when an error happens during the installation and operation of the CR machine. This is discussed, in detail in **Chapter F of the SERVICE MANUAL**.¹¹

3. Quality Control Testing On The Image Recorder, CR-LP414

Inspection procedures must be performed to verify that the machine meets all the requirements in compliance with the specifications. When the machine is to be operated as a component unit of the system, check the interface as indicated in the system instructional manual and service manual for the equipment to be connected. If any abnormality is encountered in the inspection, take necessary actions as referred to Appendix II or directed in **SERVICE MANUAL**.¹² for Image Recorder in Chapter F.

PERFORMANCE CHECKS

a. Recording Screen Format Check

Enter the service mode, perform the FLAT test pattern recording operation, and take the

following measurements to check whether the requirements are met.

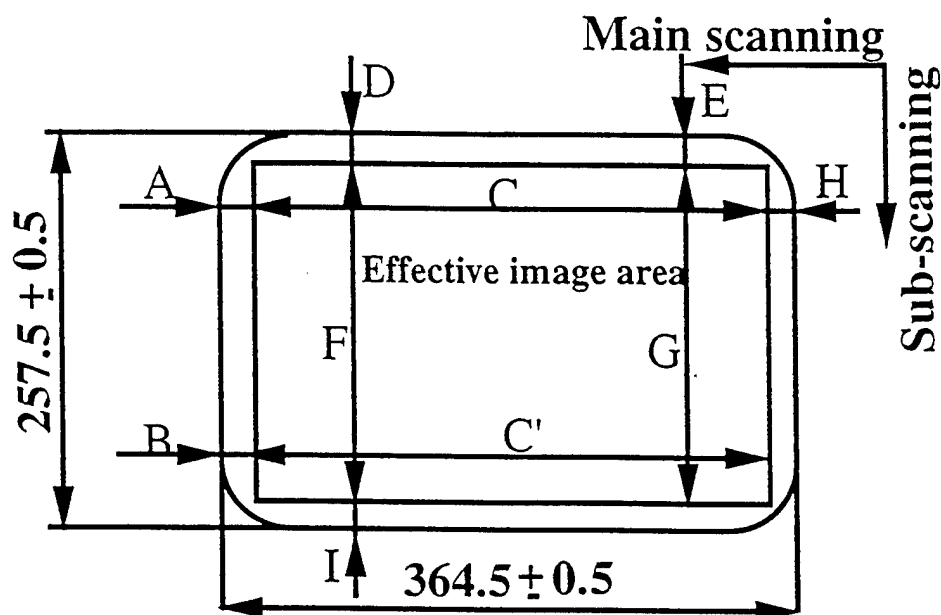


Figure 3-1 Flat Test Pattern

Table 3-1

Measurement	Requirements [mm]
A	2.2 ± 0.6
C	360 ± 1.2
C'	360 ± 1.2
D	1.6 ± 1.0
E	1.6 ± 1.0
D-E	≤ 1.2
F,G	254 ± 0.6
F-G	≤ 0.5
(H)	(2.3 typ.)
(I)	(1.9 typ.)

b. Density Check

This check must be performed while the film processor temperatures are thoroughly stabilized after machine startup. This is referred to the instructional manual for the FL-IM and execute the density control function to carry out density corrections before this check.

- (1) Press the **CHECK DENSITY** key to check the recorded density.
- (2) After completion of check, press the **YES** key to return to the normal mode.

Also check that the following requirements are met.

With a densitometer, measure the 17-step wedge output generated in step (1) above. Check whether the following requirements are met.

Table 3-2

Step	Requirements	
1	0.10 typ	(dependent on the fog density)
2	0.13 typ	(dependent on the fog density)
3	0.20 typ	(dependent on the fog density)
4	0.30	± 0.04
5	0.48	± 0.04
6	0.66	± 0.05
7	0.84	± 0.05
8	1.02	± 0.05
9	1.20	± 0.05
10	1.38	± 0.05
11	1.56	± 0.05
12	1.74	± 0.05
13	1.92	± 0.05
14	2.10	± 0.07
15	2.28	± 0.07
16	2.46	± 0.07
17	2.64	± 0.07

Repeat steps (1) and (2) in the preceeding section three times and check that the difference between the maximum density values of each step is as specified below.

Table 3-3

Step	1st sheet	2nd sheet	3rd sheet	Maximum density difference	Requirements
1					0.03
2					0.03
3					0.03
4					0.03
5					0.03
6					0.03
7					0.05
8					0.05
9					0.05
10					0.05
11					0.05
12					0.05
13					0.07
14					0.07
15					0.07
16					0.07
17					0.07

c. Image Quality Check

After the completion of image recording film density check, perform the following image quality procedures:

c.1 Checking for Unevenness and Scratches

Enter the service mode and execute the FLAT test pattern record function to check for unevenness and scratches. Generate three test pattern outputs for checking purposes. Number the generated film outputs.

- Random unevenness in scanning line direction
- Wave-like stripe unevenness
- Vertical stripes
- Cross stripes of pixels
- Vacuum cup marks
- Streaks 10 or 13 mm apart from the film upper edge
- Static marks
- Film trailing end guide rib length not exceeding 5 mm (unevenness attributable to the film processor)
- Scratches
- Other irregularities

c.2 Shading Check

Check that there is no shading in the main scanning direction of the test pattern film outputs generated in the preceding check. Measure the density at five points from the side and check that the maximum optical density difference among the five points is not greater than 0.1. This is shown in Figure 3-2.

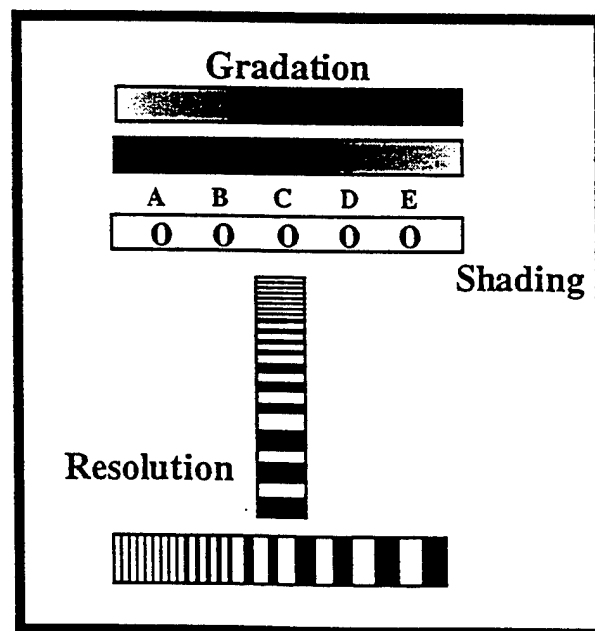


Figure 3-2 Film Location for the density measurement

c.3 Sharpness Check

Enter the service mode and execute the SHARPNESS1/SHARPNESS2 test pattern record.

Check that a 5cycle/mm pattern is clearly visible in both the main and sun-scanning directions.

c.4 Gradation Continuity Check

Note the density variation continuity of the continuous wedge of the test pattern film output generated in the preceeding section. Check that the following conditions are met.

- (1) The density gradient must be smooth.
- (2) There must be no vertical streaks due to the loss of image data bits (no pixel drop out).

d. Film Processor Processing Performance Check

Check that film outputs used for the preceeding checks are dry immediately after being discharged from the film processor.

SYSTEM FUNCTION CHECKS

This section describes the image recorder acceptance inspection procedures to be performed for image processing system function checkout purposes. Note, however, that all the following checks cannot be performed depending on the external equipment connected to the image recorder.

a. Machine Startup Check with Remote Power Control System

- (1) Make sure that the system power is off. Connect the external equipment to the image recorder.
- (2) Set the image recorder **REMOTE/LOCAL** switch to **REMOTE**, and then turn **ON** the circuit breaker on a lateral side of the enclosure. At this time, make sure that the image recorder is turned **OFF**.
- (3) Turn **ON** the external equipment. Check that the image recorder is turned **ON** normally.

Note: If two or more external equipment units are connected to the image recorder, the image recorder turns **ON** when at least one external equipment unit is turned **ON**.

b. Remote Operation Check

- (1) After completion of the steps indicated in the preceeding section, make sure that the image recorder is properly connected to the external equipment (note the external equipment display or indicator).
- (2) Enter the user expansion service suspend mode and press **TEST** key. Note the error log to make sure that no serious error has occurred.
- (3) While the system is idle, insert the tray shutter into the image recorder to intentionally cause error 602. Clear the error condition and check that the system is restored to normal.

c. Machine Shutdown Check with Remote Power Control System

After completion of the steps indicated in the preceeding section, make sure that all processed films are cleared from the image recorder. Turn **OFF** the external equipment. Check that the image recorder then enters the system shutdown sequence and normally terminates in about 3 minutes. Further, turn **OFF** the external equipment to check that the image recorder turns **OFF** in 10 minutes.

Note: If two or more external equipment units are connected to the image recorder, the image recorder does not turn **OFF** or enter the system shutdown sequence until all the external equipment units turn **OFF**.

3.1 Preventive Maintenance (PM) On The Image Recorder, CR-LP414

Procedures for maintenance operations of Image Recorder must be performed periodically to maintain the machine for good image quality and good machine operation. This is discussed extensively in **CHAPTER E OF FUJI IMAGE RECORDER SERVICE MANUAL**.¹²

4. Fuji Imaging Plates

Imaging Plates

Fuji has two basic types of plates, **Standard Type (ST)** and **High Resolution (HR)** which requires about 2.5 times the exposure of ST plates. These basic types have been produced in five generations, with later versions denoted as **I, II, III, IIIN, and V**. Generally, mixing these plates will introduce undesired variability. Each individual plate has their own characteristics and performances. Dust, scratches, and cracking around the edges of these plates will cause artifacts. Therefore, special care for handling these plates are necessary.

Cassettes

Fuji cassettes should be inspected for damage and lead-backing. The cassettes that have lead-backing to reduce back scatter may behave differently in QC testing. Even though ST and HR plates can be distinguished by a plate ID that shows through the window in the cassette, it is useful to color-code the outside of the cassette with a marker. This system of external marking works well for the FCR 7000 and FCR 9000, that can recognize the difference between ST and HR plates, but the AC's series does not distinguish between plate types. The CR plates and cassettes are different in composition from screen-film systems, so x-ray phototimers must be recalibrated. Also the size of the mammography cassettes are different from the other purpose cassettes.

a. Storage Conditions

- Store Fuji Imaging Plates under the following conditions :
 - Before unpacking : below 35°C (95°F)
 - After unpacking : 13 to 33°C (55 to 90°F)
35 to 80% Relative Humidity (RH)
- **Avoid direct sunlight, ultraviolet rays and ionizing radiation.**
- Place Imaging Plate (IP) package horizontally and avoid bending them or subjecting them to strong pressure.

b. Transport Conditions

- When transporting Imaging Plate cartons, never drop them and protect them from shock.
- Up to four cartons of Imaging Plate may be stacked.
- Keep the cartons upright and in a dry place.
- Protect them from high temperatures (45°C/113°F or more), particularly during transport.

c. **Routine Handling**

- Wear soft cotton gloves when handling Imaging Plates. Handle them carefully.
- Do not bend, knock them against other objects, or drop them. Protect them from damage or staining.

d. **Cassettes**

Imaging Plates are to be loaded in special cassettes. The Imaging Plate cassette is similar in appearance to screen-film system cassettes, but is made from special materials for optimal Imaging Plate protection, and provides a window through which the Imaging Plate bar code may be read.

e. **Secondary Erasure**

Imaging Plates will accumulate energy from X-rays, gamma-rays, and other radiation sources even if the energy is low. It is therefore necessary to erase the energy before using, especially when:

- An Imaging Plate is used for the first time in a day's work.
- A high speed exposure at a relative speed of 800 or more is required.

Refer to the **System Handling Manual** for details on providing secondary erasure.

f. **Cleaning**

- For Imaging Plate cleaning, wipe both surfaces, exercising care not to scratch the front surface.
- Remove dirt and dust from Imaging Plate surfaces by lightly wiping with a soft, clean cotton cloth.
- Do not rub hard for removal of dirt and foreign matter, rather wipe gently for removal of such from both surfaces using **ALCOHOL** (e.g., ethanol anhydride or 200 proof ethanol alcohol undenatured).
- Being anti-static, the new Imaging Plates III's Type and V's Type series do not have static build-up problems, and cleaning with Fuji AS Cleaner is unnecessary. Isopropyl alcohol is the most recommended cleaner.
- When the imaging Plate is cleaned with Fuji AS Cleaner, be sure to wipe out with a dry cloth to remove any residue. When Fuji AS Cleaner is to be used, follow the supplied instructions.

g. **Replacement**

Replace Imaging Plates when any of the following conditions exists:

- Scratches on Imaging Plate protective layer surfaces appear on film as scratch image
- Damage to Imaging Plate surfaces, backing layers, or any edges are detected.
- Phosphor layers begin to separate from supports.

h. **Other Precautions**

- Exposed but unread Imaging Plates should not be exposed to light.
- Illumination may be needed when transferring exposed but unread Imaging Plates to a magazine. In that case, use X-ray film safelights, either for blue sensitive materials or orthochromatic ones.

5. Routine Quality Control Procedures

Finally the routine daily, weekly, monthly, semi-annual, and annually quality control procedure must be performed on the CR machine on a regular basis in order to maintain the system's performance for better image quality and better machine operation. These are as follows:

Daily

- system inspection, check chemical levels used in film processor, inspect cassettes and IPs, check film supply, check film density generated from step wedge

Weekly

- clean filters and vents on system and film processor
- clean all IPs with isopropyl alcohol only
- check chemical levels in processor and add as needed
- Image Quality Testing on Image Recorder (i.e., Density check)

Monthly

- processor maintenance, including chemistry replacement and full cleaning of tanks and racks
- Image Quality Testing on Image Reader/Recorder

Semi-Annually

- full preventive maintenance as described in Fuji Service Manual¹¹⁻¹²
- Image Quality Testing on Image Reader/Recorder

Annually

- full preventive maintenance as described in Fuji Service¹¹⁻¹²
- Image Quality Testing on Image Reader/Recorder

6. References

1. Y. Tateno, T. Iinuma, and M. Takano, Computed Radiography, Springer-Verlag, Berlin, Germany, 1987.
2. J.D. Newell, Jr. and C.A. Kelsey, Digital Imaging in Diagnostic Radiology, Churchill Livingstone, Publisher, 1990.
3. R.E. Greene and J.W. Oestmann, Computed Radiography in Clinical Practice, Thieme Medical Publishers, New York, 1992.
4. J. Miyahara, K. Takahashi, Y. Amemiya, N. Kamiya, and Y. Satow, "A new type of x-ray detector utilizing laser stimulated luminescence", Nucl. Instruments Methods Phys. Res., A246, pp. 572-578, 1986.
5. W. Hillen, U. Schiebel, and T. Zaengel, "Imaging performance of a digital storage phosphor system", Med. Phys. vol. 14(5), pp. 744-751, 1987.
6. G.D. Pond, G.W. Seeley, M.T. Yoshino, et. al., "Comparison of conventional film/screen to photostimulable imaging plate radiographs for intraoperative arteriography and cholangiography", Proc. SPIE, vol. 914, p. 138, 1988.
7. A.J. Mefibow, E.R. Beranbaum, E.J. Balthazar, et al., "Computed radiography versus film-screen radiography in double-contrast gastrointestinal radiography", Radiology, vol. 169, p. 349, 1988.
8. J.A. Seibert, "Photostimulable Phosphor System Acceptance Testing", AAPM Summer School, pp. 899-929, 1991.
9. C.E. Willis, J.C. Weiser, R.G. Leckie, J. Romlein, and G. Norton, "Optimization and quality control of computed radiography", Proc. SPIE, vol. 2164, pp. 178-185, 1994.
10. M. Freedman, S.K. Mun, E. Pe, J. Weiser, J. Romlein, S.-C.B. Lo, M. Nelson, "Quality Control of Storage Phosphor Radiography Machines", SCAR, pp. 456-460, 1993.
11. Fuji AC-3 Image Reader: CR-IR318 Service Manual, Vol. 1/1, Document # 007-200-04, Fuji Photo Film Co., Ltd., 1994.
- 12- Fuji Image Recorder : CR-LP414N / FL-IM263N Service Manual, Document # 006-040-00, Fuji Photo Film Co., Ltd., 1991.

Appendix I Troubleshooting On The FCR AC-3, CR-IR318

The Fuji SERVICE MANUAL, Document No.007-200-04, 1994, Chapter F2, Page F2-1, has a flow chart that should be followed to troubleshoot the FCR AC-3 Image Reader.

Appendix II

Troubleshooting On The Image Recorder, CR-LP414

The Fuji Image Recorder, CR-LP414N/FL-IM263N SERVICE MANUAL, Document No.006-040-00, October 31, 1991, Chapter F1, Pages F1-1 thru F1-82, has flow charts that should be used to troubleshoot the Image Recorder.

Appendix III

Standard Radiographs for Future Comparison

Test radiographs, that were produced during the initial configuration and acceptance testing should be kept in this Appendix for future reference.

GAIN OF USING IRREVERSIBLE OVER ERROR-FREE DATA COMPRESSION IN DIGITAL RADIOGRAPHY

Shih-Chung B. Lo, Brian H. Krasner, Matthew T. Freedman, and Seong K. Mun

Radiology Department, Georgetown University Washington D.C. 20007

ABSTRACT

Recent clinical tests indicate that digitization of a sample size of $\approx 180\mu$ seems generally acceptable for most radiological applications in conventional radiography. Both high resolution digitizers and display systems become available in PACS implementations. However, the optimal digital contrast information still remains undetermined. Currently, many digital medical image systems produce 256 (8-bit) to 4096 (12-bit) gray levels without providing a supportive information for the selection.

In our previous studies we found that excessive spatial resolution can be partially recovered by error-free compression. However, the excessive gray dynamic resolution is not reducible as far as information content is concerned. Both information evaluation and preliminary clinical tests indicated that no information existed beyond 9-bit when digitized by a sample size of 180μ with laser film digitization or computed radiography.

In this study, we found that a high compression method can only achieve a small fraction of true compression efficacy over an error-free compression for a well-defined digital radiographic imaging system. This implies that the following two procedures contain similar digital information: (a) digitization of a 12-bit image and processing by a moderate irreversible compression (e.g., DCT type compression) and (b) digitization of the same image 8-bit followed by an error-free compression method (e.g., DPCM/arithmetic coding). Images processed by the above methods require about the same digital storage. The image quality of 12-bit with 5:1 irreversible compression is very close to that of 8-bit with 3:1 error-free compression. Higher compression efficiency (e.g., 0.5 bit/pixel) using procedure (a) would degrade the image quality particularly in edges and small structures. This is because the quantization procedure acts as a filter in the DCT compression. Without the interference of noise, the compression efficiency of using irreversible and reversible compression techniques are comparable as far as information is concerned.

I. INTRODUCTION

In the development of a total digital system for radiology, the most important criterion is to provide radiologists high quality images for diagnosis. The second most important issue needs to be addressed is "Can we deliver a cost effective digital system and still maintain current radiology practice?" It seems that the development of high speed communication channels, advanced computers, dense data storage media, and state-of-the-art display devices makes the dream nearly come true. The progress of computers and digital equipment is phenomenal. On the other hand, we do not see much progress in evolving the technology in radiological environment. This is because the consensus of using a newly developed technology is often slowly reached within the radiology society.

Image data compression is an example of this kind of technology. Although much research has been performed to demonstrate the possible use of an irreversible compression, only error-free compression techniques are seriously considered for the implementation in a PACS. Typically these researches compared the original and decompressed images or demonstrated the technique to save data storage in a digital system. Very few of them have attempted to examine the characteristics of the input information which involves image generation, signal dynamic range, and digital processing. In our previous work, we studied various levels of noise in the digital radiographs and show their impacts when using an error-free compression¹. We also showed appropriate compression methods for different image components². In this study, we did not find irreversible compression can do better than reversible compression techniques for medical images with high signal-to-noise ratio. With an image possessing much noise, the main advantage of using irreversible

was obtained by filtering the noise prior to a reversible compression. In other words, an irreversible compression can be interpreted as a digital signal processing involving a filter (hopefully it filters noises not signals) followed by an error-free coding.

II. MEDICAL IMAGE FORMATION

The purpose of image compression is to retain digital information in a compact form for storage and transmission. This task is closely associated with the formation of the digital image. In the past, many investigators have focused on the compression itself. However, we should evaluate the radiographic information for each image category in order to form the image and to store the image data effectively. In other words, we should not view image formation and compression as separate issues. In fact, image formation is the key point prior to the consideration of the compression issue. Determination of image size and gray level range would greatly impact the amount of the digital information required for an image.

The requirement of a radiographic resolution and the modality's capability in resolving power have been discussed in many papers. Over the last two decades, we have seen the improvement for various image modalities. These researchers would eventually form the consensus of the resolution of an digital modality for clinical use. For example, CT was original formed as 256×256 and gradually evolved to 320×320 and to 512×512 . MRI was original formed as 128×128 and then became 256×256 . Computed radiographic systems generate image data of 2048×2500 matrix with a $14" \times 17"$ film. Digital radiographs through digitization of a film are also made into comparable size to the computed radiography.

Unfortunately, this type of research and development was not investigated as carefully in the determination of the signal dynamic range. Among many digital radiographic modalities the signal to noise ratio of a CT image is very high. The first generation of CT systems were developed in the early 1970s. The image format of CT was progressively increased due to the recent improvement of system capability. CT numbers (pixel values) are ranged from 1000 to -1000 which correspond to physical attenuation coefficients of hard bone to air. In order to resolve 1 out of 2000 attenuation coefficients, a typical CT needs to pump about 5 times as much as exposure used in a conventional projection x-ray. The gray information provided by an MRI is limited by its weak signal. A typical 100:1 signal to noise ratio can be achieved which clearly indicates 12-bit data per pixel is over-stored for MRI.

The recent development of computed radiography (CR) using storage phosphor plate technology has been enthusiastic. Fuji CR systems provide 10-bit per pixel for their raw data, however, the processed images for diagnosis are 8-bit per pixel. Images created by DR systems (e.g., laser digitizer, drum scanner and CCD camera) are formatted into 10-bit to 12-bit per pixel. Fuji CR raw data are not available for the user. The majority of image formation and compression researches in digital radiography has used images digitized by laser scanners. Some CCD camera vendors has considered producing gray value of 16-bit per pixel. The development has become a race in the "dynamic range" in the course of film digitization. This is not a scientific way of determining a system specification in the development. CR vendors tend to like to output a greater number of bits per pixel. We have evaluated three laser scanners and found only 8-bit per pixel containing information in output dynamic range. Unfortunately, marketing decisions sometimes overwrite technical specifications in a company. We found that all vendors use 12-bit for the output of the digital values for their laser film scanner due to business reasons. In fact, medical image vendors tend to provide more information than necessary to protect themselves from being accused of losing image information and to convince customers that they have a "better product".

In the previous investigation for the evaluation of gray dynamic range in CT, MRI and three laser digitizers, we found that CT images contain 12-bit information, MRIs have 7-bit information, and all three laser scanners can only produce 8-bit information. This does not mean that the gray value resolving power of laser scanners is limited to 256 levels. This experiment indicated that films digitized by a laser scanner can only produce 8-bit information. The limitation of the dynamic range is due mainly to quantum mottle of x-ray photons deposited in 180μ pixel area on a film. This is an intrinsic noise of an x-ray procedure with film/screen techniques and may not be contributed by a film digitizer.

Considering a piece of high quality film, the information is decreased if one keeps dividing the testing region into a smaller and smaller area. Film is a media for both detecting and a viewing systems. The eye

integrates the local gray information to determine an area is darker than another or vice versa. When a high power microscope is used for viewing of a film, the gray information will be very different because gray tone from a smaller area is integrated to the eye. In addition, signal-to-noise would be decreased by a factor 2 when a tested region is evenly divided into 4 areas. A large area may have 12-bit information, but a small tested area may not have 12-bit information. Fine digitization (division) does not degrade information capturing, in fact, it enhances digital sampling without losing high frequency information. As long as the display system showing same pixel size as signals were digitized, lower gray level in digital form will be integrated by eyes without any distortion in gray information. On the other hand, if a large pixel size is used the gray information would be altered. This phenomenon applies all digital image systems.

III. ERROR-FREE VERSUS IRREVERSIBLE COMPRESSION

In this study we used several effective methods for the comparison of error-free to irreversible compression. We used polynomial interpolation/arithmetic coding^{1,3} and full-frame entropy coding/splitting gray value^{2,4} for error-free and transform coding, respectively. Twenty CT head images and 20 digital chest radiographs were selected from our clinical PACS as our objects for the compression studies. CT images represent true large dynamic medical images and digital chest radiographs represent medical images containing large noise.

True Large Dynamic Range Medical Image

(a) Error-free compression

A CT image typically consists of a circular region surrounded by air. One can clip air space off and perform a compression for the circular region. Our studies indicated that an average of 36.5% of air space can be eliminated. An additional 25% of empty space is unused in 16-bit data format (only 12-bit containing data). The data reduction efficiency using polynomial interpolation followed by arithmetic coding can further reduce 53% of CT image data. Hence a compressed CT data file only takes about 29.8% (i.e., $(1 - 53\%) \times (1 - 36.5\%)$) of the original 12-bit data or 22.4% (i.e., $29.8\% \times 75\%$) of original data storage (16-bit per pixel).

(b) Full-frame transform/Entropy coding (FFEC)

Transform methods for compression are relatively sensitive to sharp edges. CT head images often contain skull and bones which would contribute large coefficients to the entire transform domain. The purpose of decomposition image data in transform domain is to condense the "energy" of the image in a compact area. Transformation of such an image (e.g., CT head image) consisting of a large contrast area is not an appropriate decomposition method prior to a compression. This large contrast area can be extracted and can be compressed by an error-free compression with a great compression efficiency. The extracted area and the air space should be filled with an average tissue value adjacent to the area. The data reduction efficiency of this method is about 14.7% of original 12-bit CT data without creating observable artifacts⁵. A newly proposed gray value splitting and remapping method is an alternative to tackle the sharp edge problem². The compression efficiency of the split method is comparable to the sharp edge extraction method. Without preprocessing CT images, the compression efficiency ($\approx 20\%$ of original 12-bit CT data) is lower^{5,6} and company artifacts are higher than the above two methods.

(c) Compression gain in CT images

The gain of using composed methods (edge extraction and DCT transform) over error-free compression is almost double (14.4% versus 29.4%). However, composed methods pay the price of data inaccuracy. The data inaccuracy mainly involves the blurriness of structure and the unsharpness of edge due to Gibb's phenomenon. The higher the compression the stronger the phenomenon. Strictly speaking, the gain of using DCT transform over error-free compression is minimal ($\approx 20\%$ versus 29.4%). Compression efficiency and associated compression error are shown in Table I for the comprison of using error-free verus FFEC for CT image studies.

Potentially a large artifact can be shown near a high contrast boundary. Transform methods using 3-D CT data, where the third dimension is stacked with CT slices, can gain data reduction by 5% to 10% depending on the CT slice gap and number of slices used in the transform. Error-free compression using

statistical model for coding would not be beneficial by the third dimension correlation. 2-D or 3-D Lempel-Ziv coding however should be able to gain a few percentage of additional data reduction. These methods are not readily available.

Noise Involved Large Dynamic Range Medical Image

Current laser film digitization and computed radiographs contain a large amount of noise mainly due to Poisson phenomenon when digital output ranges to 12-bit per pixel. We found that little noise contribution directly results from digitization devices. In the same study, we also found that only 8-bit (rounding-off from 12-bit) out of 12-bit data contains information. The residual images of this rounding-off operation consist of noise only which were evaluated by signal-to-noise and local (10×10) correlation studies¹. MR images contain a great deal of noises too. Only 7-bit out of 12-bit data possesses information.

(a) Error-free compression

For chest digital radiographs, polynomial interpolation/arithmetic coding can reduce the digital chest radiographs down to 6.31 bit/pixel (52.4% of 12-bit image data)¹. However, 4-bit (33.3%) out of 12-bit per pixel is noise. This implies that 60% (33.3% out of 52.5%) of the compressed data containing incompressible noise. This is confirmed by using 8-bit data (round-off from 12-bit) for the compression. The polynomial interpolation/arithmetic coding can compress this image data down to 2.64 bit/pixel (i.e., 33.0% based on 8-bit/pixel) which is equivalent to 22.0% ($33.0\% \times 8 / 12$) of 12-bit/pixel.

(b) Full-frame transform/Entropy coding

Full-frame transform methods are particularly advantageous to the compression of large images due to the effectiveness of packing correlated image information in a small area^{7,8,9}. Data reduction down to 5% of 12-bit data was reported for 2K×2K chest radiographs without showing much image quality degradation. Block DCT techniques can achieve a slightly less compression efficiency with potential blocky artifacts, particularly when an image enhancement function is used (e.g., window/level, histogram equalization, or unsharp masking, etc.). Splitting method is recommended to reduce blocky artifacts and to prevent artifacts caused by sharp edges in transform domain. For small image size such as 512×512 matrix, the data can be reduced down to about 12%-10%.

(c) Compression gain in digital radiographs

The gain of using transform compression over error-free compression is about 11 times (5% versus 55%) based on 12-bit per pixel and is about 4 times (7.5% versus 32.4%) based on 8-bit per pixel. Again the data inaccuracy mainly involves the blurriness of structure and the unsharpness of edge in the decompressed large matrix size radiographs using transform coding. If one wishes to practice transform coding carefully, a moderate compression is recommended. In such a condition, the compression gain is reduced by a factor of 2 (15% versus 32.4%) based on 8-bit per pixel as original chest radiograph. By limiting the quantization error in the transform domain, the root mean-square-error of the decompressed image from 12% of data (i.e., compression ratio of 8:1 based on original of 12-bit per pixel) is around 63 which implies an average of 6-bit of data inaccuracy. On the other hand, 6-bit radiographs (round-off from 12-bit data) can be compressed to 15% of original data space used by 12-bit per pixel. By further restriction of the quantization error, the root mean-square-error of decompressed image from 55% of the data (i.e., compression ratio of 2:1) is around 18, implying an average of 4-bit of data inaccuracy. Compression efficiency and associated compression error are shown in Table II for the comparison of using error-free versus FFEC for digital chest radiography studies.

These results indicate that the transform compression techniques do not perform more compression than error-free compression. It mainly acts as a low pass filter and packs the "energy" with an error-free coding. The transform compression filters high frequency information in the transform domain, whereas the latter filters high frequency in the spatial domain. As long as this operation does not greatly impact the observation of an image, a high compression can be achieved. Since a contiguous tone image does not possess large values in the high frequency area in the transform domain, a low pass filtering would not substantially alter the structure of the image. High frequency components represent detail structure, sharp edge information and noise. High compression using transform methods would not be able to preserve them.

IV. DISCUSSION AND CONCLUSIONS

For CT images, we found that the gain of using irreversible image compression versus error-free compression is very small when a moderate error is allowed. When only small error is allowed, the error-free compression is a better choice in terms of compression efficiency and computation speed for compression and decompression.

For digital radiographs, we believe that the image formation of data dynamic range is the most important issue now. It would be not justified to capture large dynamic range of gray values with large noises and to filter them out in the transform domain for compression. A comparable method can be applied in the spatial domain by rounding off noise bits followed by a coding. We believe that gray value dynamic range is a physics issue not a compression issue. Without clearly defining the gray level range in image data, excessive noise is very costly and are not compressible by error-free techniques. More than 60% of data would be noise after an error-free compression. This is a critical issue to the development of PACS because digital radiography would consist of 60-70% data volume of total radiological information. With the implementation of an error-free compression, it would be extremely inefficient to store and to transmit large volumes of digital information and approximately half (70% $\times$ 60%) of them are useless signals. Unfortunately, the consensus of using irreversible compression has not been received for primary diagnosis at this moment. The first phase of PACS implementation most likely would be using error-free compression.

Transform domain compression techniques are commonly considered sophisticated techniques, whereas the operation of using rounding off in the spatial domain is considered to be direct and invasive to the data. As a matter of fact, the data accuracy and compression effect is much more controllable with the latter than with the former methods. For a moderate compression ratio (i.e., 2- to 2.5-bit/pixel in compressed file), error-free compression is much more effective than transform compression techniques. The transform compression methods, however, can achieve higher compression efficiency (e.g., 0.5-bit/pixel) for large size digital radiographs without causing image quality degradation on the decompressed image. When high data accuracy is required, error-free compression which is limited by entropy of the data string (or decomposed data) is the only way for the task. One may choose not to do any compression because the cost effectiveness of compression efficiency is not justified in a digital system. The main negative result of transform compression is the data uncertainty. It is predictable that edges and small structures may result in ripple artifacts. Furthermore, it is difficult for users to indicate the error range in those areas.

It is the authors' opinion that we must define the true gray level dynamic range of an image modality prior to the design a suitable compression. Digital radiographs involving large noise are very costly. In those images, 4-bit out of 16-bit is not used another 4-bit contains no information. True image information is stored in the top 8-bit range in which 68% of the data can be reduced with error-free compression. This is a very significant data reduction that only 16% of original 16-bit data space is required providing 8-bit data acquisition without contamination of digitization noise.

V. ACKNOWLEDGMENTS

This work is supported in part by NIH/NCI Grant 1R29CA59763-01 and by a Whitaker Foundation Grant. The authors are grateful to Ms. Susan Kirby for her editorial assistance.

VI. REFERENCES

- ¹Lo, SC, Krasner, B, and Mun, SK: "Noise Impact On Error-Free Image Compression," *IEEE Trans. Med. Imaging*, vol. 9, 1990, pp. 202-206.
- ²Lo, SC, Shen, A, Mun, SK, and Chen, J: "A Method for Splitting Digital Value in Radiological Image Compression," *Med. Phys.* Vol. 18(5), 1991, pp. 939-946.

³Witten, IH, Neal, RM, and Cleary, JG: "Arithmetic Coding for Data Compression," Comm. of the ACM, Vol. 30, 1987, pp.520-540.

⁴Lo, SC, Krasner, BH, Mun, SK, and Horii, SC.: "Full-Frame Entropy Encoding for Radiological Image Compression," SPIE Proc. Medical Imaging V, 1991. pp. 265-277.

⁵Lo, SC: "Radiological Image Compression," Ph.D. Desertation, UCLA, Los Angeles, CA, 1986.

⁶Lo, SC, and Huang, HK: "Radiological Image Compression: Full-Frame Bit-Allocation Technique," *Radiology*, Vol. 155, 1985, pp. 811-817.

⁷Wintz, PA: "Transform Picture Coding," *Proc. IEEE*, vol. 60, 1972, pp. 809-820.

⁸Jain, AK: "Image Data Compression: a review," *Proc. IEEE*, vol. 69, 1981, pp. 349-389.

⁹Rosenfeld, A, and Kak, AC: "Digital Picture Processing," Academic Press, 1982.

Table I. Compression results and maximum errors with CT images

Compression methods	Compression results based on 12-bit/pixel	Averaged top 1% of maximum error
Polynomial interpolation / arithmetic coding	<u>36.5% of air space</u> compression by EFC (1 - 36.5%) x 47% = 29.8 %	0
Full-frame DCT / arithmetic coding	21%	505
Split method: Alternate value contour coding for 3MSB Full-frame DCT/arithmetic coding for R9LSB	22.5%	48

Table II. Compression results and maximum errors with digital radiographs

Compression methods	Compression results based on 12-bit/pixel	Averaged top 1% of maximum error
Polynomial interpolation / arithmetic coding	use of original 12-bit data 52.4 %	0
	round-off to 8-bit/pixel 22.0 %	8
Full-frame DCT / arithmetic coding	low compression 23.2%	21
	high compression 5.7%	217
Split method: Alternate value contour coding for 3MSB Full-frame DCT/arithmetic coding for R9LSB	low compression 24.6%	19
	high compression 6.2%	134

STRUCTURAL ANALYSIS AND CODING OF MULTIMODAL MEDICAL IMAGES

O. Baudin, A. Baskurt, F. Dupont, R. Prost, R. Goutte

Ultrasound and Signal Processing Laboratory
URA CNRS 1216 - INSA Lyon - France

M. Khamadja

Department of Instrumentation and Measures
Electronic Institute, University of Constantine - Algeria

ABSTRACT

An adaptive image coding scheme based on Discrete Cosine Transform (DCT) is considered. A set of 90 features in the spatial and spectral domains leads to a subset of features which is used to automatically classify subimages, taken from a multimodal medical image data base. The classifier, based on a binary decision tree, discriminates 13 classes. In the DCT domain, a normalization matrix for each class is generated using the features computed on subimages. This matrix allows to select the significant DCT coefficients associated to a class. This method leads to a performant adaptativity for the coding scheme. The classifier is very simple and cheap in computing time. A given subimage is classified, transformed with DCT, normalized by the matrix associated to its class, quantized and coded with Huffman tables.

1. INTRODUCTION

Image analysis is a primordial step in any image processing chain. It becomes the first stage when the processing scheme is applied to several modalities of images. For example, an image coding method needs a good image understanding to be adaptive and optimal.

Several works^{1,2,3} on image analysis use only one class of images. In this paper, we analyze a multimodal image data base. This data base has been created for the project AIM (Advanced Informatics in Medicine) of the Commission of European Communities. It contains about five hundred medical images corresponding to five modalities :

- chest radiography ;
- hand radiography ;
- mammographies ;
- head computed X-rays tomography ;
- magnetic resonance images of heart.

The aim of this work is to develop an "intelligent" coding method allowing to adapt itself to these different modalities. So it seems very important to recognize and to modelize all types of structures on these modalities. Section 2 describes the analysis of the image data base and details the features used to modelize the local information in the images. In section 3, an unsupervised classifier based on these extracted features is presented. The adaptive coding procedure is then detailed (section 4). Finally, some results are given and discussed in section 5.

2. IMAGE ANALYSIS

The quality of an adaptative compression method obviously depends on how accurate the parametric segmentation is.

Thus, the choice of the parametric vector to characterize the regions of an image is critical.

In order to perform this choice as well as possible, a great number of parameters (90) issued from six families have first been computed on five 128 x 128 images of the learning set, each image being subdivided in 16 x 16 blocks. The parametrical families used are :

Family 1 : Statistical parameters of the spatial plan (ex : histogram, mean, variance, entropy,...).

Family 2 : Parameters of the cooccurrence matrix¹ (ex : contrast, entropy, energy, correlation,...).

Family 3 : Parameters of the gray level difference vector³ : (ex : mean, contrast, entropy, second angular moment,...).

Family 4 : Parameters of the run-length matrix² (ex : short run emphasis, long run emphasis, gray level non-uniformity, runlength non- uniformity, run percentage, low gray level emphasis, high gray level emphasis,...).

Family 5 : Parameters of the pixel neighborhood matrix⁴ (ex : small number emphasis, large number emphasis, non-uniformity number, second moment, entropy,...).

Family 6 : Parameters of the Fourier transformed plan⁵ : (ex : energy of the main peak, Laplacien of the main peak, highest contribution frequency, isotropy, energy percentage of quadrans 1 and 2, inertia moment of quadrans 1 and 2,...).

A decorrelation method applied to parameters within a same family allowed us to restrain the number of parameters to 16.

Finally, decorrelation on these remaining parameters (inter-family decorrelation), brought us to choose the 7 following features to constitute our parametric vector :

1) Isotropy of the spectral plan (family 6)

$$\begin{aligned}
 \sigma_u &= \sum_{u \neq 0} \sum_{v \neq 0} u^2 p(u, v) \\
 \sigma_v &= \sum_{u \neq 0} \sum_{v \neq 0} v^2 p(u, v) \\
 \sigma_{uv} &= \sum_{u \neq 0} \sum_{v \neq 0} uv p(u, v)
 \end{aligned}$$

$$\text{ISO} = \frac{|\sigma_u - \sigma_v|}{\sqrt{(\sigma_u - \sigma_v)^2 - 4\sigma_{uv}^2}} \quad \text{where}$$

(1)

The closer ISO is to zero, the more isotropic the spectrum is. This means that there is no privileged direction.

2) Inertia moment of quadrans 1 and 2 of the spectral plan (family 6)

$$INER = \sum_{u \neq 0} \sum_{v \neq 0} \sqrt{(u^2 + v^2)} p(u, v) \quad (2)$$

This parameter measures the spectrum spread around (0, 0). If the spectrum is concentrated around the origine, INER takes low values.

3) Contraste in the co-occurence matrix with a step of 2 (family 2)

$$Con(\delta) = \sum_i \sum_j (i-j)^2 P_{ij} \quad (3)$$

Con (δ) takes high values when there are abrupt changes in gray levels along the direction δ .

4) Correlation in the cooccurence matrix with a step of 2 (family 2)

$$COR(\delta) = \sum_i \sum_j (ij p(i, j) - \mu_x \mu_y) / (\sigma_x \sigma_y) \quad (4)$$

since the cooccurence matrix is symetric, wa have :

$$\mu_x = \mu_y = \sum_i iP(i)$$

$$\sigma_x^2 = \sigma_y^2 = \sum_i i^2 P(i) - \mu^2$$

COR (δ) is a measure of resemblance of the image in the direction δ .

5) Entropy of the gray level histogramme (family 1)

$$H = - \sum_{\text{signal}} p(x) \log(p(x)) \quad (5)$$

H characterizes the quantity of information (order/disorder).

6) Run length non-uniformity (family 4)

$$RLN = \sum_{j=1}^M \frac{(\sum_{i=1}^N p(i, j))^2}{S} \quad (6)$$

RUN has a low value if the runs are equally distributed throughout the lengths.

7) Non uniformity number of the neighborhood matrix (family 5)

$$NUN = \frac{\sum_{s=1}^S \left(\sum_{k=1}^K Q(k,s) \right)^2}{\sum_{k=1}^K \sum_{s=1}^S Q(k, s)} \quad (7)$$

NUN gives information on the coarseness of the image.

The notations used are :

- n : Number of pixels in the block
Is used for normalization.
- δ : Displacement vector
- P_{ij} : Value of the cooccurrence matrix element (i, j)
- $P(i)$: $P(i) = \sum_j P_{ij}$
- μ : $\mu = \sum_i iP(i)$
- σ : $\sigma^2 = \sum_i i^2 p(i) - \mu^2$
- S : Number of sections in the run length matrix $M \times N$.
- $Q(k, s)$: Value of the pixel neighborhood matrix element (k, s)
(the matrix is $K \times S$)
- $P(u, v)$: Spectrum power of element (u, v) :
 $P(u, v) = \|F(u, v)\|^2$
- $p(u, v)$: Normalized spectrum power of element (u, v) :
$$p(u, v) = \frac{P(u, v)}{\sum_{u \neq 0} \sum_{v \neq 0} P(u, v)}$$

3. BLOCK CLASSIFICATION

The seven features described in section 2, are grouped in a vector which is associated to each subimage. We then classify the subimages of the training set as follows :

- 1) An unsupervised automatic classification method is used to form clusters by grouping similar subimages and separating dissimilar ones. The number of classes (k) is fixed and then one subimage is

affected to each class and is called the kernel (A_i , $1 \leq i \leq k$). This method performs alternately the two following stages :

- Identification : subimages (x) are separated in k clusters (P_i , $1 \leq i \leq k$), with

$$P_i = \{ x / \text{for } j \neq i D(x, A_i) \leq D(x, A_j) \}$$
 where D in the Euclidean metric.
- Representation : for each cluster (P_i), a new kernel A_i is computed. This new (A_i) corresponds to the centroid.

The convergence control is realized by a criteria that minimizes the intra-class inertia.

The result is then validated with a visual comparison of subimages for each resulting class. This leads to classes for the image data base. The classification is based on all the computed features of the six families.

2) In order to realize an automatic classifier, we use the k classes obtained with the clustering method. This classifier is based on a binary decision tree constructed automatically (figure 1).

Each node of this tree is a threshold on only one feature. So, the training set is decomposed at each node to lead to terminal segments containing only elements of one class. Parameters which take decision are chosen by maximizing the Kolmogorov-Smirnov test computed on the element under each node.

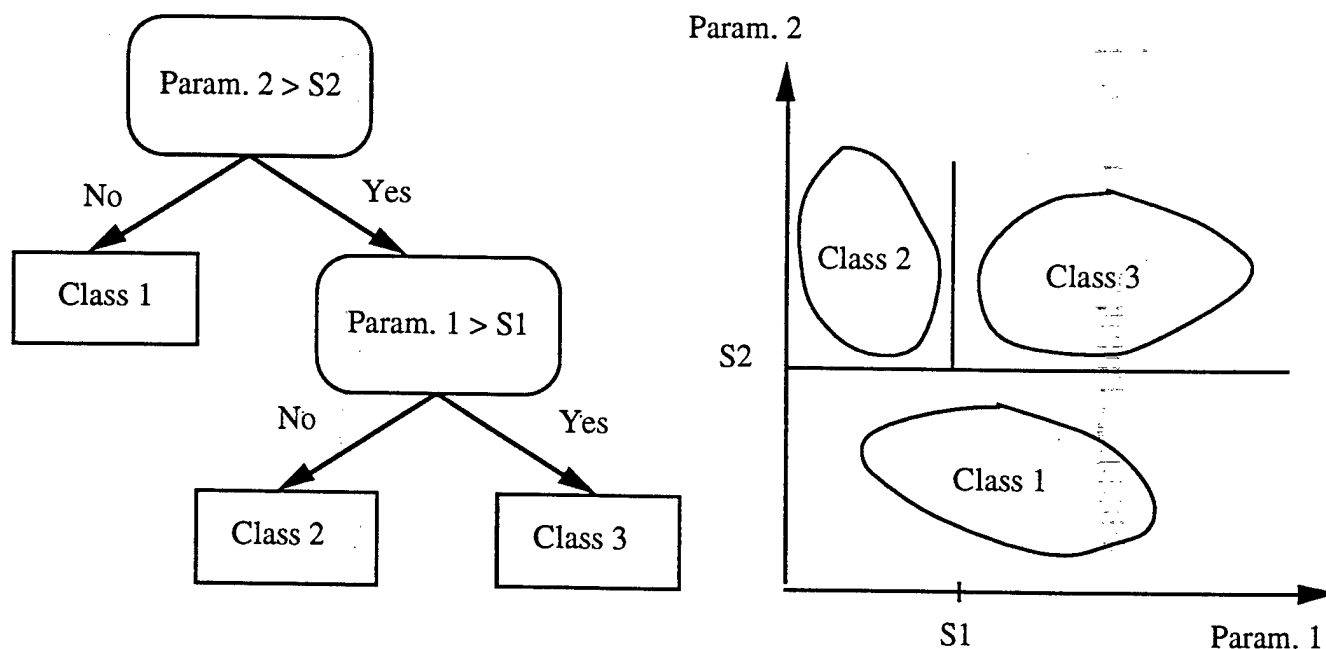


Figure 1. Classifier with two parameters and three classes.

Let us give a simple example with two classes.

(n_i) $i=1,2$ is the number of subimages in the class i

(n_1, n_2) $\rightarrow$ ($n_{1\ell}, n_{2\ell}$)

$\rightarrow$ (n_{1r}, n_{2r})

for $i = 1, 2$

$n_{i\ell} + n_{ir} = n_i$

The threshold on one feature is selected in order to maximize the Kolmogorov-Smirnov (K-S) distance :

$$|n_1 n_2 \ell - n_2 n_1 \ell|$$

This distance is computed for each feature. The decision is taken with the feature which gives the maximum K-S distance (figure 1). This image analysis approach improves processing using multimodal data bases. It leads to a better knowledge on spatial repartition of stationnarity on images. By generating a map of classes for an image, we can locally adapt the coding method.

4. CODING SCHEME

The coding scheme presented here is base on the Joint Photographic Experts Group (JPEG) algorithm⁶. This algorithm has been developed for still image coding. The original image is partitioned into pixel blocks of 8 x 8 in size. Each block is independantly transformed using the Discrete Cosine Transform (DCT). The DCT coefficients of a block are normalized (weighted) by applying a user-defined normalization matrix. This matrix is the same for all the blocks of the image. This normalization allows to select the coefficients which are weighted with low values of the normalization matrix. The normalized DCT coefficients are quantized by rounding to the nearest integer, zigzag reordered and the amplitudes and the run lengths are coded with Huffman tables.

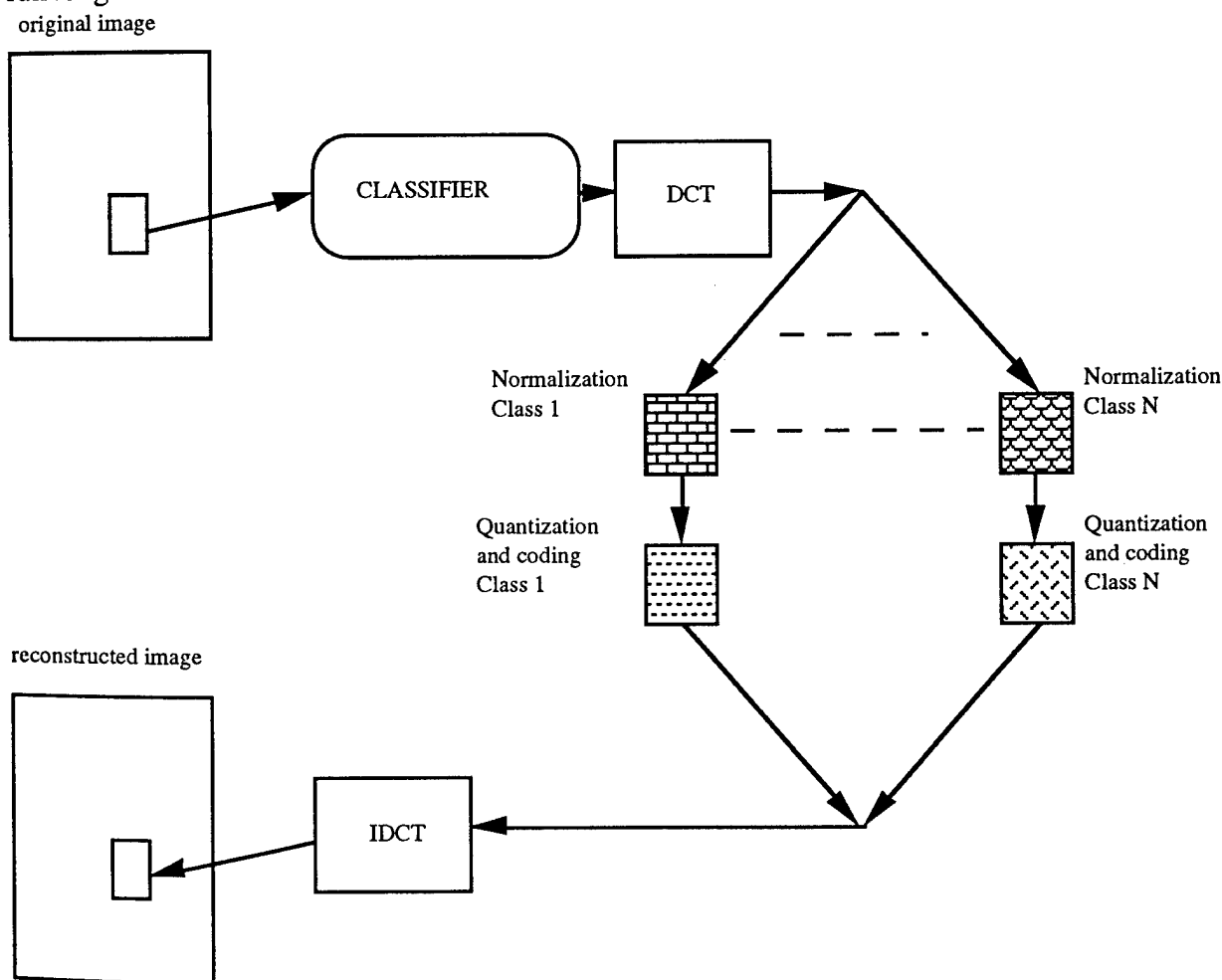


Figure 2. Adaptive coding scheme

In this study, we use the map of classes generated by the image analysis step presented above, to locally adapt the JPEG approach. Figure 2 shows how the coding is processed. Each 16×16 block (this size has been chosen using the mean correlation length of the image data base) is classified through the binary decision tree (figure 1). It is then normalized with a matrix associated to its class. The quantization and coding is similar to JPEG algorithm detailed above. Note that one normalization matrix is associated to one class and that it depends on statistical properties of subimages clustering in each class.

5. RESULTS

We present here some results obtained in this study. A training set of 320 blocks is used to generate the classifier. It is formed by taking 64 blocks of 16×16 size from typical images of each modality. 90 features are computed; 16 of them are retained after intra-family decorrelation and only 7 parameters are chosen after inter-family decorrelation (section 2).

13 classes are chosen with the whole scheme of analyse. Figures 3.a and 3.b show the subimages 16×16 corresponding respectively to class C2 and class C9 for a head computed X-rays tomography. C2 includes the uniform blocks (low contrast ; high non uniformity number for the neighborhood matrix (coarseness of the texture)). Class 9 presents a noisy structure, uniformly distributed, with a finer texture.

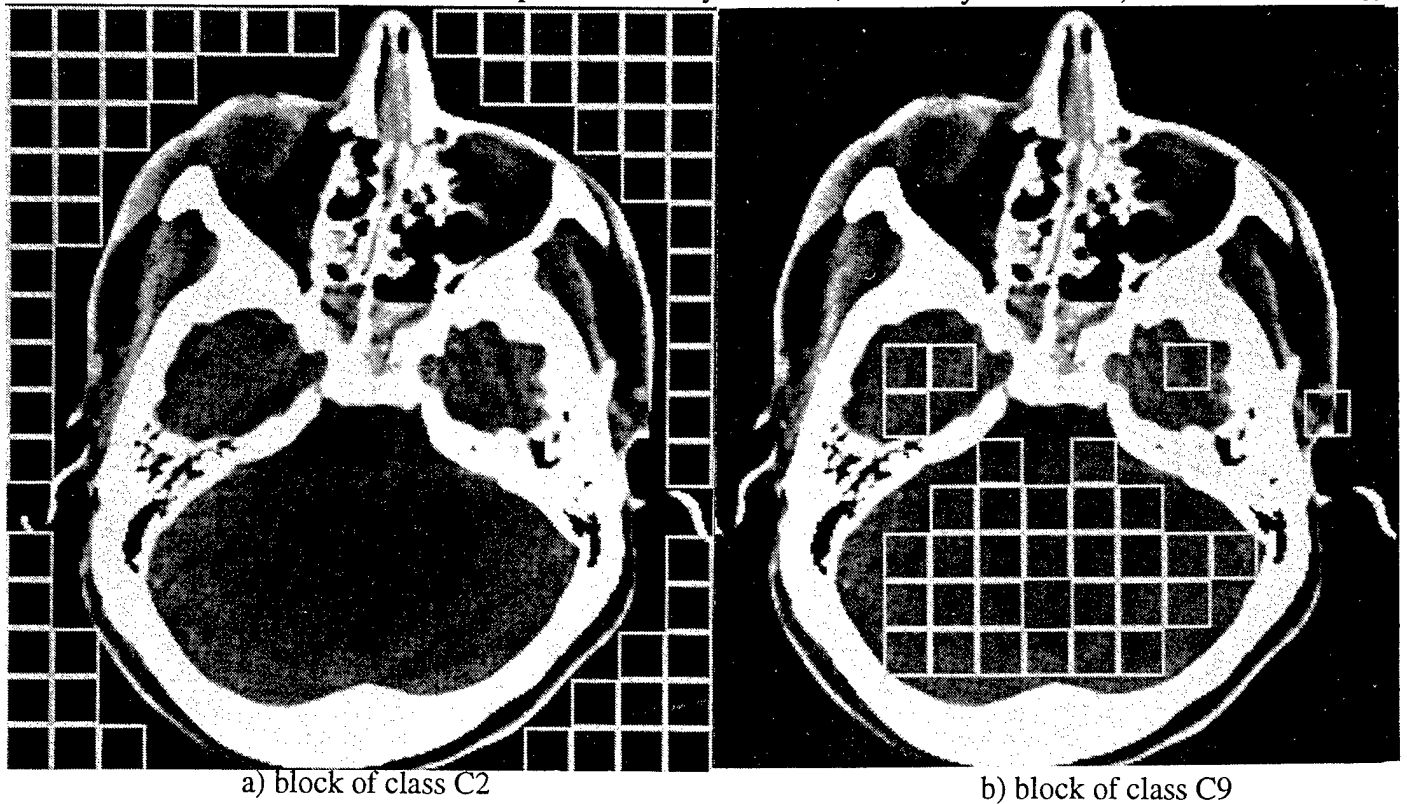


Figure 3. Head computed X-rays tomography

4). 13 normalization matrix are generated using the statistical properties of the DCT domain (section

Finally, the coding scheme (figure 2) is applied on each 16×16 block of a mammogram ($512 \times 512 \times 8$ bits) which is not included in the training set. Figure 4.a presents the original image. The reconstructed one with a compression ratio of 24 (0.33 bits/pixel) and a signal to noise ratio of 47.7 dB is presented in figure 4.b.

The quality of this image gives no visual difference with the original mammogram. The experts have distinguished the same pathologies in the two images. This result is very encouraging, but it has to be validated computing all the multimodal image data base and analyzing the results by experts in mammography. At present, our study is in this validation step.

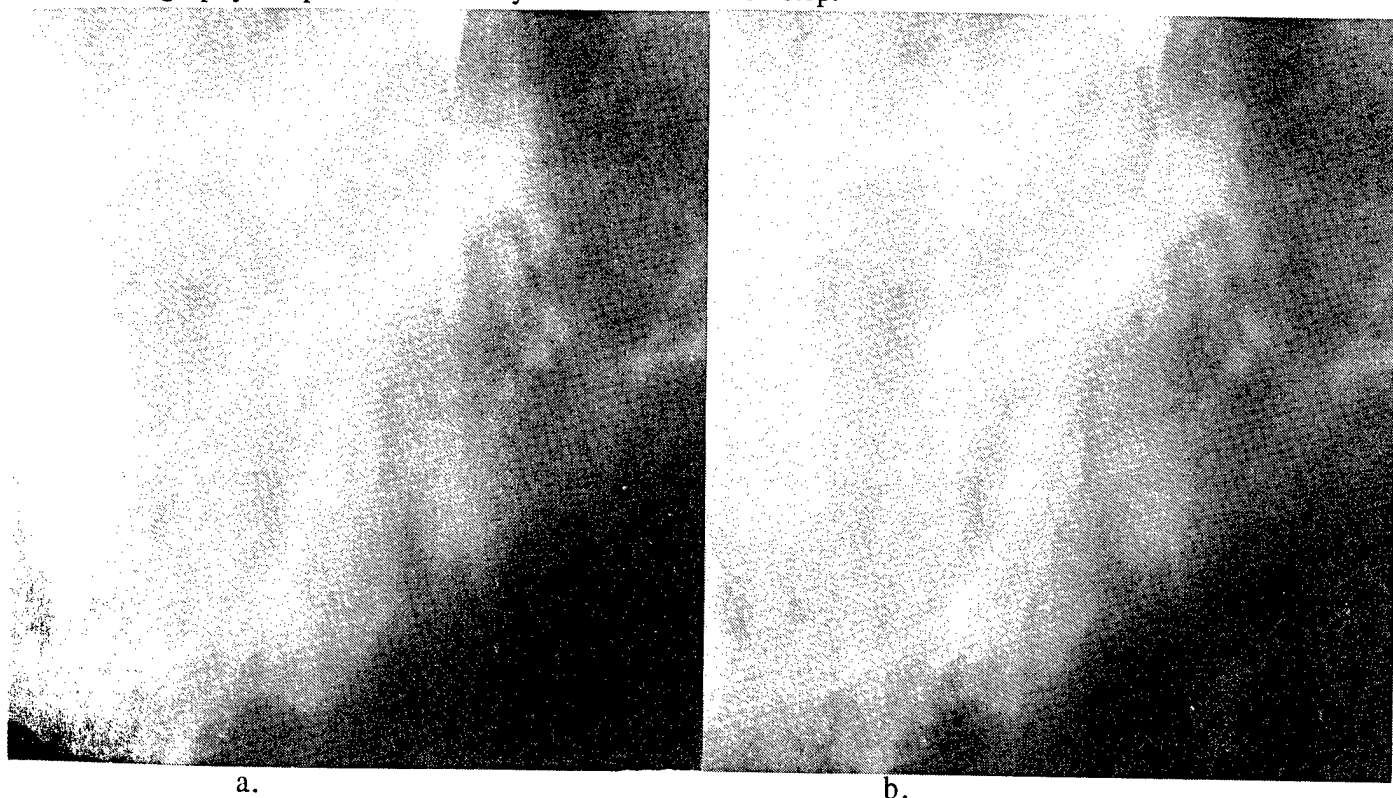


Figure 4.a. Original mammogram ($512 \times 512 \times 8$ bits) 4.b. reconstructed image: compression ratio = 24 (0.33 bits/pixel) ; SNR = 47.7 dB.

6.CONCLUSION

We have presented an adaptive coding scheme improved with an important analysis process of the multimodal image data base. This coding scheme adapts itself to local stationnarity of medical images forming the data base. An automatic adaptation is possible using a simple classifier which discriminates subimages in classes. For each class, we have generated a normalization matrix used in the DCT domain. The application of this matrix allows to select the significant DCT coefficients, taking into account the local properties (a local signature based on the computed features) of each class.

The results obtained on mammograms are very satisfactory. Reconstructed images present a high SNR values (section 5) for high compression ratio. The method will be validated by experts after analyzing the results obtained on all the images of the data base.

7. ACKNOWLEDGEMENTS

The authors wish to thank Nicolas Daron and André Magras for their contributions in this work.

This work is supported in part by the Commission of European Communities within the Advanced Informatics in Medicine program (AIM) EurIPACS project number A2009.

8. REFERENCES

1. R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. Syst., Man, Cybern., vol. SMC-3, pp. 610-621, Nov. 1973.
2. M.M. Galloway, "Texture classification using gray level run lengths," Computer Graphics and Image Processing, vol. 4, pp. 172-179, June 1975.
3. J.S. Weszka, C.R. Dyer and A. Rosenfeld, "A comparative study of texture measures for terrain classification", IEEE Trans. Syst., Man, Cybern., vol. SMC-6, pp. 269-285, April 1976.
4. C. Sun and W.G. Wee, "Neighboring gray level dependance matrix for texture classification", Computer vision, graphics and image processing, vol. 23, pp. 341-352, 1983.
5. S.S. Liu and M.E. Jernigan, "Texture analysis and discrimination in additive noise", Computer vision, graphics and image processing, vol. 49, pp. 52-67, 1990.
6. G.P. Hudson, "Report on the joint photographic expert group meeting", 25-27 jan. 1988, KTAS, Copenhagen, Denmark, ISO/IEC/JTC1/SC2/WG8 N175, Rev. 2, May 1988.

Interactive Wavelet-Based 2-D and 3-D Image Compression

Armando Manduca

Dept. of Physiology and Biophysics
Mayo Clinic and Foundation
Rochester, MN 55905

ABSTRACT

We have developed a software module which performs 2-D and 3-D image compression based on discrete wavelet transform / subband coding techniques. The software allows the user to interactively determine the tradeoff between compression ratio and fidelity (by viewing the results) and to interactively define specific regions of the image - of any size and shape - that are to be preserved with full fidelity while the rest of the image is compressed (again, viewing the results). The compression achieved is superior to the JPEG standard algorithm. We present sample results on a variety of medical images and direct comparisons with JPEG results. We also show examples of the improvements gained by true 3-D compression of a 3-D image (as opposed to 2-D compression of each slice), discuss human visual system response issues, and describe extensions of the current approach to still more efficient compression schemes.

1. INTRODUCTION

Real world images expressed in a wavelet basis usually carry most of their information in a small number of wavelet coefficients, and a large number of coefficients can be truncated to zero (effectively thrown away) with relatively little effect on the appearance of the image. Since only the important coefficients need to be stored, a very natural form of image compression is possible. This is not lossless compression, but there is often a very good tradeoff between the image appearance and the required storage space. Several characteristics make wavelet-based compression particularly interesting: (1) the amount of compression, and the fidelity of the resulting image, are directly determined by the number and accuracy of coefficients kept, so the user can be allowed to determine these, (2) there are theoretical reasons why wavelets should be fundamentally better for image compression than other common transforms, including the Discrete Cosine Transform (DCT), and (3) it is easy to implement a mechanism by which user-specified arbitrary regions of the image are preserved with full fidelity while compression is carried out in the rest of the image.

We have developed a software module in the biomedical image analysis and display package ANALYZE¹ which performs wavelet-based compression on both 2-D and 3-D gray scale images. Using this software, a user can specify certain areas of interest to be preserved at maximum fidelity, while the rest of the image, which only provides context, is compressed. This compression is carried out interactively, and the user can view the result and decide whether to increase or decrease the compression ratio. 3-D images can either be compressed a slice at a time or, if memory allows, with a full 3-D transform, taking full advantage of structure along the third dimension. The compression ratios presently achieved compare very favorably with the JPEG standard. Such a tool may be useful in applications such as teleradiology, or in any situation where data storage or transmission bandwidth is limited and only certain portions of the images must be viewed at the highest resolution. In the sections below, we present examples of the tool's operation, results on a variety of medical images, and discuss various issues arising from the compression scheme.

2. DISCRETE WAVELET TRANSFORMS

Discrete wavelet transforms have recently gained wide application in many areas of signal and image processing; some examples in the imaging domain are spatial filtering, edge detection, feature extraction and texture analysis. Excellent review papers on these subjects are available^{2,3,4,5}. The work presented here is based on the development and algorithms presented by Press² and Simoncelli and Adelson⁶. Like the Fast Fourier Transform (FFT), the Discrete Wavelet Transform (DWT) is a fast, linear operation that operates on a data vector whose length is an integer power of two, transforming it into a numerically different vector of the same length². Also like the FFT, the DWT is invertible and orthogonal, and represents the information in an image in a different way. In the case of the FFT, the basis functions are sines and cosines; in the case of the DWT, they are a

hierarchical set of "wavelet functions" which satisfy certain mathematical criteria and are all translations and scalings of each other. Such a wavelet basis is in a sense intermediate between image data and the frequency spectrum - while image data is fully localized in space but totally unlocalized in frequency, and frequency data is fully localized in frequency but unlocalized in space, a given wavelet function is partially localized both in space and in frequency (or characteristic scale)².

There are infinitely many possible wavelet functions, with various classes more or less suited to particular applications, and perhaps the most important distinction is a tradeoff between how compact they are and how smooth they are. The more compact (and less smooth) ones are best suited for image compression, since they better represent edges and their transforms are the fastest to calculate (calculation time is proportional to the length of the filter). We have implemented two sets of wavelet functions: (1) the class discovered by Daubechies⁷ (we often use the simplest and most compact member of that class, termed DAUB4, because it's the fastest to calculate), and (2) the 5-, 9-, and 13-tap filters presented by Simoncelli and Adelson⁶. We have not yet systematically addressed the question of characterizing which wavelet functions are best for compressing which images and why. The examples below all use the 9-tap filter since it empirically seems to give slightly better performance.

To compute a 1-dimensional DWT^{2,6}, a pair of quadrature mirror filters are defined from the underlying wavelet function, and both are applied to the signal and subsampled by a factor of two. This splits the signal into two components, each of half the original length, with one containing the low-frequency or "smooth" information and the other the high-frequency or "difference" information. The process is performed again on the smooth component, breaking it up into "high-low" and "low-low" components in turn, and this is repeated several times. This yields a multi-level hierarchy, with the initial "difference" component having many points carrying the information in the upper half of the frequency range (so these coefficients are well-localized in space but only slightly localized in frequency), and each higher level having half as many points as the previous level and only half the spread in frequency (so each higher level is progressively more localized in frequency and less localized in space). The final "smooth" signal has only a relatively few points carrying the information in the lowest frequency band. There is a good deal of overlap between this type of wavelet transform and subband coding, and the particular scheme described here is in fact identical to hierarchical subband coding with octave width subbands⁶.

Like the FFT, the DWT is separable, and to apply it to a 2-D image one applies it to each dimension in turn. Usually, the filtering is performed once for each dimension, creating four subimages (low-low, low-high, high-low, high-high in x and y respec



Figure 1. The classic Lena image and the logarithms of the amplitudes of its DWT.

tively), and then recursively performed only on the low-low subimage. The DWT of the classic Lena image is shown in Figure 1. Note the hierarchical scales in each dimension, and how vertical features (high frequency in x, low in y) are captured in the lower right subimages while horizontal features (low in x, high in y) are captured in the upper left subimages and diagonal features (high in both x and y) in the upper right subimages at each level. This process partitions frequency space into octave-spaced oriented subbands⁶. A 3-D DWT can be defined as the obvious extension of the 2-D case, and a similar process with recursion only on the "low-low-low" subimage can be used to calculate a 3-D DWT of a volume image.

3. IMAGE COMPRESSION WITH WAVELETS

Real world images, since they tend to have internal morphological consistency, tend to have first-order correlations (locally similar luminance values), second-order or dipole correlations (e.g., oriented edge continuation), and higher-order correlations (e.g., texture)⁸. It is precisely these correlations that distinguish real-world images from random noise, but this distinction is not exploited in the standard pixel-by-pixel image representation⁸. Wavelet transforms do a good job of efficiently encoding image structure by exploiting these correlations, so most of the information in the image is carried in a relatively small number of coefficients (this is the basic idea behind all transform-based coding methods). There are interesting analogies between wavelet transforms and the way the visual cortex processes incoming visual data in higher animals and humans.

Figure 2 shows the histograms of the Lena image (left) and the amplitudes of the coefficients of its wavelet transform (right). The image histogram is broad and multimodal, and has high entropy. Conversely, most of the transform coefficients have zero or near-zero values (83% of the coefficients fall in the lowest of the 256 bins), and there is a small tail of coefficients with significant amplitudes. It is this latter set that is carrying most of the information in the image, and the large number with small or zero values can be ignored or approximated with little effect on the image quality.

The DWT as described above has very desirable properties for image compression since the coefficients are localized in space (to varying degrees), form a multiscale representation of the image (with a constant scale factor, leading to localized frequency subbands with equal widths on a logarithmic scale), and have some orientation specificity⁶. By contrast, the coefficients in a full-frame DCT have no spatial localization and describe equal-width (rather than equal log-width) frequency subbands, while those in a block DCT have an abrupt spatial localization at a single scale, and the equal-width frequency subbands are not well-localized (due to the abrupt block edges)⁶.

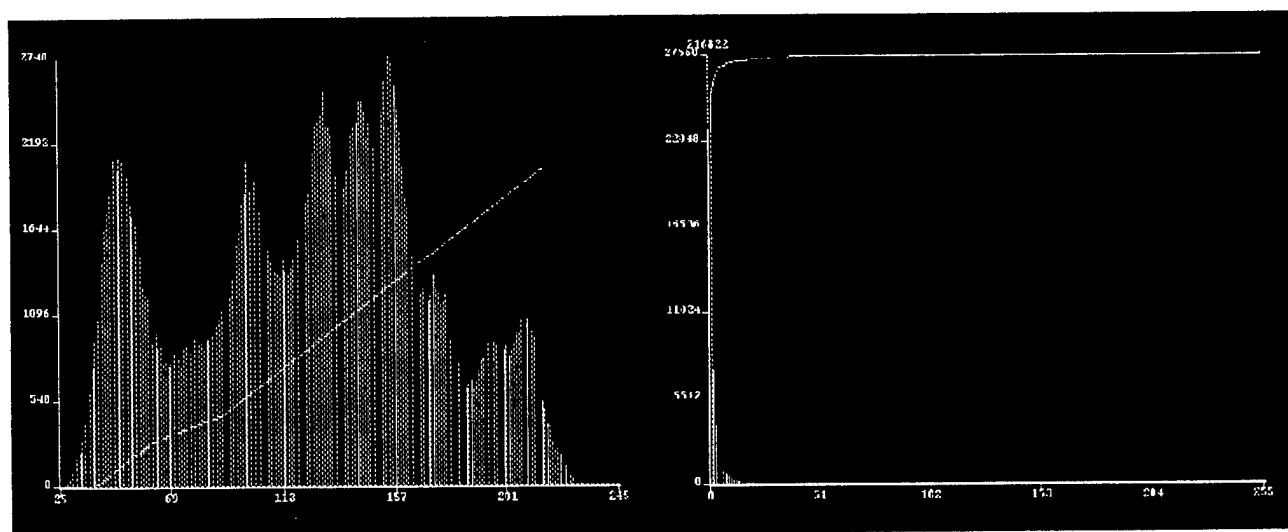


Figure 2. The histograms and cumulative histograms of the Lena image (left) and the amplitudes of its DWT (right). In the latter, the bin with value 0 is off the scale and contains 216822 or 83% of the total values.

3.1. Implementation

The compression scheme currently implemented is based on uniformly quantizing the wavelet coefficients by dividing by a user-specified quantization parameter and rounding off (typically, a large majority of coefficients with very small values are quantized to zero by this step). The zeroes in the resulting sequence are then run-length encoded, and the entire sequence is Huffman encoded. Better coding schemes (e.g. arithmetic coding) are possible and may be implemented in the future. If the quantization parameter is increased, more coefficients are quantized to zero, the remaining ones are quantized more coarsely, the representation accuracy decreases, and the compression ratio increases; if the parameter is decreased, the reverse happens.

To use the program, the user loads an image, selects which wavelet function to use, and invokes a 2-D or 3-D forward transform. The user then specifies the quantization parameter for a trial compression, and the program performs the quantization, does the inverse transform, displays the resulting image, and reports the RMS error and the compression percentage. This process is quite rapid (typically 4 sec. for a 512x512 image on a Sparcstation 2 with the 9-tap filter) and can be repeated while varying the quantization parameter (and hence the compression percentage) as desired. The user can also look at the difference image between the original and the compressed result, or at a map showing which coefficients are quantized with non-zero values. The compressed coefficients can be saved to disk at any time.

3.2. Sample Compressions

Figure 3 shows the Lena image and a compressed version at 8.00% of the original size, which corresponds to 12.5:1 compression and 0.64 bits per pixel (bpp). The RMS difference between the original and compressed images is 3.91. The performance is significantly better than the JPEG standard algorithm¹⁰ in terms of both RMS error and subjective image quality (see below). Figure 4 shows an MRI image and an 8:1 compression (RMS error = 3.50, 1.00 bpp). Figures 5 and 6 show a portion of a chest X-ray (this is 16-bit data; all the other examples are 8-bit data) and a 25:1 compression (RMS error = 11.77, 0.64 bpp). Figures 7 and 8 show a transmission light microscope image of pitting in bone tissue and a 20:1 compression (RMS error = 3.11, 0.40 bpp). Figures 9 and 10 show a rendered image of a wrist and a 10:1 compression (RMS error = 5.74, 0.80 bpp).

3.3. Comparisons with JPEG

In all cases tried, this scheme achieves significantly better compression than the JPEG standard algorithm in terms of both RMS error (which is of course not a good measure of image quality) and subjective quality (although this has not been verified by ROC studies). Figures 11 and 12 show corresponding wavelet and JPEG results for the Lena image at 33.3:1 compression and for a magnified portion at 20:1 compression. In both cases, the artifacts are much worse for the JPEG compressions, and the block edge effects are very objectionable. These are purposefully high compression ratios in order to better show the artifacts introduced by the two methods; at lower compression ratios it becomes more difficult to discern differences.

3.4. Human Visual System Response

Most image compression schemes take advantage of known properties of the human visual system (HVS) and attempt to suppress information which a human viewer would not perceive anyway. For example, JPEG (and DCT-based compression schemes in general) use a "quantization matrix" which quantizes the higher frequencies more coarsely, in accordance with the HVS spatial frequency sensitivity¹⁰. Sometimes more sophisticated schemes are used; e.g., Lewis and Knowles⁹ adjust the quantization parameter for individual transform coefficients based not only on spatial frequency but also on background luminance, edge proximity, and texture masking - all modeling well-documented HVS effects. Such attempts to obtain maximum compression without perceptually altering the image are quite appropriate when the image is meant only to be viewed, especially at the same size as the original. They become more problematical if the compressed image is to be later magnified, manipulated, processed, or automatically interpreted in some way - since information that is not visible to the human in the original image may still be quite relevant; in fact, extracting such information may be the point of the processing. It is also less clear what to do when the data is 3-dimensional and may be viewed from different orientations, or obliquely, or surface or volume rendered. Also, in many medical data sets the resolution is commonly different in the third dimension, further complicating matters. Thus, we prefer to not take advantage of HVS response effects in the general case, although we intend to offer such a scheme as a user-selectable option and are currently researching various alternatives.



Figure 3. The Lena image and a compressed version at 8.0% of the original size.

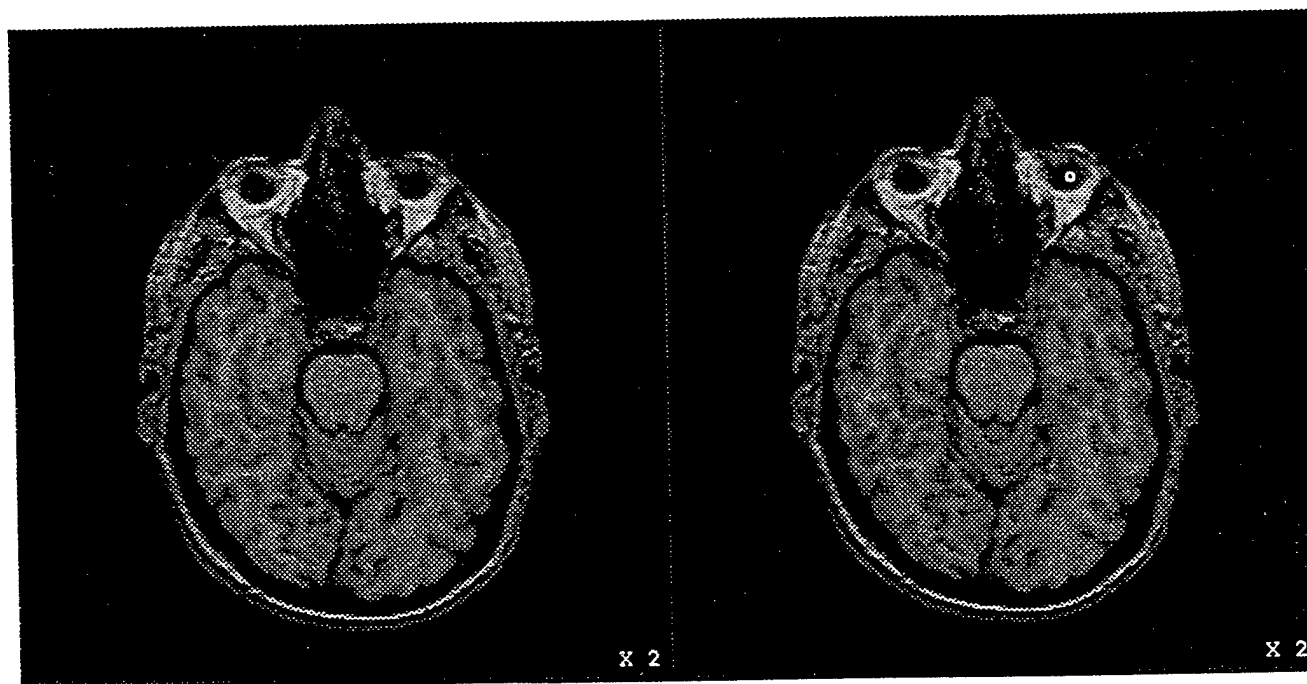


Figure 4. An MRI image and a compressed version at 12.5% of the original size.

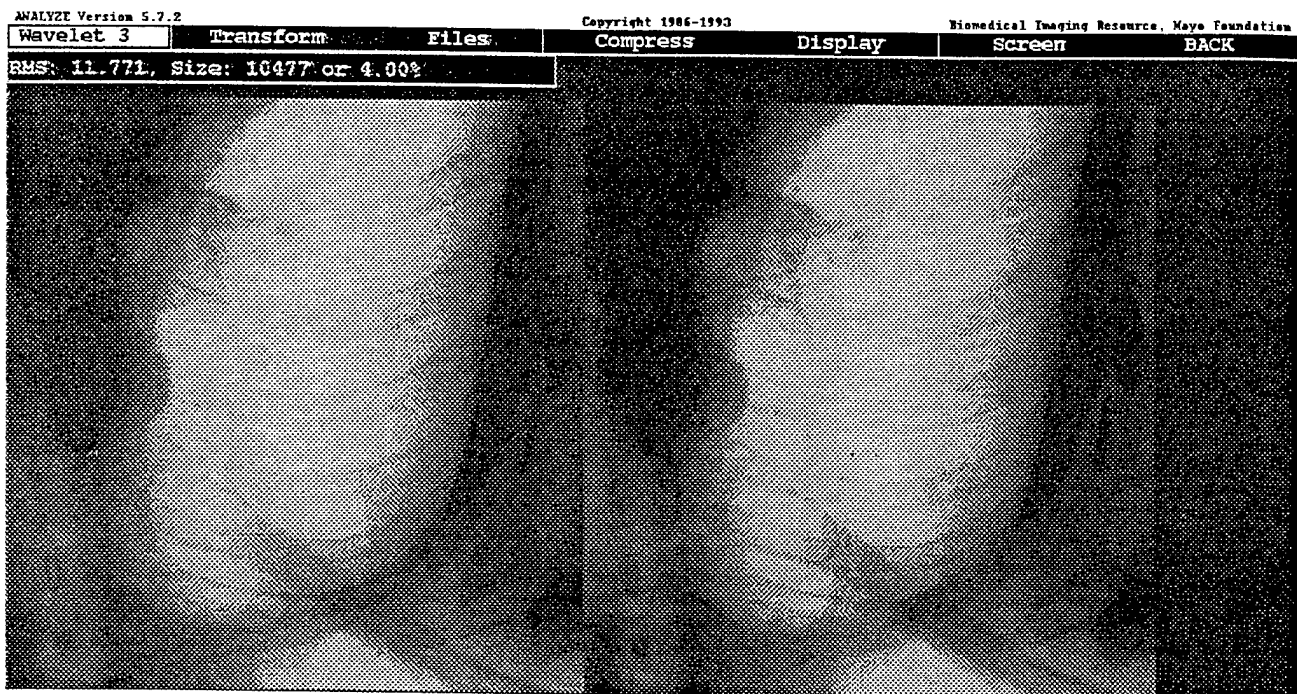


Figure 5. A portion of a chest X-ray and a compressed version at 4.0% of the original size.

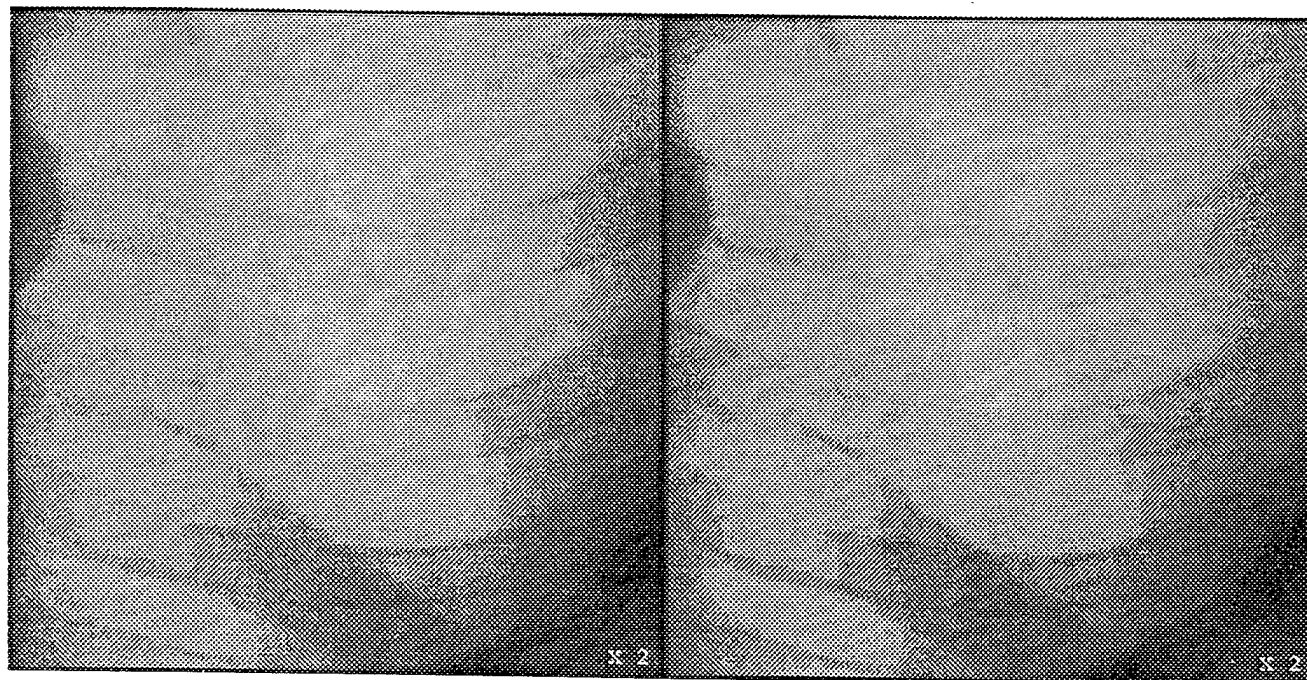


Figure 6. A magnified view of one section of the original and compressed images above.

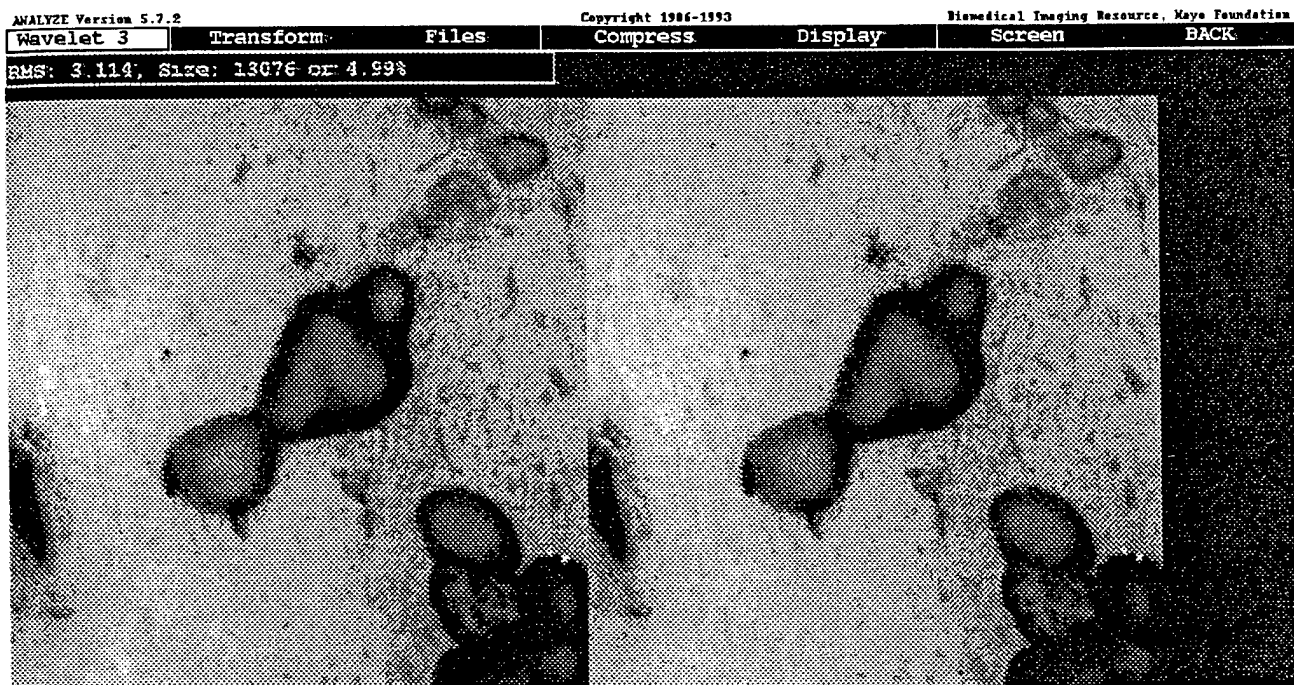


Figure 7. A transmission light microscope image of pitting in bone tissue and a compressed version at 5.0% of the original size.

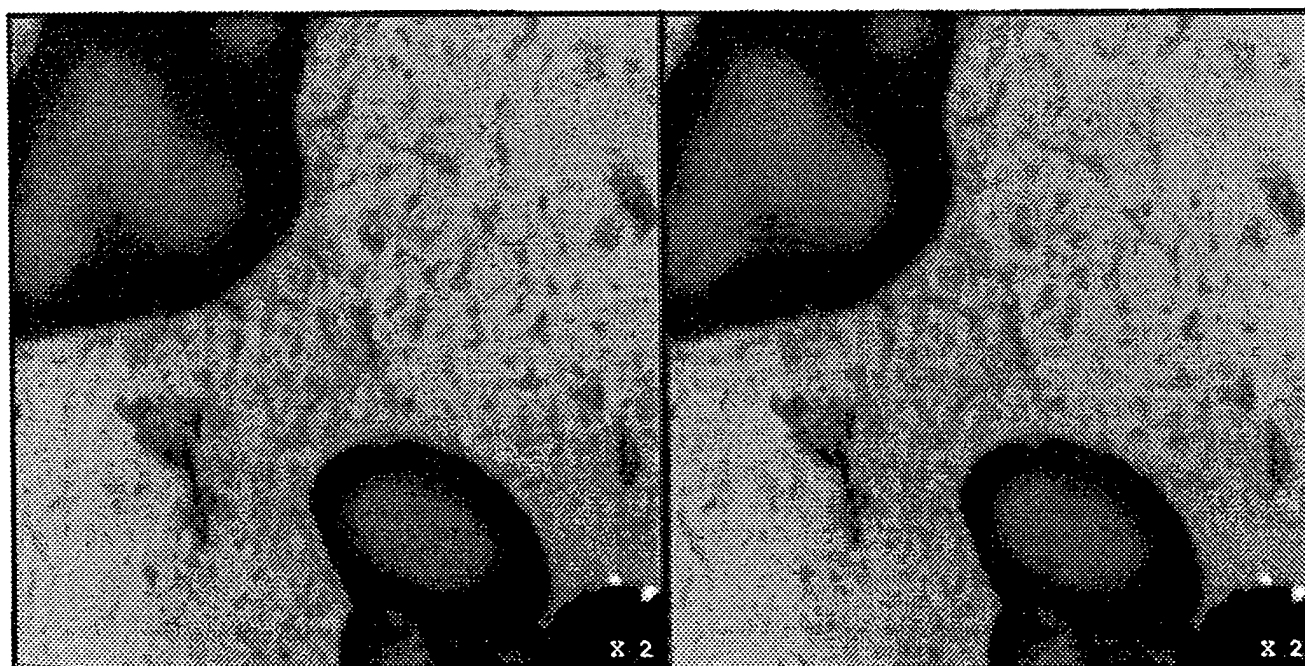


Figure 8. A magnified view of one section of the original and compressed images above.

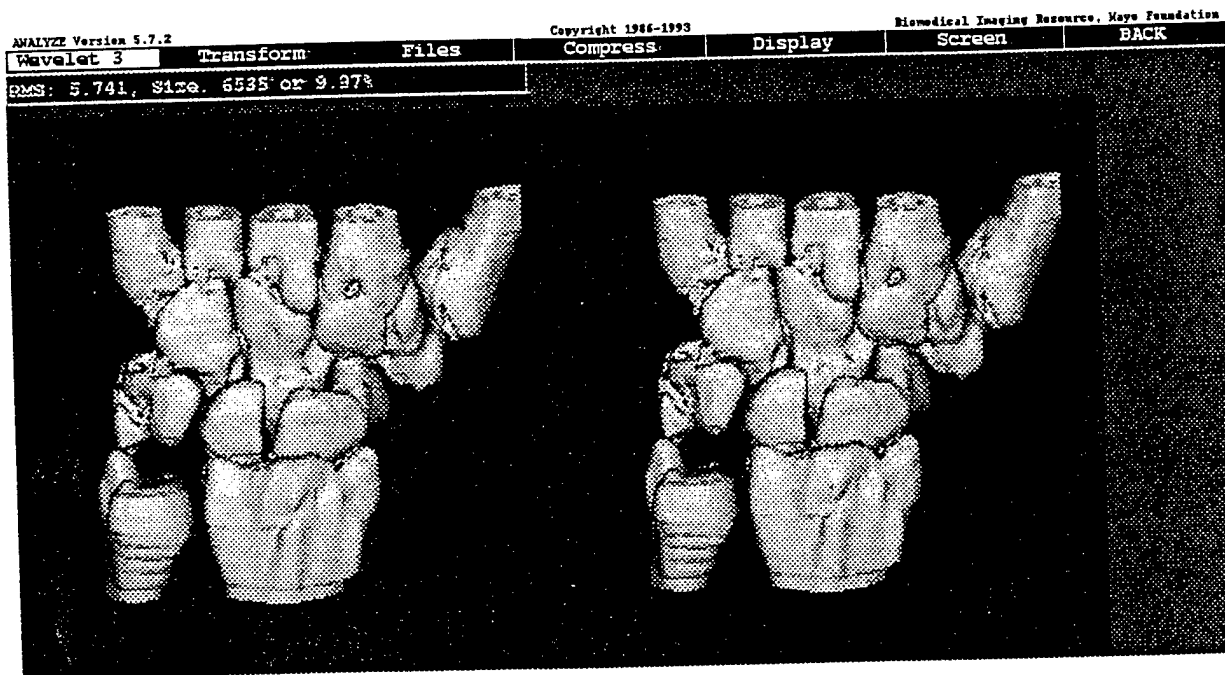


Figure 9. A rendered image of the bones of the wrist and a compressed version at 10% of the original size.

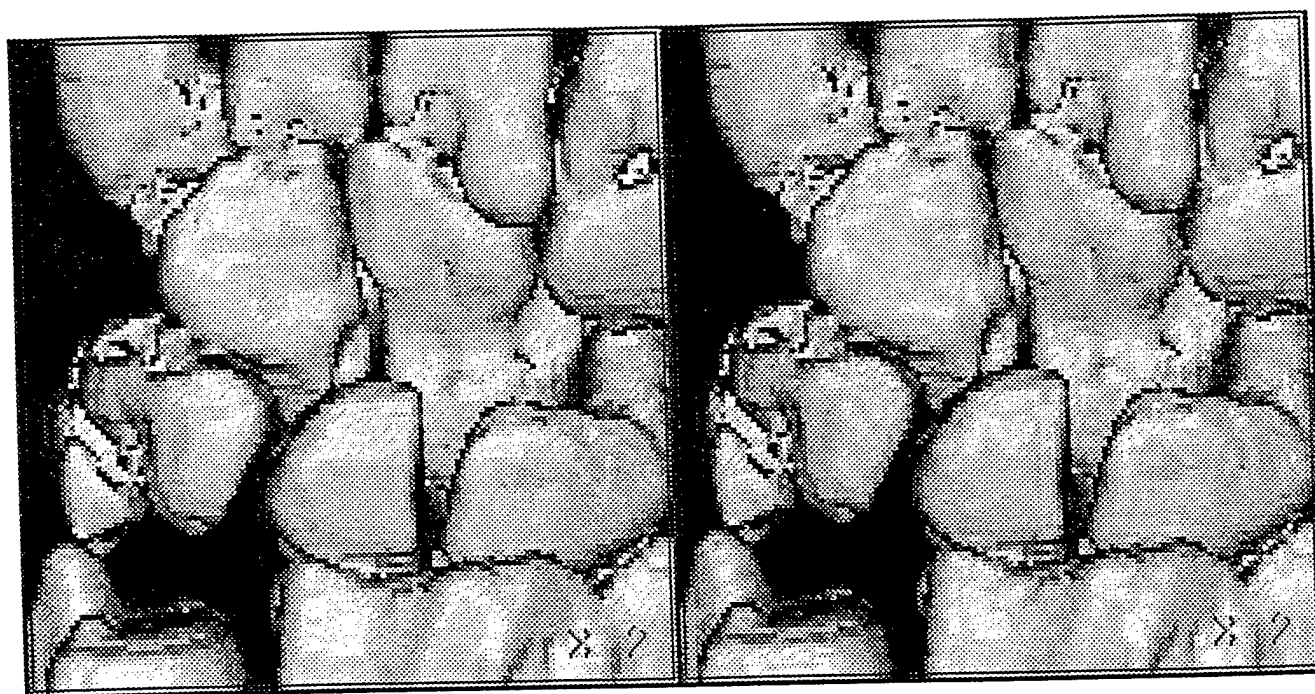


Figure 10. A magnified view of a section of the original and compressed images above.



Figure 11. Compressions of the Lena image to 3.0% of the original size with the wavelet algorithm (left) and the standard JPEG algorithm (right).



Figure 12. A magnified view of a section of the Lena image compressed to 5.0% of the original size with the wavelet algorithm (left) and the standard JPEG algorithm (right).

3.5. Compression of 3-D Images

As mentioned above, a 3-D DWT can be defined as the obvious extension of the 2-D case. The procedure described here can thus be used to perform true 3-D compression of a volume image, taking full advantage of structure along the third dimension, as illustrated in Figure 13. Two sample slices of a 256x256x32 MRI volume image are shown in the left column, and the results of a full 3-D 20:1 compression and individual 2-D 20:1 compressions are shown in the middle and right columns respectively (again, a high compression ratio has been chosen to better show the artifacts introduced). The 3-D compression is significantly better, with much less serious artifacts, as one would expect since the redundancy in the third dimension can be exploited. The RMS errors are 1.92 for the 3-D compression and 2.52 and 2.55 for the 2-D compressions. The obvious drawback is that operating with the full 3-D transform of large volume images requires large amounts of computer memory and processing time (a 256x256x32 transform requires 50 seconds on a Sparcstation 2 for the 9-tap filter).

3.6. Related Advanced Techniques

Simoncelli and Adelson⁶ describe two approaches which are more advanced than the scheme here: one is based on biorthogonal wavelet functions and the other on hexagonal, non-separable filters. The biorthogonal wavelets involve one set of filters (long, relatively slow to compute) for the initial compression, and a second set (very small and fast to compute) for the decompression. The hexagonal filters offer improved orientation selectivity and seem to provide somewhat better performance than the separable filters used here. They also describe a 3-D extension using rhombic dodecahedral filters. It is also known that higher compression ratios can be achieved with a generalization of wavelet transforms known as wave packets⁵, although at the expense of increased processing time. Mallat and Zhong¹¹ describe a more complicated approach based on the evolution of wavelet local maxima across scales and an iterative reconstruction of images from these maxima alone which may yield even higher compression ratios.

3.7. Preserving Arbitrary Regions

The partially localized nature of the wavelet transform means that the value of any one pixel in the image depends on only a small number of wavelet coefficients. Thus, it is possible to specify an arbitrary region of the image and prevent that region from being badly degraded during the compression process by simply representing more accurately (i.e., quantizing more finely) the wavelet coefficients which map to that region, and noting which coefficients which are treated in this way. In a previous implementation¹², with a somewhat different compression scheme (the wavelet coefficients were sorted by amplitude, a chosen percentage of the largest ones quantized and the rest truncated to zero), such an approach worked well. All coefficients corresponding to the arbitrary regions were kept, in addition to the specified percentage of the largest ones, and those regions were preserved with maximum fidelity as the rest of the image was compressed.

We have since realized, however, that even better results are obtained with a very naive approach: to the compressed image, we append a compressed version of a residual image which contains the difference values between the compressed and original images in the regions of interest and which is zero elsewhere. This residual image compresses very well, and restores the regions of interest to precisely the original values. It is this implementation which is currently in place, and the user is allowed to define any number of such regions, either as rectangular areas or by tracing out arbitrary shapes on the image. If the image is 3-D, the user also specifies to which slices each such region is applicable. The effects of preserving a region in this manner are shown in Figure 14. The top row shows the wrist image, a badly degraded 40:1 compression, and the difference between the two. In the bottom row, a region has been selected to be preserved. The compressed image, which now totals 10.2% of its original size, preserves the original detail in this area, and the difference image shows clearly that there are no differences between the compressed image and the original in the designated area but that the large differences remain elsewhere. This approach, of course, is independent of the use of wavelets for the initial compression, and works equally well with any other type of image compression. We are investigating whether the localized nature of the wavelet transform can be exploited to preserve arbitrary regions in an even more efficient manner with the current compression scheme.

4. ACKNOWLEDGEMENTS

Thanks are due to Dr. Richard Robb, for making available the ANALYZE package and allowing me to add this module to it, and to the entire Biomedical Imaging Resource programming and support staff for their help.

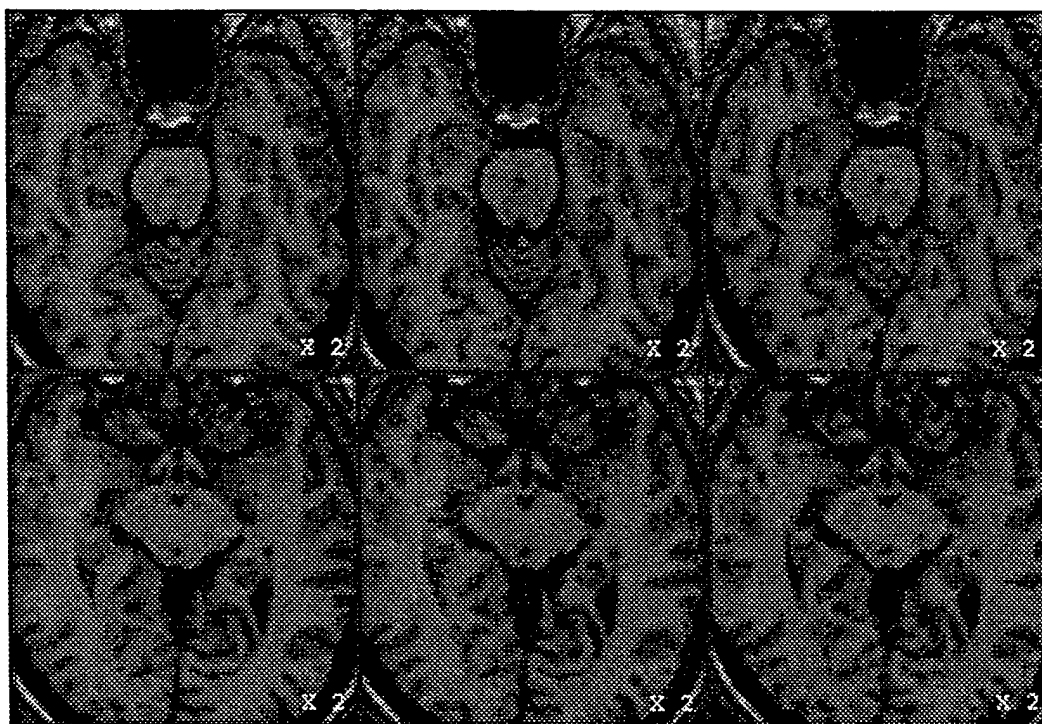
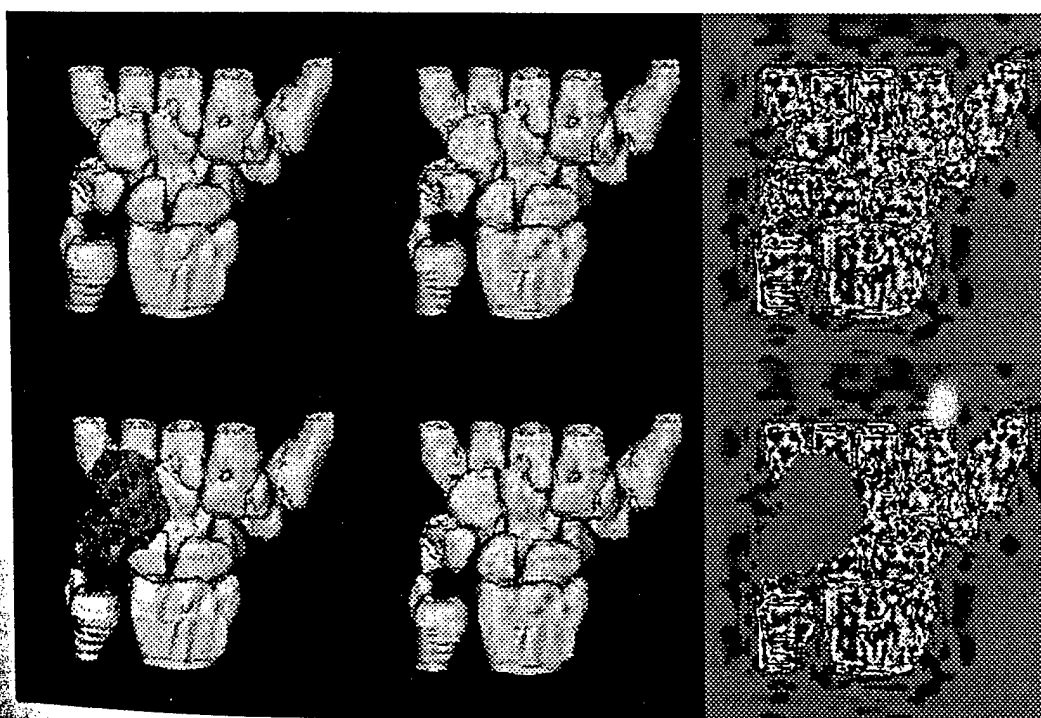


Figure 13. Two sample slices of a 3D MRI data set (left column), the slices after 3-D compression to 5.0% of original size (middle column), and the slices after individual 2-D compression to 5.0% of original size (right column).



Top row: the rendered wrist image, a 40:1 compression, and the difference between the two. Bottom row: the wrist image with a region designated to be preserved, the resulting compressed image (now requiring 10.2% of the original size), and the difference between the two.

5. REFERENCES

1. Robb, R. 1990, "A Software System for Interactive and Quantitative Analysis of Biomedical Images", in K.H. Hohne, H. Fuchs, and S.M. Pizer (eds.), *3D Imaging in Medicine*, p. 333, NATO ASI Series Vol. F60, Springer-Verlag, Berlin.
2. Press, W. 1991, "Wavelet Transforms: A Primer", MIT Center for Astrophysics Preprint Series No. 3184.
3. Pentland, A. 1990, "Equilibrium and Interpolation Solutions Using Wavelet Bases", MIT Media Lab Technical report No. 143.
4. Rioul, O. and Vetterli, M. 1991, "Wavelets and Signal Processing", *IEEE Signal Processing Magazine*, Oct. 1991, p. 14.
5. Wickerhauser, M. 1991, "Picture Compression by Best-Basis Sub-Band Coding", preprint, Yale Univ.
6. Simoncelli, E. and Adelson, E. 1991, "Subband Transforms", in J.W. Woods (ed.), *Subband Image Coding*, p. 143, Kluwer Academic Publishers, Norwell, MA.
7. Daubechies, I. 1988, "Orthonormal Bases of Compactly Supported Wavelets", *Comm. on Pure and Applied Mathematics*, 41, p. 909.
8. Daugman, J. 1988, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression", *IEEE Trans. ASSP*, 36, no. 7, p. 1169.
9. Lewis, A. and Knowles, G. 1992, "Image Compression Using the 2-D Wavelet Transform", *IEEE Trans. Image Processing*, 1, p. 244.
10. Rabbani, M. and Jones, P. 1991, *Digital Image Compression Techniques*, SPIE, Bellingham, WA.
11. Mallat, S. and Zhong, S. 1992, "Characterization of Signals from Multiscale Edges", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14, p. 710.
12. Manduca, A. 1992, "A Wavelet-Based Tool for Medical Image Compression", in Y. Kim (ed.), *Medical Imaging VI: Image Capture, Formatting and Display*, Proc. SPIE, vol. 1653, p. 495.

Classification Of Microcalcifications In Radiographs Of Pathological Specimen For The Diagnosis Of Breast Cancer

Y. Chris Wu, Shih-Chung B. Lo, Matthew T. Freedman,
Akira Hasegawa, Rebecca A. Zuurbier, Seong K. Mun

Georgetown University, Department of Radiology, ISIS Center, Washington, DC 20007

ABSTRACT

Mammography is a proven method for the detection of breast cancer at an early stage. However, approximately 10-30% of breast cancers are missed by mammography screening. Moreover, only about 20% of patients who undergo biopsy prove to actually have cancer. Any improvement in this percentage should have a major impact on the diagnostic efficacy and cost of medical care. Microcalcifications are common findings in benign conditions of the breast, but are also an important sign leading to the detection of breast cancer. The classification of microcalcifications for the diagnosis of breast cancer has been an important and yet difficult task in mammography. The close resemblance between microcalcifications associated with benign and malignant conditions has resulted in radiologists classifying many breast calcifications into an indeterminate category. Improvements in the classification of calcifications would decrease the number of unnecessary breast biopsies and thereby the cost of breast cancer detection. Furthermore, differing levels of expertise among radiologists or simple oversight may lead to failure to diagnose breast cancer or inconsistent recommendations for courses of action. A computer vision system that can classify microcalcifications objectively and consistently can be useful to aid radiologists in the diagnosis of breast cancer.

A convolution neural network (CNN) was employed to classify benign and malignant microcalcifications in the radiographs of pathological specimen. The input signals to the CNN were the pixel values of image blocks centered on each of the suspected microcalcifications. The CNN has been shown to be capable of recognizing different image patterns. Digital images were acquired by digitizing radiographs at a high resolution of $21\ \mu\text{m} \times 21\ \mu\text{m}$. Eighty regions of interests (ROIs) selected from digitized radiographs of pathological specimen were used for the training and testing of the neural network system. The performance of the neural network system was analyzed using the ROC analysis.

1. INTRODUCTION

Breast cancer is one of the leading causes of death among women. However, there is clear evidence that early diagnosis and subsequent treatment can significantly improve the chance of survival for patients with breast cancer (refs. 1-3). Mammography has become one of the major diagnostic procedures with a proven capability for detecting early-stage, clinically occult breast cancers (refs. 4-7). Earlier detection has resulted in decreased mortality (ref. 8). However, breast cancers in their early stage are small and frequently their radiographic appearance differs only subtly from that of normal tissue or benign abnormalities. Because of this subtlety, the potential for misclassification and oversight ("miss") by radiologists is substantial. Only 10-30% of cases that have mammographically suspicious findings and are subjected to biopsy prove to be malignant (ref. 9). Approximately 10-30% of patients with breast cancer are misdiagnosed by mammography (have the cancer missed or not detected on their mammograms) (refs. 10-14). Besides the subtle nature of radiographic lesions associated with breast cancer, many errors in radiological diagnoses can be attributed to human factors such as subjective or varying decision criteria, distraction by other image features, and of simple oversight (refs. 15-17). Studies suggest that these errors may occur even with experienced radiologists (refs. 18-19). These errors may be reduced by the use of automated detection schemes that locate and classify possible lesions, thereby alerting the radiologist to examine these areas with particular attention. Moreover, the automated detection schemes can serve as a "second radiologist", similar to the double reading by two radiologists that is commonly practiced in diagnostic radiology to increase diagnostic efficacy.

Microcalcifications are commonly considered to be important signs of breast cancer. It has been reported that 30-50% of breast cancers detected radiographically demonstrate microcalcifications on mammograms (refs. 20-25). Up to 90% of cases of ductal carcinoma in situ present with microcalcifications (ref. 26). The correlation between the presence of microcalcifications and the presence of breast cancer suggests that accurate detection of microcalcifications will improve the efficacy of mammography as a diagnostic procedure.

Microcalcifications occur in malignant and benign conditions. Some microcalcifications are characteristically benign or are associated with a benign process. Calcified fibroadenomas have a typical "popcorn" configuration appearing coarse and solitary. Milk of calcium demonstrates sedimentation (ref. 27). Vascular calcifications have a tram track appearance, typical of vascular calcifications seen in other areas of the body. Dermal calcifications tend to be smooth and round with lucent centers. Secretory calcifications are thick, smooth, cigar-shaped, and usually non-branching.

Some microcalcifications associated with malignancy have a typically granular or linear appearance. They usually occur in clusters consisting of greater than 15 particles (ref. 28). The particle size is small (less than 1 mm) and shape is irregular (ref. 29).

Some clusters of microcalcifications have neither the typically benign nor typically malignant configurations described above. These "indeterminate" microcalcifications present a significant diagnostic problem and require careful analysis. Features supporting benignity include uniform size and density of the calcium flecks, as is seen in sclerosis adenosis (refs. 30). Benign microcalcifications tend to be uniformly dense or scattered, without a segmental or linear distribution (ref. 31). The number of microcalcifications per cm^2 has been shown to be the most important predictor of malignancy, with clusters consisting of less than 10 microcalcifications per cm^2 having a high chance of benignity. Clusters consisting of microcalcifications numbering greater than 15 per cm^2 have a higher chance of malignancy (ref. 32).

The task of detection and classification of microcalcifications for the diagnosis of breast cancer is a difficult one. The inability to correctly predict cancer is not only due to the overlap in appearance between microcalcifications associated with benign and malignant conditions. Dense breasts, improper technical factors or simple oversight by radiologists may contribute to the failure to detect microcalcifications. Differing levels of confidence and training among interpreting radiologists may lead to inconsistent recommendations for management.

Radiologists classify breast microcalcifications into one of three groups: benign, likely malignant and indeterminate. Most patients with indeterminate types of calcifications undergo a breast biopsy to exclude cancer. Any method that would correctly classify benign types of calcifications previously considered indeterminate would decrease the frequency of biopsy and therefore the cost of detection of breast cancer.

Several investigators have been developing computer programs for the automated detection of microcalcifications on mammograms (refs. 33-36). Chan, *et al.* (ref. 37) showed that the computer program can detect subtle microcalcifications that may be missed by radiologists, indicating that it is a promising approach to the automated detection of microcalcifications. More recently, Wu, *et al.* (ref. 38) applied an artificial neural network (ANN) to detect microcalcifications. The ANN, trained by using the power spectrum of regions of interests (ROI) containing microcalcifications, was able to eliminate 50% of false-positive detections of a rule-based scheme (refs. 37, 39) while preserving more than 95% of the true-positive detections. The neural network achieved an A_z value of 0.85 for the detection of clustered microcalcifications. Several other computer schemes for detection of microcalcifications were also reported by Astley, *et al.* (ref. 40), based on likelihood estimators, by Grimaud, *et al.* (ref. 41), using mathematical morphology tools, and by Karssemeijer (ref. 42), using a stochastic method based on Bayesian decision theory.

As stated earlier, microcalcifications can be associated with either benign or malignant processes. It is important to distinguish different types of microcalcifications after they have been identified by a detection

scheme. Accurate classification of microcalcifications into benign and malignant groups would help improve the sensitivity of the diagnosis as well as reduce the number of unnecessary biopsies.

As the first step in the process of developing an automated computer scheme for classification of microcalcifications, a neural network system will be developed to classify microcalcifications in the radiographs of pathological specimen. Microcalcifications in the radiographs of pathological specimen are more clearly represented than those in regular mammograms. In pathological specimen, underlying tissue around microcalcifications is less than that present in normal mammograms. Therefore, the scatter radiation recorded on films is reduced, resulting in improved contrast and geometrical unsharpness. Larger exposure can be used to achieve greater signal to noise ratio and thereby improve image quality. It is a simplified version of the difficult task of the classification of microcalcifications in mammograms to use radiographs of specimen. Once proved successful, the algorithm will become the foundation for developing schemes for classification of microcalcifications in regular mammograms.

In recent years, rapid progress of research on artificial neural networks (ANN) (ref. 43) has been reported extensively in the field of computer science and many applied fields. Neural networks address detection, classification, and decision-making problems not by pre-specified "conventional" algorithms, but rather by "learning" from examples presented repeatedly. The popularity of neural networks is primarily due to their apparent ability to make decisions and draw conclusions when presented with complex, noisy, or partial information and to adapt their behavior to the properties of the training data. Neural networks are capable of parallel-processing a large amount of information simultaneously and have been shown (ref. 43) to be a useful tool for pattern recognition in fields where conventional algorithmic approaches and rule-based expert systems may not be successful.

2. MATERIALS AND METHODS

The overall approach for the classification of microcalcifications using a CNN system is shown in Fig. 1. The radiographs of pathological specimen were digitized by a high resolution digitizer. Regions of interest (ROI) containing microcalcifications were selected. These ROIs were preprocessed and used as input to the CNN system. Finally, the classification results were examined by the ROC analysis.

The classification of microcalcifications is based primarily on the fact that microcalcifications associated with malignant processes generally have more irregular shape with fuzzy boundaries and are less uniform in density and size. They are usually grouped into multi-particle clusters. The microcalcifications associated with benign processes, on the other hand, usually tend to have smoother and well-defined boundaries, rounded shape, and uniform densities and sizes. The neural network system will be trained to recognize the characteristics of each type of microcalcification.

2.1. Acquisition of Mammograms

The selected radiographs of breast biopsy specimen will be digitized with an image resolution of $21\mu\text{m} \times 21\mu\text{m}$ per pixel by a CCD camera digitizer (DBA Systems Inc.). With the high resolution digitization, the morphological information of microcalcifications can be preserved, which enables the neural network system to differentiate different types of microcalcifications on the basis of their geometrical shapes and density patterns. Figure 2(a) shows a cluster of microcalcifications in original radiographs of pathological specimen. Shown in Fig. 2(b) are microcalcifications after being digitized with the high resolution digitizer. The shapes and density patterns of the microcalcifications are much better defined in Fig. 2(b) than those in Fig. 2(a).

Eighty ROIs that contains microcalcifications (40 benign and 40 malignant) were selected from digitized radiographs of pathological specimen in this study. The information concerning the classification of microcalcifications ("truth") were obtained from the results of biopsy examination. Background trend correction were first employed to alleviate the interference of the background trend with microcalcification signals.

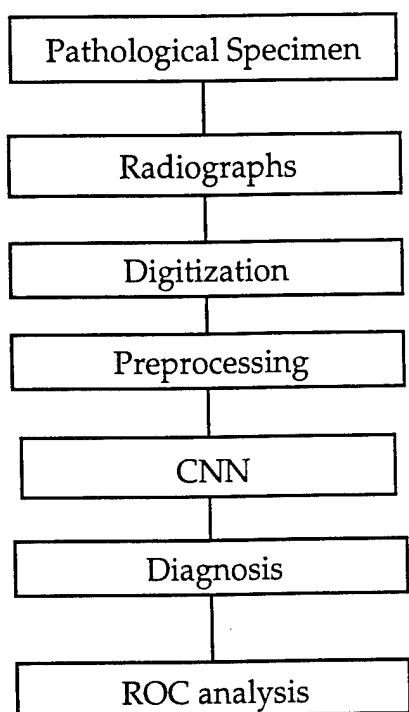


Figure 1. Overall approach for the classification of microcalcifications using CNN.



(a)



(b)

Figure 2. Microcalcifications shown in the pathological specimen (a) and shown after being digitized with a high resolution ($21\text{ }\mu\text{m} \times 21\text{ }\mu\text{m}$) digitizer (b).

2.2. Convolution Neural Network

The structure of the CNN is shown in Fig. 3 (ref. 44). The input to the CNN are ROIs of matrix size of 64×64 pixels, containing benign or malignant type of microcalcifications. Only one hidden layer is used in this study. The connections between input and hidden layer are grouped into seven different kernels based on the structure of Fukushima's neocognitron (ref. 45). There are two output units in the output layer, with each unit corresponding to benign or malignant class of microcalcifications. The hidden layer and the output layers are fully connected. Backpropagation and generalized delta rule are used in the training process (ref. 46).

In the training process of the CNN, each image block were rotated and reflected such that the number of training data will be increased eight fold. The rotation and reflection would represent different orientations of microcalcifications in mammograms. The training with additional orientation can effectively make the CNN rotational invariant.

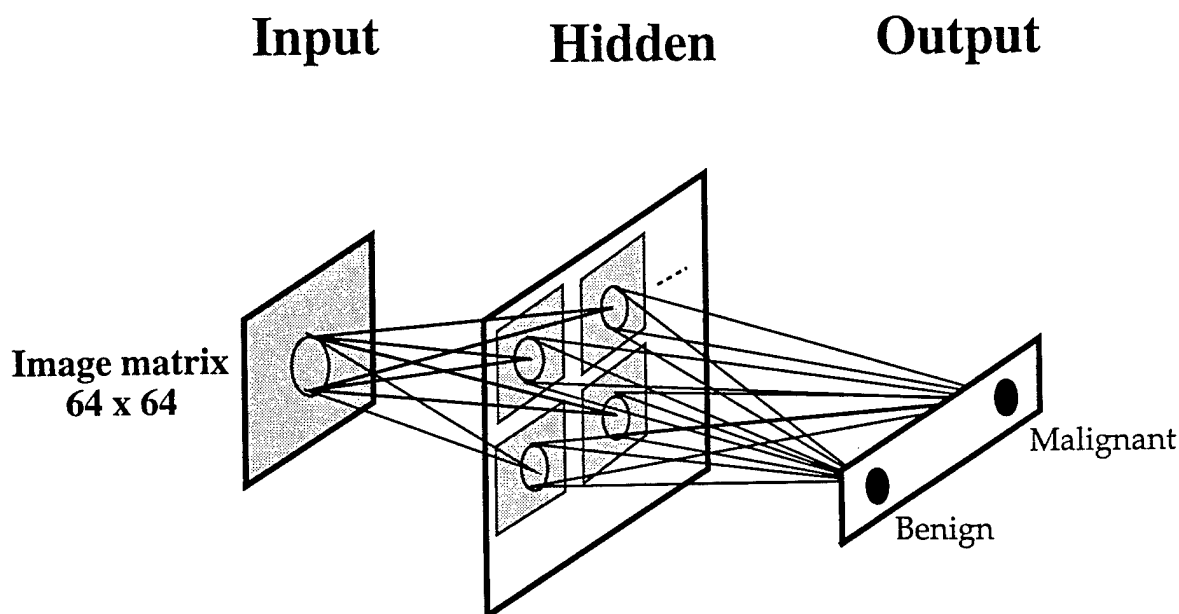
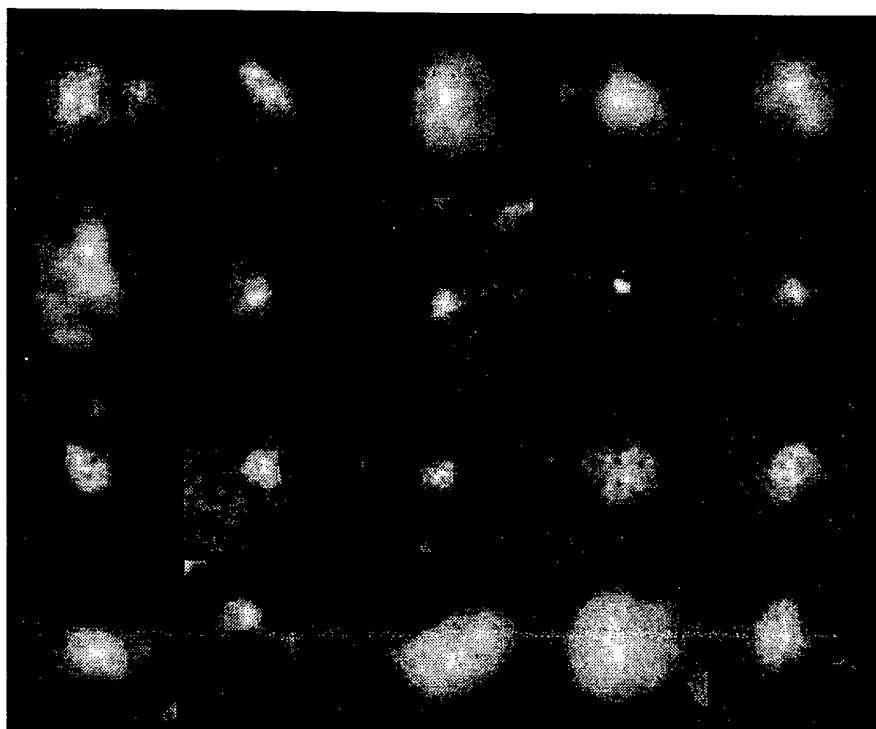
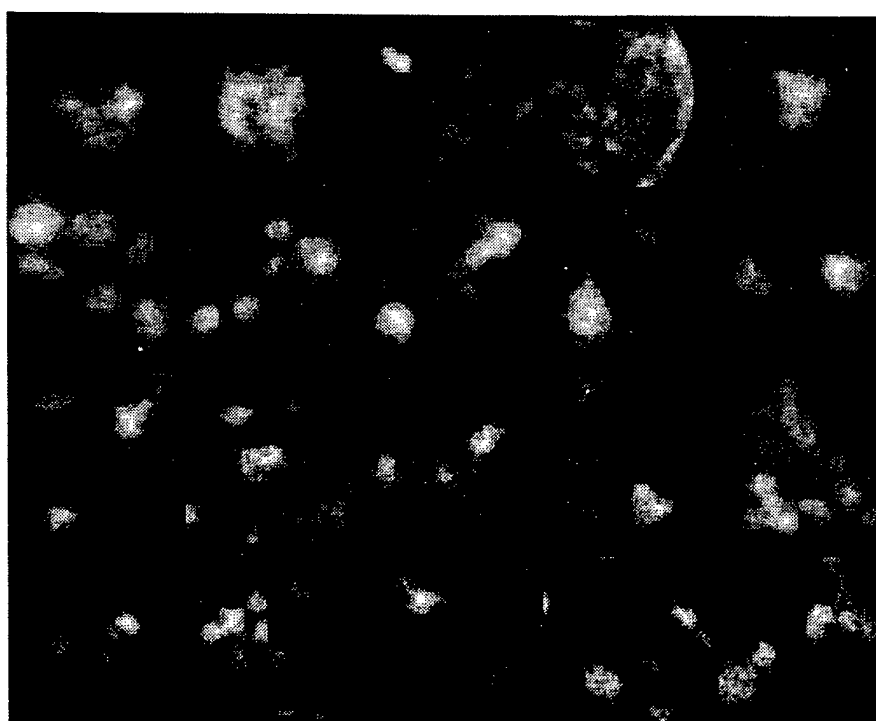


Figure 3. Structure of a convolution neural network.

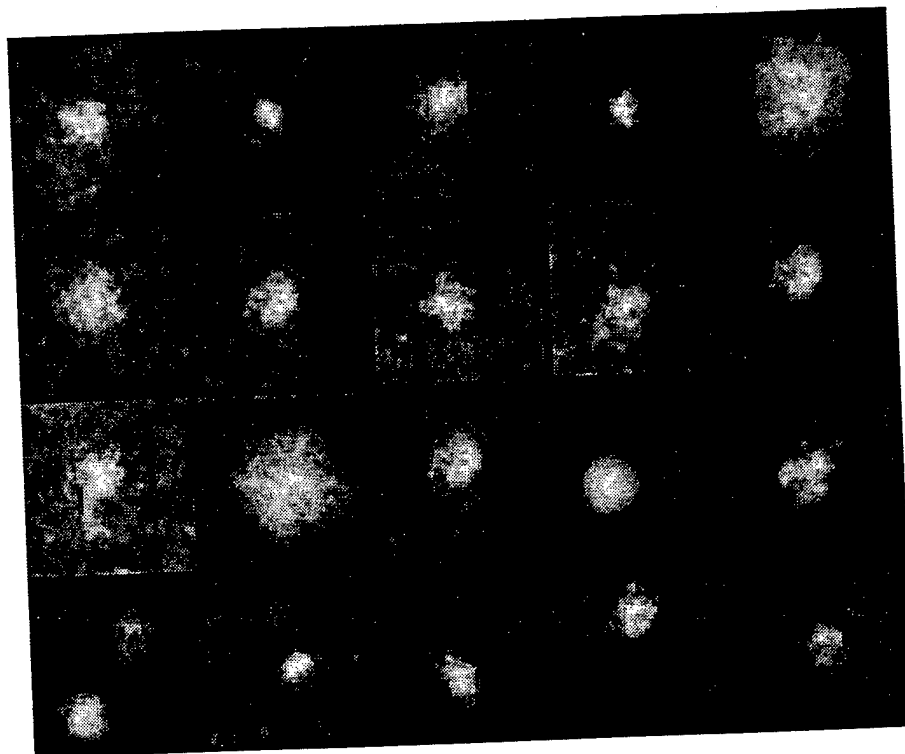


(a)

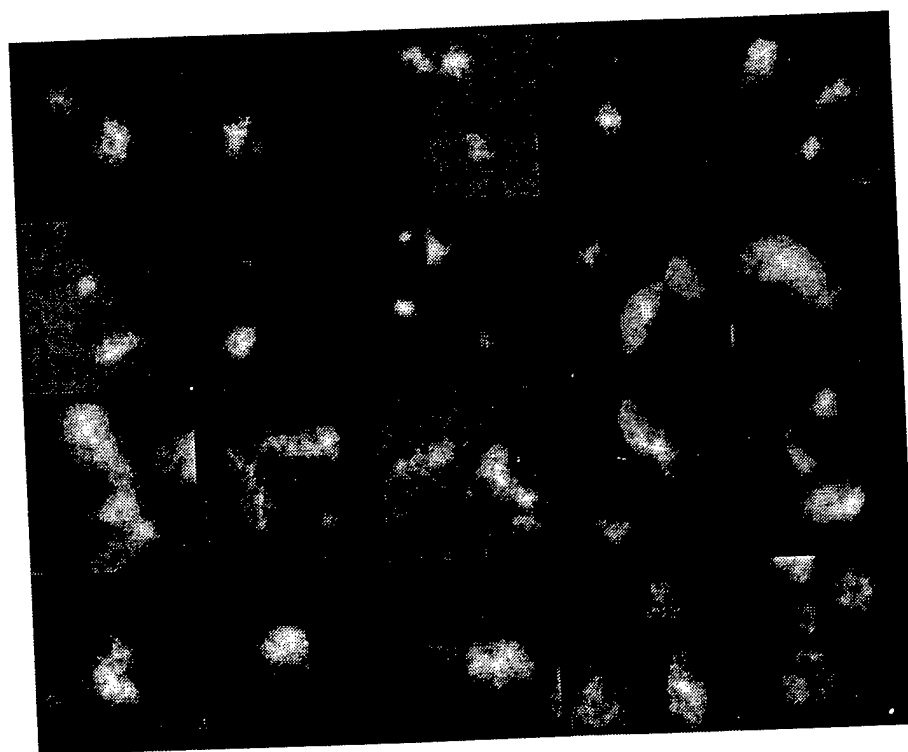


(b)

Figure 4. Training database of the convolution neural network that contains (a) 20 benign microcalcifications and (b) 20 malignant microcalcifications.



(a)



(b)

Figure 5. Testing database of the convolution neural network that contains (a) 20 benign microcalcifications and (b) 20 malignant microcalcifications.

3. RESULTS

3.1. Database

Forty ROIs (20 benign and 20 malignant) were randomly selected to train the CNN. Figure 4 shows the training database of the 40 ROIs. There are substantial variations in sizes among benign or malignant microcalcifications. Another 40 ROIs, shown in Fig. 5, were selected to test the trained CNN.

3.2. Performance of CNN

The output values from the two output units were analyzed and shown in Table I. The CNN achieved a sensitivity of 75% at false-positive rate of 25% (specificity = 75%).

Table I. Output of The Neural Network

	Negative Output from CNN	Positive Output from CNN
Benign ROIs (20)	15	5
Malignant ROIs (20)	5	15

3.3. ROC Analysis

Receiver Operating Characteristic (ROC) analysis (ref. 47-48) was employed to examine the output values from the neural network. The LABROC4 algorithm (ref. 49) developed by Metz *et al.* was used to fit ROC curves to the continuous data from the output of CNN. The ROC curve obtained is shown in Fig. 6. The area under ROC curve (A_z) achieved by the CNN was 0.83.

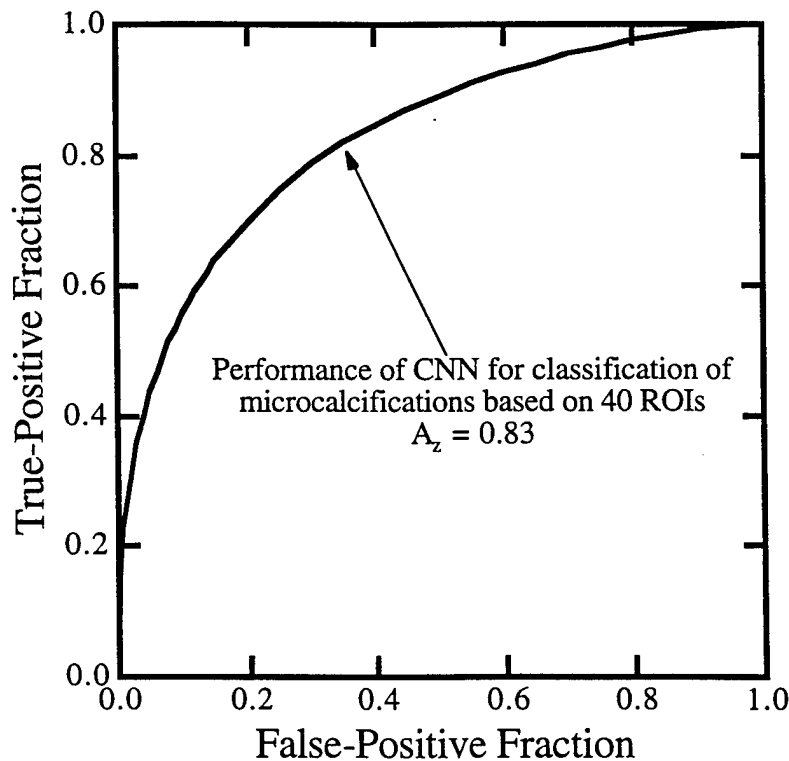


Figure 6. ROC analysis of the performance of CNN in classifying benign and malignant microcalcifications.

4. CONCLUSIONS

Even though they were based a relatively small training and testing database, they have been very promising, the results obtained in this study indicated the potential usefulness of CNN in classification of microcalcifications in mammograms digitized with high resolutions. Both a large training and testing database are necessary in order to train and evaluate the performance of the neural network sufficiently and reliably. We will be expanding our database significantly in the future. A hybrid neural network (HNN) (ref. 50) will be employed to classify microcalcifications based on the input of both image data and image features (ref. 51) that will be automatically extracted.

5. ACKNOWLEDGMENTS

This research was supported in part by an US Army Grant (DAMD17-93-J-3007). The content of the manuscript does not necessarily reflect the position or policy of the government. The authors wish to thank DBA Systems for the use of its digitizer.

6. REFERENCES

1. American Cancer Society, "Mammography 1982: A statement of the American Cancer Society," CA Cancer J. Clin. **32**, 226-230 (1982).
2. R. G. Lester, "The contribution of radiology to the diagnosis, management, and care of breast cancer," Radiology **151**, 1-7 (1984).
3. D. B. Kopans, J. E. Mayer, and N. Sadowsky, "Medical progress: Breast imaging," N. Engl. J. Med. **310**, 960-967 (1984).
4. American Cancer Society, "Mammography guidelines 1983: Background statement and update of cancer related checkup guidelines for breast cancer detection in asymptomatic women age 40 to 49," CA Cancer J. Clin. **33**, 255 (1983).
5. L. H. Baker, "Breast cancer detection demonstration project: Five-year summary report," CA Cancer J. Clin. **32**, 194-225 (1982).
6. NCRP Report No. 85, *Mammography — A User's Guide* (National Council on Radiation Protection and Measurement: Washington, D.C., 1986).
7. S. A. Feig, "Decreased cancer mortality through mammographic screening: Results of clinical trials," Radiology **167**, 659-665 (1988).
8. R. R. Millis, R. Davis, A. J. Stacey, "The detection and significance of calcifications in the breast: A radiological and pathological study," Br. J. Radiol. **49**, 12-26 (1976).
9. F. M. Hall, J. M. Storella, D. Z. Silverstone, G. Wyshak, "Nonpalpable breast lesions: Recommendations for biopsy based on suspicion of carcinoma at mammography," Radiology, **167**, 353-358 (1988).
10. L. W. Bassett, R. H. Gold, *Breast Cancer Detection: Mammography and Other Methods in Breast Imaging* (Grune & Stratton, New York, 1987).
11. C. J. Baines, A. B. Miller, C. Wall, et al, "Sensitivity and specificity of first screen mammography in the Canadian National Breast Screening Study: a preliminary report from five centers," Radiology **295**-298 (1986).

12. S. R. Pollei, F. A. Mettler, S. A. Bartow, G. Moradian, M. Moskowitz, "Occult breast cancer: prevalence and radiographic detectability," Radiology **163**, 459-462 (1987).
13. I. Anderson, "What can we learn from interval carcinomas?" Recent Results Cancer Res **90**, 161-163 (1984).
14. J. E. Martin, M. Moskowitz, J. R. Milbrath, "Breast cancer missed by mammography," AJR **132**, 737-739 (1979).
15. M. D. Vernon, *The Psychology of Perception* (Penguin, Middlesex, England, 1962).
16. W. J. Tuddenham, "Visual search, image organization and reader error in roentgen diagnosis," Radiology **78**, 694-704 (1962).
17. M. J. Smith, *Error and Variation in Diagnostic Radiology* (Charles C. Thomas Publisher, Springfield, IL, 1967).
18. J. L. Lehr, G. S. Lodwick, C. Farrel, M. O. Braaten, P. Virtama, E. L. Koivisto, "Direct measurement of the effect of film miniaturization on diagnostic accuracy," Radiology **118**, 257-263 (1976).
19. B. J. Hillman, L. L. Fajardo, T. B. Hunter, *et al.*, "Mammogram interpretation by physician assistants," AJR **149**, 907-911 (1987).
20. E. A. Sickles, "Mammographic detectability of breast microcalcifications," AJR **139**, 913-918 (1982).
21. J. N. Wolfe, "Analysis of 462 carcinomas," AJR **121**, 846-853 (1974).
22. W. A. Murphy, K. DeSchryver-Kecsckemeti, "Isolated clustered calcifications in the breast: Radiologic-pathologic correlation," Radiology **127**, 335-341 (1978).
23. E. R. Fisher, R. M. Gregorio, B. Fisher, C. Redmond, F. Vellios, S. C. Sommers, "The pathology of invasive breast cancer," Cancer **36**, 1-85 (1975).
24. J. W. Black, B. Young, "A radiological and pathological study of the incidence of calcifications in diseases of the breast and neoplasms of other tissue," Br. J. Radiol. **38**, 596-598 (1965).
25. M. Lanyi, *Diagnosis and Differential Diagnosis of Breast Calcifications* (Springer-Verlag Berlin Heidelberg, 1988).
26. P. Stomper, J. L. Connolly, J. E. Meyer, J. R. Harris, "Clinically occult ductal carcinoma in situ detected with mammography," Radiology **172**, 235-241 (1989).
27. M. J. Homer, A. G. Cooper, E. R. Pile-Spellman, "Milk of calcium in breast microcysts: manifestation as a solitary focal disease," Am. J. Radiol **150**, 78-9 (1988).
28. F. Dido, J. Crowe, R. Zollinger, *et al.*, "Biopsy of the breast for mammographically detected lesions," Surgery Gynecology and Obstetrics **171**, 449-455 (1990).
29. L. H. Leviathan, D. M. Within, E. G. Harrison, "Calcification in breast disease: mammographic-pathologic correlation," AJR **2**, 29-39 (1964).
30. R. L. Egan, M. B. McSweeney, C. W. Sewell, "Intramammary calcifications without an associated mass in benign and malignant diseases," Radiology **137**, 1-7 (1980).

31. D. F. Roses, J. Mitnick, M. H. Harris, *et al.*, "The risk of carcinoma in wire localization biopsies of mammographically detected clustered microcalcifications," Surgery **110**, 877-86 (1991).
32. I. M. Freundlich, T. B. Hunter, G. W. Seeley, *et al.*, "Computer assisted analysis of mammograph clustered microcalcifications," Clinical Radiology **40**, 295-298 (1989).
33. H.-P. Chan, K. Doi, S. Galhotra, C. J. Vyborny, H. MacMahon, P. M. Jokich, "Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography," Med Phys **14**, 538 (1987).
34. H.-P. Chan, K. Doi, C. J. Vyborny, K. L. Lam, R. A. Schmidt, "Computer-aided detection of microcalcifications in mammograms: Methodology and preliminary clinical study," Invest Radiol **23**, 664 (1988).
35. B. W. Fam, S. L. Olson, P. F. Winter, F. J. Scholz, "Algorithm for the detection of fine clustered microcalcifications on film mammograms," Radiology **169**, 333 (1988).
36. D. H. Davies, D. R. Dance, "Automatic computer detection of clustered calcifications in digital mammograms," Phys Med Biol **35**, 1111 (1990).
37. H.-P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K.-L. Lam, T. Ogura, Y. Wu, H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," Invest Radiol **25**, 1102-1110 (1990).
38. Y. Wu, K. Doi, M. L. Giger, R. M. Nishikawa, "Computerized detection of clustered microcalcifications in digital mammograms: Applications of artificial neural networks," Med Phys **19**(3), 555-560 (1992).
39. R. M. Nishikawa, Giger, M. L., Doi, K., Vyborny, C. J., and Schmidt, R. A., "Computer-aided detection of clustered microcalcifications using digital mammograms", Med. Bio. Eng. Com. (accepted for publication).
40. S. Astley, C. Taylor, C. Boggis, M. Wilson, T. Ellison, "Automated detection of microcalcifications in mammograms," Radiology **177**(P), 288 (1990).
41. M. Grimaud, S. Muller, F. Meyer, "Automated detection of microcalcifications in mammograms," Radiology **177**(P), 288 (1990).
42. N. Darssemeijer, "A stochastic method for automated detection of microcalcifications in digital mammograms. In: A. C. F. Colchester, D. J. Hawkes, *Information processing in medical imaging*, Springer-Verlag, New York, 227-238, 1991.
43. R. Hecht-Nielsen, *Neurocomputing* (Addison-Wesley Publishing Company, Reading, MA, 1990).
44. S.-C. B. Lo, J.-S. Lin, M. Freedman, S. K. Mun, "Computer-assisted diagnosis of lung nodule detection using artificial convolution neural network," Medical Imaging VII **1898**, SPIE Medical Imaging 1993 Symposium, Newport Beach, CA, 14-18 February, 1993.
45. K. Fukushima, S. Miyake, "Neocognitron: A New Algorithm for Pattern Recognition Tolerant of Deformations and Shift in Position," Pattern Recognition **15**(6), 455-469 (1982).
46. D. E. Rumelhart, J. L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, 1986.

47. C. E. Metz, "ROC methodology in radiologic imaging," Invest Radiol **21**, 720-733 (1986).
48. C. E. Metz, "Basic principles of ROC analysis," Seminars in Nuclear Medicine **VIII**(4), 283-298 (1978).
49. C. E. Metz J.-H. Shen, B. A. Herman, "New methods for estimating a binormal ROC curve from continuously-distributed test results," Presented at the 1990 Joint Meetings of the American Statistical Society and the Biometric Society: Anaheim, CA, August 1990.
50. Y. C. Wu, S.-C. B. Lo, M. T. Freedman, S. K. Mun, "Automatic detection of lung nodules with used of parametric and nonparametric neural networks," Presented at the 79th scientific assembly and annual meeting of RSNA, Radiology, 189 (p), 272 (1992).
51. Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, "Application of neural networks in mammography for the diagnosis of breast cancer," Radiology **187**, 81-87 (1993).



Classification of Microcalcifications in Digital Mammograms for the Diagnosis of Breast Cancer

Osamu Tsujii^{1,2}, Akira Hasegawa², Chris Y. Wu², Shih-Chung B. Lo²,
Matthew T. Freedman², Seong K. Mun²

¹ Medical Equipment Business Group, CANON INC., Kawasaki-shi, Japan 211

² Department of Radiology, Georgetown University Medical Center, Washington, DC 20007

ABSTRACT

We are developing a computer program for automated classification of clustered microcalcifications associated with benign and malignant processes in mammograms. Accurate classification of microcalcifications into benign and malignant groups would help improve diagnosis sensitivity as well as reduce the number of unnecessary biopsies. In this study we investigate the effectiveness of several image features and the proposed neural network, the Trend-Oriented Radial Basis Function (TRBF). Our database is composed of 47 benign and 81 malignant region of interest (ROI) images, a total of 128 ROIs, which are selected from 50 micron x 50 micron digitized whole mammograms manually. Each 256 x 256 pixel ROI image contains clustered microcalcifications. First we extract 16 image features which are calculated from a binarized microcalcifications image and its two morphological dilation images. These features are based on 3 morphological criteria: (1) size and/or shape of the calcifications, (2) size or shape of the "cluster", and (3) number of microcalcifications. Secondly, we apply Karhunen-Loeve(K-L) expansion to 16 dimension feature space in order to reduce the dimension of the problem. Nextly we select two dimensional K-L features, which are outcomes of K-L expansion, through a calculation of the Euclidean distance measure. Finally we classify them based on two dimensional K-L features using the proposed neural network. The proposed TRBF neural network has three layers and a new learning algorithm. Its cost function for the learning process consists of a sum of squares error and a sum of inverse RBFs width. Through the learning process, the center and the width of each RBF and the weights between hidden and output layers are decided. The performances of the K-L features and the TRBF neural network are evaluated through the Round-Robin method in which one sample is tested after learning based on the rest of the 127 samples. The two dimensional K-L features are more distinguishable than the raw two dimensional feature combination. The proposed TRBF network is able to define the trend of the distribution than former RBF networks. According to the receiver operating characteristic (ROC) curve, this system indicates a better performance than one trained radiologist.

Keywords: Mammograms, Microcalcification, Classification, Feature selection, Karhunen-Loeve expansion, Euclidean distance measure, Neural Network, Radial Basis Function, Round-Robin method, Receiver Operating Characteristic

1. INTRODUCTION

In the United States, breast cancer is the leading cause of death in women between 40 and 55 years of age. At present the mammogram is the only proven method for detecting minimal breast cancer [1]. One important indicator of breast cancer is the presence of clustered microcalcification [2]. Clustered microcalcifications can be seen on mammograms in 30%-50% cases of breast cancer [3]. However most mammographic calcifications are benign. Accurate classification of microcalcifications into benign and malignant groups would help improve diagnosis sensitivity as well as reduce the number of unnecessary biopsies. Wu et al. applied a convolution neural network to classify microcalcifications in radiographs of pathologic specimens[4]. The neural network achieved an Az value of 0.90.

What features are useful to distinguish benign from malignant calcifications? Various investigators have attempted this distinction. Roselli-Del-Turco et al. used three features to distinguish benign from malignant calcifications[5]: (1) size, shape or density of the calcifications, (2) size or shape of the "cluster", and (3) number of microcalcifications. Their analysis of the morphological criteria, which led to a distinction between benign and malignant biopsy results, is one of the most thorough that has been published to date (Table 1). However it does not seem that there are conclusive features which could distinguish benign from malignant calcifications very well. Figures 1 and 2 are example images of mammograms of benign and malignant calcifications. According to the spatial disposition result of Roselli-Del-Turco, both images seem to be malignant. In fact, the Figure 1 image is a benign case.

Most classification systems consist of four subsystems: measure, preprocess and/or transformation, feature selection, and classification. In the design of a classification system, a common assumption is that all input features play an important discrimina

tory role in the classification and are essential for a specified performance. However this may not always be true in practical application. If the designer does not have confidence in what the effective features are, some features may be redundant or not as important as others. Some features also might be noise-corrupted.

Our purpose in this paper is to investigate the effectiveness of our feature selection method and the proposed neural network as detailed below:

(1) Image feature selection method for classification of clustered microcalcifications

It is not easy to estimate appropriate image features on this classification problem. First we propose some image feature candidates. Secondly, some redundancy of image features is eliminated through K-L expansion. Finally we determine some eligible image feature combinations through the Euclidean distance measure of normalized K-L features.

(2) Trend-Oriented Radial Basis Function neural network

In this case, a more powerful neural network to regularize the class distribution is desired. We propose applying the original cost function which would be minimized at the network training. We present the learning equations of the centers and widths of each RBF based on the Gradient-descent method.

In Section 2, we propose 16 image features which might be redundant, and two dimensional K-L feature combination is determined through our proposed selection method. In Section 3, we apply the Trend-Oriented Radial Basis Function to classify the data patterns with two dimensional feature. The performance is evaluated through the ROC analysis.

Table 1 Distribution of mammographically evidenced microcalcifications by morphological criteria in 217 cancers and in 111 benign lesions: positive predictive value (PPV) of each radiologic criterion

	CANCER 1,110 217		BENIGN 2,016 111		PPV (%)
Total evaluated cases					
Cases with macrocalcifications					66.2
Mammographic morphological criteria	N	%	N	%	
Spatial disposition					
Isolated	40	18.4	42	37.8	48.8
Clustered	146	67.3	69	69.2	67.9
Widely diffused	31	14.3	0	-	100.0
Total number					
1 to 4	145	66.8	81	72.9	64.2
5 to 10	22	10.2	25	22.9	46.8
Over 10	50	23.0	5	4.5	90.9
Number/cm2					
1 to 10	34	15.7	27	24.3	55.7
11 to 20	44	20.3	27	24.3	62.0
21 to 50	87	40.0	48	43.2	64.4
Over 50	52	24.0	9	8.2	85.2
Morphology					
Dotlike	151	69.6	93	83.8	61.9
Sticklike	33	15.2	16	14.4	67.3
Ramified	33	15.2	2	1.8	94.3
Shape					
Regular	25	11.5	62	55.8	28.7
Irregular	192	88.5	49	44.2	79.7
Radiologic density					
Low	103	47.5	60	54.0	63.2
High	114	52.5	51	46.0	69.1
Mammographic capacity					
Absent	61	28.1	67	60.4	47.7
Next to Ca++*	33	15.2	4	3.6	89.2
Ca++ inside opacity	121	55.8	36	32.4	77.1
Ca++ at border	2	0.9	4	3.6	33.3
Maximum diameter					
0.1-0.3 mm	47	21.7	27	24.4	63.5
0.6-0.9 mm	64	29.5	42	37.8	60.4
Over 0.9 mm	106	48.8	42	37.8	71.6
Average diameter					
0.1-0.3 mm	48	22.1	27	24.3	64.0
0.4-0.5 mm	88	40.5	56	50.4	61.1
0.6-0.9 mm	51	23.5	9	8.1	85.0
Over 1.0 mm	30	13.8	18	16.2	62.5

Ca++* means microcalcifications.

From Roselli-Del-Turco et al.: The significance of mammographic calcification in early breast cancer detection, Radiol Med 72:7-12, 1986.



Figure 1 Benign mammogram.



Figure 2 Malignant mammogram.

2. METHOD

2.1. Preprocessing

Figure 3 shows the overview of the entire method. The purpose of the preprocessing is to get a binarized image of clustered microcalcifications. Our image features are fully based on the size, shape, disposition and the number of microcalcifications, not on the density. The density will be changed because of the overlap. Figures 4 and 5 are an original gray scale image and a binarized microcalcifications image, respectively. Our preprocessing algorithm is regarded as below.

(1) Subtraction of the averaged image from the original image

The averaged image is made with 23 x 23 pixel size kernel and then is subtracted from the original image to eliminate the background trend.

(2) Binarization through histogram quantization

The average subtracted image, the outcome of (1), is quantized to 32 levels based on its histogram. The maximum level is only used as a candidate of a clustered microcalcification.

(3) Opening to remove the line artifact

The 3x3 kernel morphological opening filter is applied to the histogram quantized image (3) to remove line artifacts which arise when microcalcifications are in the ductal structure.

(4) Dilation (3)

The 3x3 kernel size morphological dilation filter is applied to the outcome of process (3) to enlarge objects.

(5) And (2) with (4)

The AND operation between the outcome (2) and the outcome (4) is done to preserve the shape of microcalcifications.

(6) Labeling to remove small microcalcifications

The labeling process with the outcome (5) is done to eliminate small objects whose sizes are less than 5 pixels. In this process all microcalcifications in which pixel size is less than five should be eliminated.

2.2. Feature extraction

In this study our image features are based on three images; one, which is a binarized clustered microcalcifications image (Figure 5), is the outcome of preprocessing, the other two are morphological dilation images which come from a binarized clustered microcalcifications image (Figure 7 and 8). These morphological kernel sizes are 5x5 and 25x25 which are estimated from the actual images. Below is a list of 16 features which are called raw features.

(01) number of microcalcifications N

(02) area (pixel count) of microcalcifications in Image1 [S1]

(03) area (pixel count) of microcalcifications in Image2 [S2]

(04) area (pixel count) of microcalcifications in Image3 [S3]

(05) $(S2-S1)/N$ which is intended to be an average irregular pixel count

(06) $(S3-S1)/N$ which is intended to be one of the disposition features. When microcalcifications are close together, this feature should be small.

(07) $(S2-S1)$ which is a total pixel count of the irregularity

(08) $(S3-S1)$ which is intended to be one of the disposition features

(09) (biggest microcalcification pixel count) - (average microcalcification pixel count) on Image1

(10) (average microcalcification pixel count) - (smallest microcalcification pixel count) on Image1

(11) biggest microcalcification pixel count on Image1

(12) ellipsity of biggest microcalcification on Image1

(13) weighted average ellipsity of microcalcifications $\sum (S_n \times \text{ellipsity}_n) / S1$

(14) length of semimajor axes of the ellipse which best fits to the disposition of microcalcifications on Image1 (Figure 6)

(15) ax / bx where bx is the length of semiminor axes of the best-fit ellipse which fits to the disposition of microcalcifications on Image1 (Figure 6)

ROI Images Database consist of 47 benign and 81 malignant.

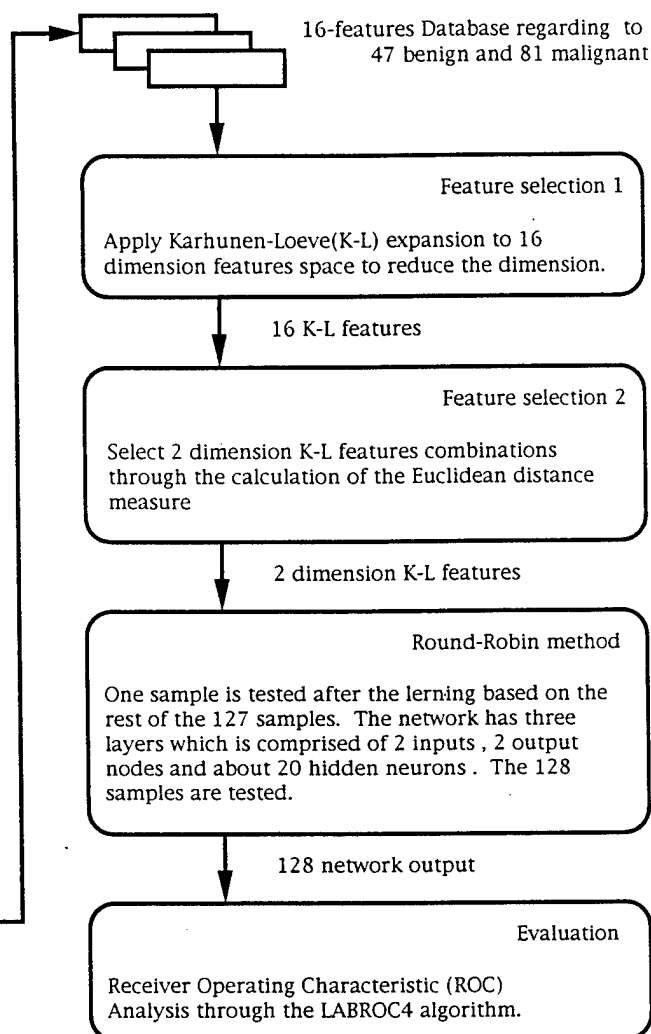
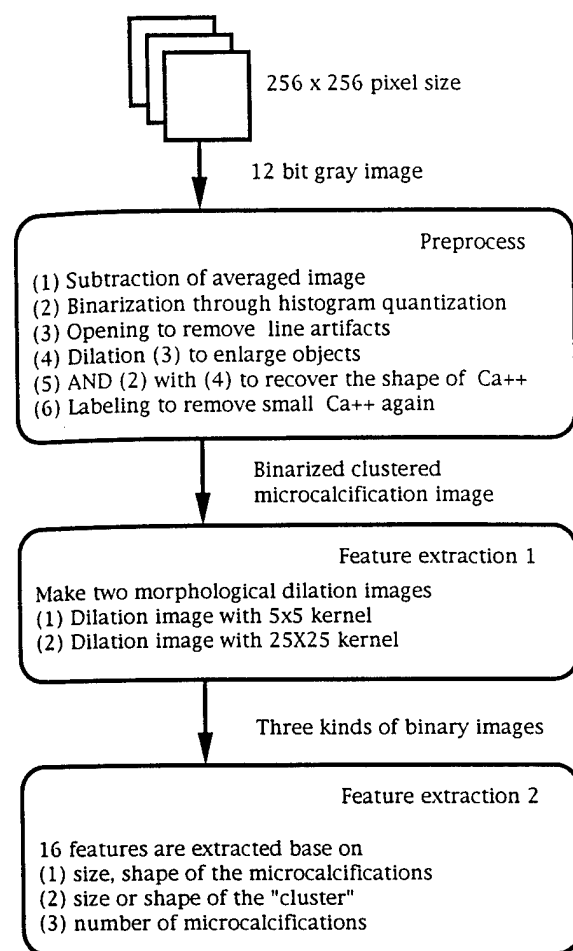


Figure 3 The overview of the method.



Figure 4 Original gray scale image.



Figure 5 Binarized microcalcifications image.

(16) $a \times b$

where Image1 is the binarized clustered microcalcification image, Image2 is the 5x5 kernel dilated image, and Image3 is the 25x25 kernel dilated image. We know that these image features may be redundant or not as important as others, and such a redundancy should be removed in the next feature selection process.

2.3. Feature selection

The reduction of the dimensionality of a program in order to deliver a system for a real-world application has always been a main concern of researcher. In this study there is not a large enough database and we are not sure which features are most effective. There are two basic approaches to reducing dimensionality. One approach is to use a transformation technique such as the Fourier transformation or the K-L expansion. Another approach is to select a subset of features by evaluating the features based on the available data.

Features can be evaluated in combination or individually. A combinational feature evaluation measures the separability of a subset of features which takes into account the combining effect of a set of features. Two distance measures, probabilistic distance and Euclidean distance, are commonly used to evaluate the separability in the feature space. The Euclidean distance measure of 2 features is defined as

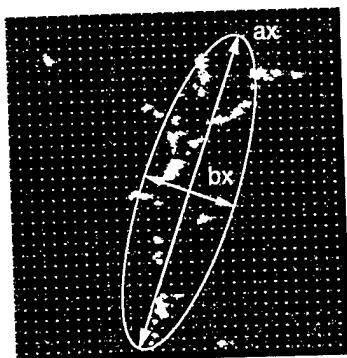


Figure 6 Best-fit ellipse and its semimajor axes ax and semimajor axes bx.

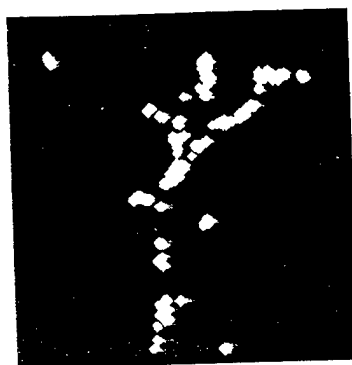


Figure 7 Dilation with 5x5 kernel.

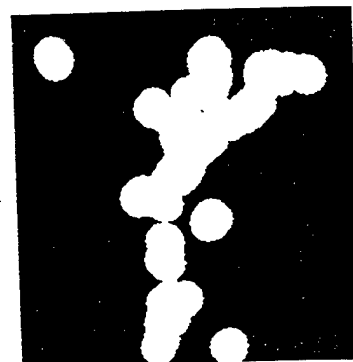


Figure 8 Dilation with 25x25 kernel.

$$EDM(F_1, F_2) = p(\omega_k) \times p(\omega_l) \times \frac{1}{N(\omega_k) \times N(\omega_l)} \times \sum_{p=1}^{N(\omega_k)} \sum_{q=1}^{N(\omega_l)} [x_p^{(k)} - x_q^{(l)}]^T [x_p^{(k)} - x_q^{(l)}] \quad (1)$$

where F_1, F_2 are features, $p(\omega_k)$ and $p(\omega_l)$ are prior probability of occurrence of class ω_k and ω_l respectively, $N(\omega_k)$ is the number of patterns in class ω_k , and $x_p^{(k)}$ are two element vectors from class ω_k . This equation 1 accumulates the distance between every pair of two-element patterns that is from a different class.

Three reasons we decide the dimensionality had better be two:

- (1) The database is not large enough to solve many-dimensional problem which introduce more degrees of freedom.
- (2) For the two dimensional problem it is easier to estimate the probability of solving the problem through the data distribution chart.
- (3) With the two dimensional problem, it is easier to evaluate the performance of the proposed neural network.

We use the Euclidean distance measure to evaluate the separability of two dimensional feature combinations. The number of combinations is 124. Figure 9 shows the Euclidean distance measure of 124 two dimensional combinations from our 16 raw features. The highest separability combination which indicates a maximum Euclidean distance measure is the combination between the raw feature 02 [S1] and 14 [ax], in which S1 is the area of microcalcifications in the binarized clustered microcalcification image, ax is the length of semimajor axes of the best-fit ellipse which fits to the disposition of microcalcifications in a binarized microcalcification image. But there are several peaks of the Euclidean distance measure in Figure 9, so that it is redundant to select one two dimensional combination. Then we apply K-L expansion to the covariance matrix of this database to eliminate the redundancy of raw image features. Table 2 shows eigen values and eigen vectors of the distribution of this database. The K-L01 feature in which the eigen value is maximum indicates the maximum covariance direction of this database. The K-L01 feature mainly consists of raw image features 01, 02, 03, 04, 07, 08, 13 and 16. Figure 10 shows the Euclidean distance measure of 124 two dimensional combinations from 16 K-L features. The most separable combination which indicates a maximum Euclidean distance measure is the combination between K-L01 and K-L02. However the result of Figure 10 does not precisely mean that the most separable combination is this combination between K-L01 and K-L02, because the eigen values of these two feature vectors are two of the top value vectors which have the maximum covariance direction. Therefore we normalized each K-L feature to be [0..1] in its own direction. Figure 11 shows the Euclidean distance measure of 124 two dimensional combinations from 16 normalized K-L features. The two of the most separable combinations are the combination between the normalized K-L01 and the normalized K-L05 and the combination between the normalized K-L01 and the normalized K-L04. These two Euclidean distance measures are much bigger than the other combinations so that there is not much redundancy in selecting the eligible combination at this time. Figure 12, 13 and 14 show the distributions of feature combinations between the raw features 01 and 14, the normalized K-L05, and the normalized K-L 01 and the normalized K-L04, respectively. In Figures 12~14, the diamonds and the circles denote a benign pattern and a malignant pattern, respectively. These results show that the Euclidean distance measure of normalized K-L features are a more appropriate indicator of feature selection than the Euclidean distance measure of raw features of K-L features.

Table 2 Eigen Values and Eigen Vectors of the Distribution of this Database

Eigen Values =

(0.00, -0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.01, 0.01, 0.02, 0.03, 0.04, 0.06, 0.12, 0.33)
K-L16 K-L15 K-L14 K-L13 K-L12 K-L11 K-L10 K-L09 K-L08 K-L07 K-L06 K-L05 K-L04 K-L03 K-L02 K-L01

Eigen Vectors =

Raw Feature																
01	0.00	-0.00	0.00	-0.04	-0.55	0.44	0.41	0.39	0.07	-0.03	0.02	-0.10	0.06	-0.09	-0.17	0.33
02	0.18	-0.31	0.04	0.21	-0.07	0.45	-0.36	-0.45	-0.07	0.03	0.00	-0.19	0.16	-0.08	0.26	0.33
03	-0.45	0.64	-0.19	-0.10	-0.13	-0.15	-0.11	-0.24	-0.06	-0.00	0.05	-0.16	0.14	-0.10	0.12	0.35
04	0.28	0.35	0.56	0.16	0.42	0.02	0.03	0.13	0.21	-0.18	0.10	0.05	0.06	-0.00	-0.05	0.38
05	0.00	-0.00	0.00	-0.46	0.31	0.25	0.18	0.21	-0.35	0.04	0.32	0.01	0.05	-0.07	0.54	-0.04
06	-0.00	0.00	0.00	-0.12	-0.30	0.01	-0.02	-0.18	0.48	-0.33	0.28	0.46	-0.11	0.27	0.35	-0.04
07	0.24	-0.35	0.10	-0.35	-0.18	-0.64	0.08	-0.06	-0.05	-0.04	0.09	-0.13	0.13	-0.11	0.01	0.37
08	-0.26	-0.33	-0.53	0.15	0.45	-0.01	0.06	0.17	0.23	-0.19	0.11	0.07	0.05	0.00	-0.08	0.37
09	-0.47	-0.21	0.37	-0.15	0.05	0.01	0.03	0.05	0.20	-0.08	-0.59	-0.05	-0.20	-0.15	0.29	0.07
10	-0.09	-0.04	0.07	0.68	-0.18	-0.30	0.13	0.27	-0.12	0.10	0.15	-0.03	0.02	-0.02	0.49	-0.02
11	0.56	0.25	-0.44	-0.01	0.01	-0.03	0.05	0.09	0.15	-0.05	-0.47	-0.05	-0.17	-0.13	0.32	0.05
12	-0.00	0.00	-0.00	0.02	-0.01	-0.00	0.04	0.06	-0.36	-0.50	-0.26	-0.16	0.21	0.67	0.03	0.03
13	0.00	0.00	0.00	-0.11	0.07	-0.00	0.04	0.06	0.42	0.61	-0.03	-0.26	0.26	0.51	0.12	0.02
14	0.00	-0.00	-0.00	0.06	0.06	0.01	0.38	-0.33	-0.19	0.19	0.07	-0.03	-0.70	0.28	-0.03	0.27
15	-0.00	0.00	0.00	-0.06	-0.03	0.00	-0.34	0.23	0.19	-0.24	0.30	-0.66	-0.41	0.02	0.00	-0.12
16	-0.00	0.00	0.00	-0.09	-0.12	-0.01	-0.58	0.42	-0.22	0.24	-0.06	0.37	-0.23	0.18	-0.03	0.32
	K-L16	K-L15	K-L14	K-L13	K-L12	K-L11	K-L10	K-L09	K-L08	K-L07	K-L06	K-L05	K-L04	K-L03	K-L02	K-L01

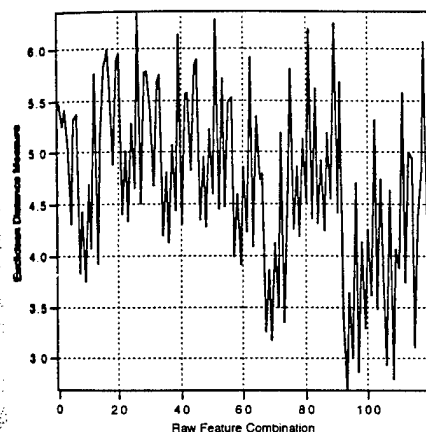


Figure 9 The Euclidean distance measure of 124 two dimensional combinations from our 16 raw features.

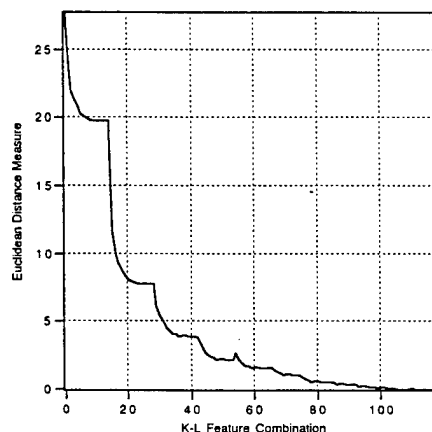


Figure 10 The Euclidean distance measure of 124 two dimensional combinations from 16 K-L features.

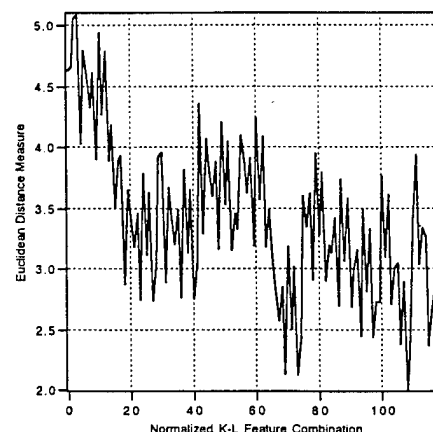


Figure 11 The Euclidean distance measure of 124 two dimensional combinations from 16 normalized K-L features.

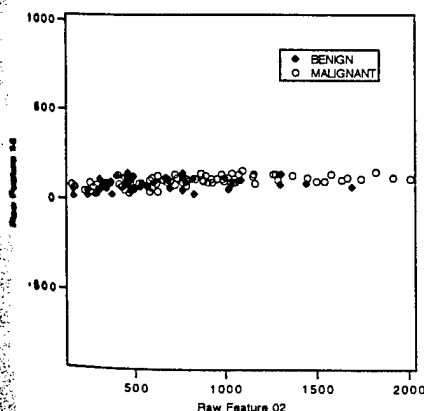


Figure 12 The distributions of raw feature 02 and 14.

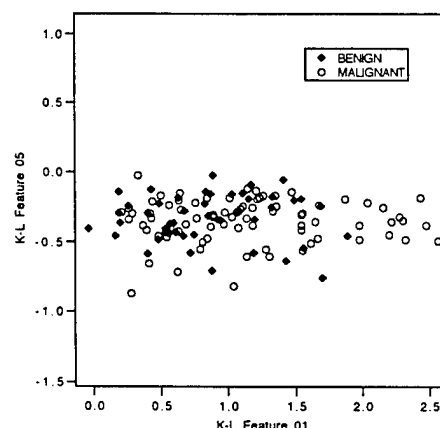


Figure 13 The distributions of K-L Feature 01 and 05.

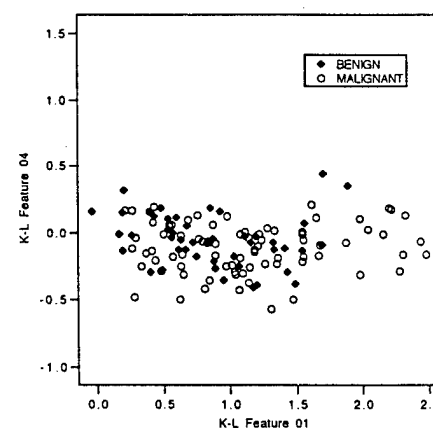


Figure 14 The distributions of K-L Feature 01 and 04.

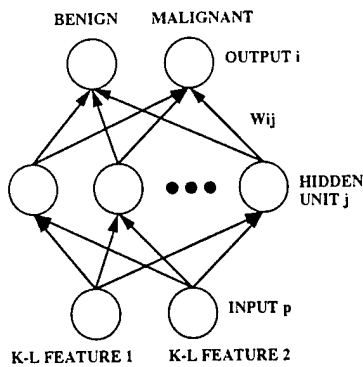


Figure 15 Architecture of a Radial Basis Function network.

2.4. Neural network

2.4.1. Background of RBF neural network and diagram

Here we review briefly the central features of the Radial Basis Function network. For a more extensive discussion see Broomhead and Lowe (1988)[6]. The network has a three layer feedforward architecture as shown in Figure 15. Input vectors x are propagated to the hidden unit (hidden unit) each of which computes a hyperspherical function of x , so that the output of the j th hidden unit is given by

$$\phi_j = \phi(\|x - y_j\|) \quad (2)$$

where y_j is the center of the Radial Basis Function for unit j , and $\|\cdot\|$ denotes a distance measure that is generally taken to be the Euclidean norm. The nonlinear function ϕ can be chosen in variety of ways and can in principle vary from one hidden unit to the next. For example, we have taken a Gaussian nonlinearity.

$$\phi(x) = \exp\left(-x^2/\sigma^2\right) \quad (3)$$

The outputs of the network are formed from the weighted sum of the outputs from the hidden units:

$$Z_i = \sum_j \omega_{ij} \phi_j \quad (4)$$

where the synoptic weights ω_{ij} are adaptive variables that are set during the learning phase. Training data are supplied to the network in the form of pairs (x_p, t_p) of input and target vectors, where $p = 1 \dots P$ labels the individual training pairs. The learning algorithm aims to minimize the sum-of-squares error defined by

$$E = \frac{1}{2} \sum_p \sum_i (Z_{ip} - t_{ip})^2 \quad (5)$$

where $Z_{ip} = Z_i(x_p)$ denotes the output of unit i when network is presented with input vector x_p . At a minimum of E we have

$$\frac{\partial E}{\partial \omega_{ij}} = 0 \quad (6)$$

Unlike the widely used technique of error backpropagation (Rumelhart and McClelland 1986)[7] the learning algorithm for RBF networks corresponds to the solution of linear problem. Therefore the training of the network is a fast procedure as detailed below.

$$\omega_{ij} = \sum_k (M^{-1})_{kj} \left\{ \sum_p \phi_{kp} t_{jp} \right\} \quad (7)$$

where the matrix M , which is the covariance matrix of the transformed data, is defined by

$$M_{kj} = \sum_p \phi_{kp} \phi_{jp} \quad (8)$$

where $\phi_{kp} = \phi_k(x_p)$ and $\phi_{jp} = \phi_j(x_p)$.

2.4.2. Former RBF neural networks

An important consideration in setting up an RBF network is the choice of the number, center y_j and width σ_j of the Radial Basis Function (i.e., the hidden unit). The most natural choice is to let each data sample point in the training set correspond to a Radial Basis Function center. In this case the number of degrees of freedom in the network equals the number of items of data, and the network function fits exactly through each data point. If the data appear regular, but are contaminated by noise, the network will learn all the details of the individual data points, rather than representing the underlying trends in the data. This phenomenon is sometimes called overfitting. There are three main ways to avoid overfitting. The first, regularization (Tikhonov and Arsenin 1977; Bishop 1991)[8], reduces the "number of parameter measurements" (MacKay 1992) in a large model (e.g., the full model) by adding a weight penalty term to the minimization criterion. For example, minimization of the energy E_S

$$E_S = \frac{1}{2} \sum_p \sum_i (Z_{ip} - t_{ip})^2 + \lambda \sum_j \omega_{ij}^2 \quad (9)$$

is zero-order regularization. The regularization parameter λ has to be chosen a priori or estimated from the data. The second way to avoid overfitting is to explicitly limit the complexity of the network by allowing only a subset of the possible centers to partici

pate. This method has the added advantage of producing parsimonious networks. Broomhead and Lowe (1988) suggested choosing such a subset randomly from the training inputs. Chen et al.(1991)[9] used forward selection to choose the centers of the hidden units to produce parsimonious networks. The third way to avoid overfitting is to find the approximated solution of centers and σ s after limiting the complexity of the network. At a minimum of E_s we have

$$\frac{\partial E_s}{\partial Y_j} = 0 \quad (10)$$

$$\frac{\partial E_s}{\partial \sigma_j} = 0 \quad (11)$$

Gradient-descent is probably the simplest approach for attempting to find the solution to this problem, though, of course, it is not guaranteed to converge (Tomaso Poggio and Federico Girosi 1990). In the gradient-descent method the values of y_j and σ_j that minimize E_s are regarded as below:

$$\Delta Y_j = -\beta \frac{\partial E_s}{\partial Y_j} \quad (12)$$

$$\Delta \sigma_j = -\beta \frac{\partial E_s}{\partial \sigma_j} \quad (13)$$

where β is a learning parameter which is related to the rate of convergence. We compare this third method with our proposed method to evaluate the regularization performance.

2.4.3. The proposed Trend-Oriented RBF (TRBF)

The propose Trend-Oriented RBF, to avoid overfitting, aims to find the approximated solution of centers and σ s after limiting the complexity of the network. But the cost function E_T of the TRBF is different from the former networks.

$$E_T = \frac{1}{2} \sum_p \sum_i (z_{ip} - t_{ip})^2 + \lambda \sum_j \frac{1}{\sigma_j^2} \quad (14)$$

By adding an inverse σ^2 penalty term to the minimization criterion, the regularization is more accelerated. The regularization parameter λ has to be chosen a priori or estimated from the data. By adopting this cost function E_T each neuron tends to regularize the problem while satisfying the permissible sum-of-squares error. In the gradient-descent method the delta values of y_j and σ_j that minimize E_T are regarded as below:

$$\Delta Y_j = -\beta \frac{\partial E_T}{\partial Y_j} \quad (15)$$

$$\Delta \sigma_j = -\beta \frac{\partial E_T}{\partial \sigma_j} \quad (16)$$

where β is a learning parameter which is related to the rate of convergence.

Figures 16, 17 and 18 show the classification performance based on the feature combination between K-L01 and K-L04

K-L 04



K-L 01

Figure 16 Full size RBF. Training Error = 0.0. Training Error comes from the square root of E (Eq. 5).

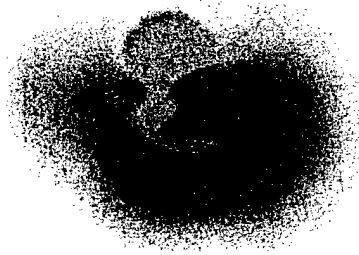
K-L 04



K-L 01

Figure 17 Former RBF. Training Error = 83.0. Training Error comes from the square root of E (Eq. 5).

K-L 04



K-L 01

Figure 18 Proposed TRBF. Training Error = 88.2. Training Error comes from the square root of E (Eq. 5).

K-L 04



Figure 19 $\lambda = 0.001$. Training Error = 83.4, where the number of neuron is 20. Training Error comes from the square root of E (Eq. 5)

K-L 04



Figure 20 $\lambda = 0.003$. Training Error = 88.2, where the number of neuron is 20. Training Error comes from the square root of E (Eq. 5)

K-L 04



Figure 21 $\lambda = 0.01$. Training Error = 90.7, where the number of neuron is 20. Training Error comes from the square root of E (Eq. 5)

after the learning phase regarding the full size RBF, the former RBF, and the proposed TRBF, respectively. The full size RBF has the same number of neurons as the number of training data. And the positions of the full size RBF neurons are identical to the positions of the training data. The former RBF has the converging method of equations 10 and 11 in which the cost criteria is not E_s (Eq. 9) but E (Eq. 5). Both the former RBF and the proposed TRBF have 20 neurons. The dark dots and the white dots denote benign pattern and malignant patterns, respectively. The brightness of the classification areas are the opposite, the gray and dark areas denote benign and malignant classes, respectively. The black dots denote the neuron positions. The white area is classified as neither benign nor malignant.

3. RESULTS

The performances of K-L features and the TRBF neural network are evaluated through the Round-Robin method, where one sample is tested after the learning based on the rest of the 127 samples. We use the K-L feature combination between K-L01 and K-L04. The LABROC4 algorithm[10] developed by Metz et al. is used to fit the ROC curve to the continuous data from the TRBF output. As you can see in Figure 18, our neural network outputs benign value [0, 1] and malignant value [0, 1] separately. If the output value is zero, the data is classified as negative. For example if both outputs are zero, it means that its data is classified as neither benign nor malignant. However, the LABROC4 algorithm does not allow such separated outputs, so the output conversion must be applied to make the combined output in [0, 1]:

$$t_f = 0.5 - (t_b - t_m) \times 0.5, t_b \geq t_m \quad (17)$$

$$t_f = 0.5 + (t_m - t_b) \times 0.5, t_m \geq t_b$$

where t_f is the combined output, t_b is the output of the benign node, and t_m is the output of the malignant node. If the combined value t_f is 0, the calcification is benign. If the combined value t_f is 1, the calcification is malignant.

3.1. Comparison based on regularization parameter λ

The λ parameter is able to control the degree of regularization. The bigger the λ parameter, the more regularization. Figures 19, 20 and 21 show the classification results with various λ parameters. The λ parameters in Figures 19, 20 and 21 are 0.001, 0.003, and 0.01, respectively. Figure 22 shows an ROC curve at various λ parameters. The appropriate regularization is required to achieve a high performance of Az on the ROC analysis.

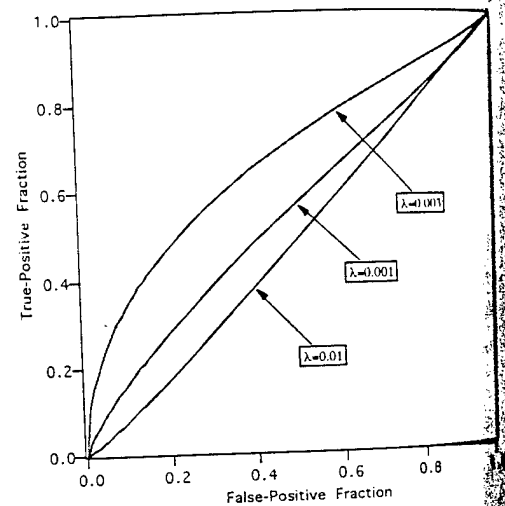


Figure 22 ROC curve at various parameters $A_z = 0.55$ at $\lambda = 0.001$, $A_z = 0.68$ at $\lambda = 0.003$, $A_z = 0.49$ at $\lambda = 0.01$

K-L 04



K-L 01

Figure 23 12 neurons. Training Error = 93.8. where the λ parameter is 0.003. Training Error comes from the square root of E (Eq. 5)

K-L 04



K-L 01

Figure 24 20 neurons. Training Error = 88.2. where the λ parameter is 0.003. Training Error comes from the square root of E (Eq. 5)

K-L 04



K-L 01

Figure 25 24 neurons. Training Error = 80.4. where the λ parameter is 0.003. Training Error comes from the square root of E (Eq. 5)

3.2. Comparison based on the number of neurons

The degree of regularization deeply depends on the number of neurons. According to the size of this database, the upper limit of the number of neurons seems to be around 20. Because the number of the free parameter of the proposed TRBF is 100 when the number of neuron is 20. The network generally becomes more regularized at the decrement of the number of neuron. But if there is not enough neurons to form the trend of the distribution, this situation cannot be called regularization. Figure 23, 24 and 25 show the classification results with various number of neurons. The number of neuron in Figure 23, 24 and 25 are 12, 20, and 24, respectively. It seems that there is not enough neurons in Figure 23. Figure 26 shows ROC curve at various number of neurons. The appropriate number of neurons is required to achieve a high performance of Az on the ROC analysis.

4. DISCUSSION AND CONCLUSION

We propose the feature selection method and the Trend-Oriented Radial Basis Function (TRBF). The two dimensional K-L features are more distinguishable than the raw two dimensional feature combination. The proposed TRBF network has a better ability to define the trend of the distribution than the former RBF. Figure 27 shows the comparison of ROC analysis between our proposed system and one trained radiologist. According to ROC curve this system indicates a better performance than one trained radiologist. But we do not mean that these selected features are better than this radiologist. As you can see in Figure 14, the pattern distribution of K-L Feature 01 and 04, two classes, benign and malignant, are almost overlapped. It is true that the proposed TRBF neural network boldly regularizes this pattern distribution and that its regularization sometimes works better than this radiologist.

Our feature works are (1) more theoretical discussion of the appropriate number of neuron and regularization parameter (2) the gathering of more data samples (3) more investigation of how to use ROC analysis against RBF type neural network. The RBF type neural network has a problem using ordinary ROC analysis. Unlike the widely used technique of error backpropagation, the RBF type neural network may have the area which is not classified as any class. If the pattern falls on such non-classified area, every output should be negative. As

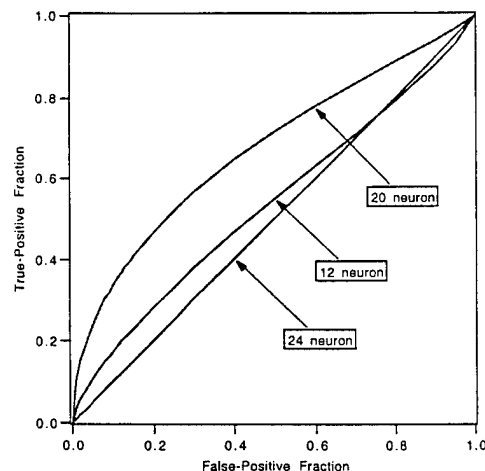


Figure 26 ROC curve at various λ parameters Az = 0.54 at 12 neurons, Az = 0.68 at 20 neurons, Az = 0.50 at 24 neurons

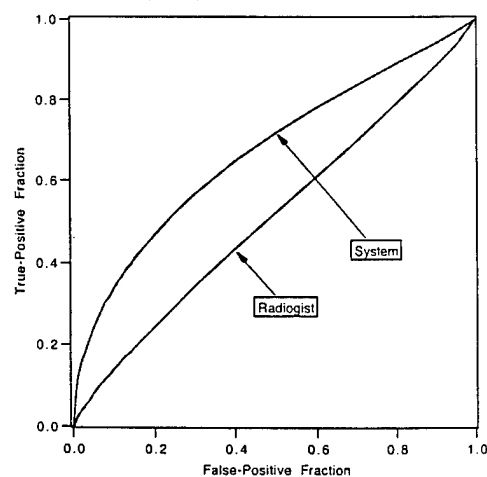


Figure 27 ROC comparison between system and one radiologist. The system consists of 20 neurons and the regularization $\lambda = 0.003$

we discuss in the equation 17, the system should make its output 0.5 in such a case. When the data pattern falls on the area which is overlapped with two classes, the system also outputs 0.5. In the former case, that pattern might have to be rejected. We can only say here that another ROC analysis is required for the RBF type neural network.

5. ACKNOWLEDGMENTS

This work is supported in part by a US Army Grant (DAMD17-94-V-4015). The content of this manuscript does not necessarily reflect the position or the policy of the U.S. government. The authors gratefully acknowledge the editorial support of Miss Susan Kirby.

6. REFERENCES

- [1] National Center for Health Statistics. Vital Statistics of the United States, 1987. Vol. 2. Mortality. Part A, DHHS Publication No. (PHS) 90-1101 (Government Printing Office, Washington DC, 1990)
- [2] G. Svane, E. J. Potchen, A. Sierra, and E. Azavedo, Screening Mammography: Breast Cancer Diagnosis in Asymptomatic Women, 1993, Mosby-Year Book
- [3] E. A. Sickles, Mammographic features of 300 consecutive nonpalpable breast cancer, Am. J. Roentgenol, 146, pp661-663, 1986
- [4] Y. Chris Wu, M. T. Freedman, A. Hasegawa, R. A. Zuurbier, S-C B. Lo, and S. K. Mun, Classification of Microcalcifications in Radiographs of Pathologic Specimens for the Diagnosis of Breast Cancer. Acad Radiol 1995;2: pp199-204
- [5] R. D. Turco et al, The significance of mammographic calcifications in early breast cancer detection, Radiol Med 72: pp7-12, 1986
- [6] D. S. Broomhead, and D. Lowe Multi-variable functional interpolation and adaptive networks, Complex Anal. Ser. B2, 205, 1988
- [7] D. E. Rumelhart and J. L. McClelland, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundation. The MIT Press, Cambridge, MA.
- [8] C. Bishop, Improving the Generalization Properties of Radial Basis Function Neural Networks, Neural Comp. 3, pp579-588, 1991
- [9] S. Chen, C. F. N. Cowan, and P. M. Grant, Orthogonal least squares learning for radial basis function networks. IEEE Transact. Neural Networks 2(2), pp302-309, 1991
- [10] Tomaso Poggio and Federico Girosi, Networks for Approximation and learning. Proceeding of the IEEE, Vol. 78, No. 9, pp1481-1497, September 1990

Implementation and Testing of Digital Mammography for MDIS Environment

December 31, 1995

**Georgetown University Medical Center
Department of Radiology
Imaging Science and Information Systems (ISIS) Center
2115 Wisconsin Ave., NW, Suite 603
Washington, D.C. 20007**

<http://www.isis.imac.georgetown.edu>

Project leaders

Seong K. Mun, Ph.D., Director of ISIS Center
Matthew Freedman, M.D., M.B.A., Clinical Director of ISIS Center

Principal Investigator

Jyh-Shyan Lin, Ph.D.
Research Associate
☎ 202/687-7953
Fax: 202/784-3479
E-mail: jslin@isis.imac.georgetown.edu

Table of Contents

1. INTRODUCTION	1
1.1. Objective	1
1.2. Project Schedule	1
1.3. Participants	1
2. EXPERIMENTAL SYSTEM CONFIGURATION	2
2.1. Set up of an Experimental System	2
2.2. Study of Physics Characteristics of Data Acquisition Modules	4
3. DATA COLLECTION AND MANAGEMENT	5
3.1. Case Selection and Data Collection Protocol	5
3.2. Data Management	5
3.3. Image Processing Protocol	5
4. COMPARATIVE READING	7
4.1. Preference Study of Hard Copy and Soft Copy of Screen Films	7
4.2. Questionnaire for Comparative Reading	9
5. QUESTIONNAIRE RESPONSES AND DATA ANALYSIS	10
5.1. Questionnaire Responses	10
5.2. Comments of the Five Radiologists	22
6. CONCLUSIONS	23
7. ADDENDA	24
7.1. Acronym / Symbol Definition	24
7.2. References	24

List of Figures

Figure 1	Experimental System Configuration -----	3
Figure 2	Comparative Reading -----	8
Figure 3.1	Questionnaire Responses of Radiologist 1 -----	13
Figure 3.2	Questionnaire Responses of Radiologist 2 -----	15
Figure 3.3	Questionnaire Responses of Radiologist 3 -----	17
Figure 3.4	Questionnaire Responses of Radiologist 4 -----	19
Figure 3.5	Questionnaire Responses of Radiologist 5 -----	21

List of Tables

Table 1	Database of 50 Clinical Relevant Cases -----	6
Table 2.1	Questionnaire Responses of Radiologist 1 -----	12
Table 2.2	Questionnaire Responses of Radiologist 2 -----	14
Table 2.3	Questionnaire Responses of Radiologist 3 -----	16
Table 2.4	Questionnaire Responses of Radiologist 4 -----	18
Table 2.5	Questionnaire Responses of Radiologist 5 -----	20

1. INTRODUCTION

1.1. Objective

The objective of this project was to evaluate image quality of digital mammography based on computed radiography (CR) and high resolution film digitizer and develop a strategy to implement digital mammography into the military diagnostic imaging systems (MDIS) environment.

1.2. Project Schedule

Start Date: October 1, 1994

End Date: December 31, 1995

There are total 15 months which are divided into 5 quarters.

First Quarter (October 1, 1994 - December 31, 1994)

Budget request: 40%

- Goals:
1. Set up of an experimental system.
 2. Study of physics characterization of data acquisition modules.
 3. Data collection and database management.

Second Quarter (January 1, 1995 - March 31, 1995)

Budget request: 20%

- Goals:
1. Collection of 10 clinically relevant cases in all modalities.
 2. Comparative reading of 10 cases in hard copy and soft copy.

Third Quarter (April 1, 1995 - June 30, 1995)

Budget request: 20%

- Goals:
1. Collection of 10 clinically relevant cases in all modalities.
 2. Comparative reading of 10 cases in hard copy and soft copy.
 3. Technical evaluation of MDIS database to determine suitability for digital mammography.

Fourth Quarter (July 1, 1995 - September 31, 1995)

Budget request: 10%

- Goals:
1. Collection of 20 clinically relevant cases in all modalities.
 2. Comparative reading of 20 cases in hard copy and soft copy.
 3. Determination of MDIS workstation functionality for digital mammography.

Fifth Quarter (October 1, 1995 - December 31, 1995)

Budget request: 10%

- Goals:
1. Comparative reading of 10 cases in hard copy and soft copy.
 2. Data analysis.
 3. Final report.

1.3. Participants

- Case selection
 - Rebecca Zuurbier, M.D., Director of Breast Imaging, GUMC
 - Jaquelyn Hogge, M.D., Mammographer, GUMC
- Collection of Fuji CR9000 and screen film mammograms - Dot Artz, R.T., R.M.
- Digitization of screen film mammograms - Two part time students of Georgetown University
- Data management and image processing - Jyh-Shyan Lin, Ph.D., ISIS Center, GUMC
- Vicom display workstation - Akira Hasegawa, Ph.D., ISIS Center, GUMC
- Comparative reading - Five board-certified radiologists of Radiology Department, GUMC
 1. Matthew T. Freedman, M.D., M.B.A. - Clinical Director of ISIS Center
 2. Rebecca Zuurbier, M.D. - Director of Breast Imaging
 3. Jaquelyn Hogge, M.D. - Mammographer
 4. Wendelin Hayes, D.O. - Associate Professor of Radiology
 5. Curtis Green, M.D. - Radiologist

2. EXPERIMENTAL SYSTEM CONFIGURATION

2.1. Set up of an Experimental System

The experimental system configuration is shown in Figure 1. The key components are: a Fuji CR9000 (FCR9000), a Lumisys film digitizer, a data acquisition and system management (DASM) host Sun workstation, a digitizer host Sun workstation, and a Vicom high resolution display workstation.

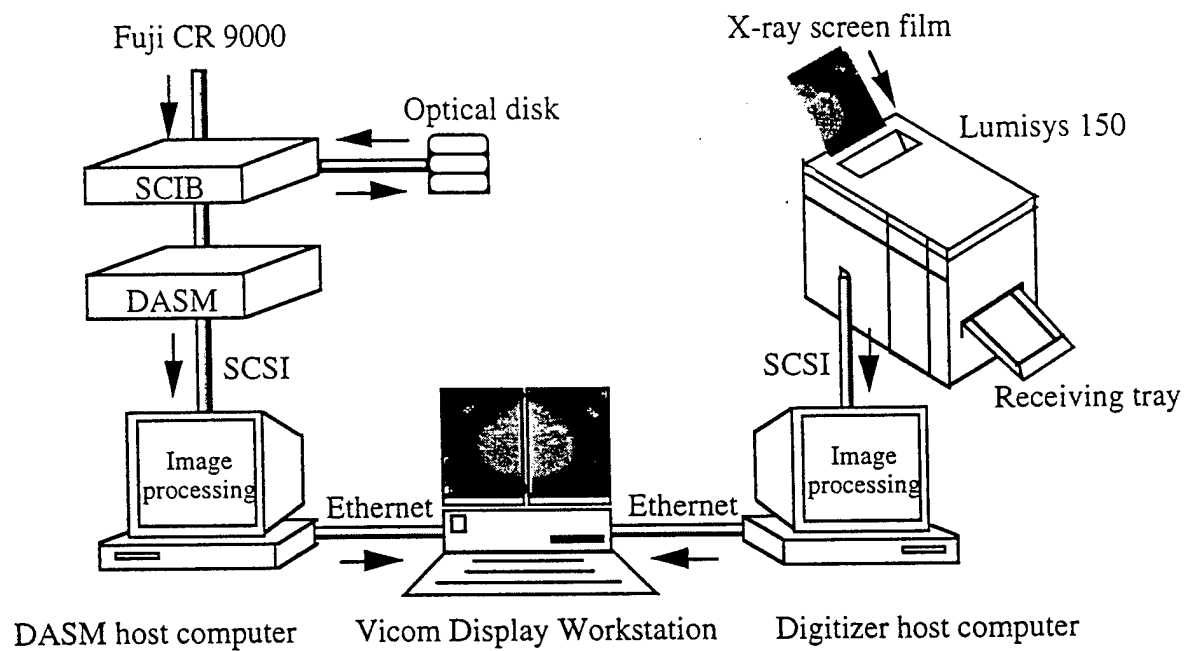
The Fuji Computed Radiography 9000 System

The DASM controls the data transfer from the FCR9000 to the host Sun workstation through the small computer system interface (SCSI) cable. Each CR image contains a 2048 bytes image header. The CR images were then processed and transferred to the Vicom display workstation. A second channel interface board (SCIB) was installed for the purpose of simultaneously interfacing with DASM and an optical disk device. Original CR images including biopsy specimen were backed up in the optical disks.

The Lumisys 150 High Resolution Laser Film Digitizer

Conventional screen films (SF) were digitized by a Lumisys high resolution film digitizer. Digitized SF images were transferred to the digitizer host Sun computer through the SCSI, processed, and then transferred to the Vicom display workstation.

- The Lumisys 150 meets the film digitizer requirements by MDIS which are variable spot size, different film sizes (from 8" $\times$ 10" up to 14" $\times$ 17"), minimum spot size of 210 microns, and 10 bits minimum dynamic range (i.e., 10 bits per pixel). The density resolution and precision is a linear function from 0 to 3.5 optical density (OD). The Lumisys 150 film digitizer can determine the film size automatically. It can convert films sized from 8" $\times$ 10" up to 14" $\times$ 17" into digital images of 2048 pixels $\times$ 2560 lines with 12 bits per pixel (10 Mbytes). The 8" $\times$ 10" films can also be digitized at 50 μ m spot size which is equivalent to 4096 pixels $\times$ 5120 lines with 12 bits per pixel (= 40 Mbytes). The laser spot size can be auto-adjusted accordingly.
- The Lumisys 150 does not provide automatic sheet feeder. A more advanced Lumisys 200 digitizer will have the function of automatic sheet feeder which is able to hold and automatically feed at least 10 x-ray films of intermixed sizes. The digitizer generates two file formats: raw binary image and TIFF compressed images for image preview. The image previewing software is the x-window view (XV) software program run on the SUN digitizer host computer. The Lumisys 150 does not support the American College of Radiology (ACR)-National Electrical Manufacturers Association (NEMA) standard.
- Phantoms, such as CIRS, ACR, CDMAM, step wedge, and SMPTE, were digitized at 50 micron and 100 micron and displayed on the Vicom display workstation. The Lumisys digitized phantoms showed satisfactory quality.
- The Lumisys high resolution digitizer (Model: 150) is set up to digitize conventional SF at 100 μ m resolution. Currently, the Lumisys digitizer is now located in the mammography reading room in Georgetown University Medical Center (GUMC) and is being used routinely to digitize clinical cases. Each clinical case consists of four recent and four previous mammograms with pathology reports.
- The Vicom display workstation is a Pixar-based system which has four high resolution MegaScan monitors. Each display monitor can display image data of 2048 $\times$ 2560 $\times$ 8 bits. The display workstation provided window-and-level function for adjustment of image brightness and contrast. Each monitor can be manipulated individually.



DASM: data acquisition and system management
 SCSI: small computer system interface
 SCIB: second channel interface board

Figure 1. Experimental System Configuration.

2.2. Study of Physics Characteristics of Data Acquisition Modules

- **Quality control (QC) on a Fuji printer and a Fuji CR9000**

- QC of 18 cm × 24 cm image plate (IP):

Every morning each 18 cm × 24 cm IP received secondary erasure through FCR 9000 to ensure that a properly erased IP was in each cassette before any images were obtained. This process reduces possible residual radiation in the plate and ensures the best possible image.

- QC of Fuji 9000 laser printer (LP):

Every morning a cleaning film was run through the LP, then the density and check density were adjusted. A step wedge was produced from both steps. After the step wedge pattern is obtained from the check density, the optical density values was displayed on the LP display. The values and graph similar to sensitometry graph for a film processor were recorded.

- QC of mammography film processor:

Normal sensitometry graphing, cassette cleaning, exposure monitoring were performed according to ACR standard.

The quality control for mammography film processor and Fuji LP was performed by in-house engineering.

- **Lumisys Film Digitizer and the Vicom Display Monitors**

- Lumisys film digitizer:

We have performed maintenance of the Lumisys scanner periodically. The maintenance includes dust removal of the reflection mirror, the pinch roller, and the aperture of the detector, and adjustment of the voltage of both the main and the reference detectors [1]. The digitizer is also calibrated based on a Lumisys test pattern. The maintenance is to assure the digitization quality and calibration of the digitizer.

- Vicom workstation:

The Vicom system has four 2k × 2.5k display monitors. Each monitor can display a full data set of a 8 × 10" mammogram digitized at 100 micron. The brightness is 60 foot-Lamberts and the gray level is 256 shades. Since there is no single set of window (i.e., contrast) and level (i.e., brightness) values for every image display, the look-up table of the system was preset to a set of window and level values visually determined by a physicist, a registered mammographer, and a scientist. The purpose is to adjust the two monitors such that they have similar intensity and contrast. The contrast and brightness can be adjusted by using a trackball during the study of comparative reading.

3. DATA COLLECTION AND MANAGEMENT

3.1. Case Selection and Data Collection Protocol

The database used in this study included 50 cancerous and non-cancerous cases. All the cases contained bilateral breast images, radiology reports, and related pathology reports. Biopsy was used as the standard proof of cancer or non-cancer. Each set of bilateral mammograms contained either left-right cranio caudal (CC) or left-right mediolateral oblique (MLO) views. We collected 50 CR soft copies, 50 CR hard copies, 50 SF hard copies, and 50 soft copies of digitized films. Table 1 lists the collected database of the 50 clinically relevant cases. The table shows fake patient identification numbers (ID) for patient confidentiality. A similar table with real patient name and ID is available and kept securely at GUMC. In the table, "BRST" represents breast, "BX" represents biopsy, "MAMMO" means mammogram, and "POS FOR CA" means positive for cancer. One half of the cases contained microcalcifications, one half of cases contained masses, and some cases of architecture distortions were included. Films were digitized in such a way that most patient demographic data were excluded.

3.2. Data Management

- The Lumisys model 150 film digitizer was set up at 100 micron scanning mode.
- Actual size of hard copy:
 - Hard copy SF: 8 inch $\times$ 10 inch.
 - Hard copy CR: 11 inch $\times$ 14 inch.
- Actual size of soft copy:
 - 100 micron/pixel for the CR soft copies.
An image size is 8,404,136 bytes (including image header) per CR soft copy.
 - 100 micron/pixel for the soft copies of digitized SF.
An image size is 10,485,760 bytes (including image header) per digitized SF.
- The soft copies of SF and CR both contain 2,048 bytes image header.
- Storage of soft copies: DEC optical disks (595 MB per optical disk).
A set of two digitized films (left and right breasts) and two soft copies of CR (left and right breasts) has approximately 40 Mbytes, hence for 50 cases of processed and non-processed images we need $50 \text{ cases} \times 40 \text{ Mbytes/set} \times 2 = 4 \text{ Gbytes} = 8 \text{ DEC optical disks}$. For other ISIS research projects, such as the classification of benign and malignant microcalcifications, we have also digitized the mammograms at 50 μm resolution using the Lumisys film digitizer. Digitized image data were stored on optical disks as well as a Hewelett-Packard (HP) jukebox.

3.3. Image Processing Protocol

- Decode the image headers of Lumisys digitized images.
- Encode image data - rearrange display scanning lines.
- Attach Pixar II image header.
- Move images to the fast disks of Vicom workstation for display.

Table 1 Database of 50 Clinical Relevant Cases

ID #	DATE	DIGITAL VIEWS	BRST FOR BX	MAMMO FINDINGS	PATHOLOGY RESULTS	POS FOR CA
1	6/8/95	Bilat. CC	Left	Ca++	1mm ductal CA +DCIS	yes
2	9/29/95	Bilat. CC	Right	mass	fibroadenoma	no
3	12/12/94	Bilat. CC	Left	mass	fibrocystic changes	no
4	12/8/94	Bilat. CC	Right	spic. mass with Ca++	intraductal CA	yes
5	1/24/95	Bilat. CC	Right	Ca++	invasive intraductal CA	yes
6	7/13/95	Bilat. CC	Right A73	small mass	ductal hyper. Ca++	no
7	12/20/94	Bilat. CC	Right	spic. mass	ductal hyper. Ca++	no
8	3/3/95	Bilat. MLO	Right	mass	fibrocystic changes	no
9	6/12/95	Bilat. CC	Left	mass with Ca++	small fibroadenoma	no
10	6/1/95	Bilat. CC	Left A111	mass with Ca++	intraductal CA	yes
11	8/24/95	Bilat. CC	Bilat.	Ca++	intraductal CA	yes
12	7/24/95	Bilat. CC	Right A90	Ca++	fibrocystic changes	no
13	6/19/95	Bilat. CC	Right A48	multiple masses	8mm infiltrating ductal CA	yes
14	6/2/95	Bilat. CC	Left	mass	stromal fibrosis	no
15	6/23/95	Bilat. CC	Right	micro Ca++	small fibroadenoma	no
16	6/12/95	Bilat. CC	Left	mass	1. radial scar 2. ID CA	yes
17	6/23/95	Bilat. CC	Left A58	Spic. mass with Ca++	radial scar with ID CA	yes
18	7/31/95	Bilat. CC	Right A113	Ca++	cystic change Ca++	no
19	7/11/95	Bilat. CC	Right A98	scattered Ca++	normal	no
20	3/10/95	Bilat. CC	Right A55	Ca++	intraductal CA	yes
21	8/2/95	Bilat. CC	Right A113	posterior Ca++	3cm papill. ductal CA	yes
22	4/24/95	Bilat. CC	Right A64	dense mass	intraductal CA	yes
23	12/12/94	Bilat. CC	Right	Ca++ at prev. surg. site	fibrosis	no
24	2/3/95	Bilat. CC	Left	Ca++	intraductal CA insitu atyphyper.	yes
25	3/6/95	Bilat. MLO	Left	Ca++	fibroadenoma	no
26	12/29/94	Bilat. CC	Left	mass with ?Ca++	atypical intrduct. hyperp.	no/yes
27	6/29/95	Bilat. CC	Right	mass/prev. CA	invas. ductal CA with Ca++	yes
28	7/20/95	Bilat. CC	Right A152	mass	intraductal CA	yes
29	12/8/94	Bilat. CC	Right	spic. mass with Ca++	intraductal CA	no
30	6/22/95	Bilat. CC	Right A70	Ca++	intraductal papill.	no
31	2/27/95	Bilat. CC	Right	subtle Ca++	ductal hyper. Ca++	no
32	12/8/94	Bilat. CC	Right	mass	ductal hyperplasia	no
33	10/9/95	Bilat. CC	Right+Left	masses	infiltrating ductal CA	yes
34	8/17/95	Bilat. CC	Left	SUBTLE microCa++	extensive ductal CA insitu	yes
35	12/16/94	Bilat. CC	Right	Ca++ dense tissue	intraductal CA	yes
36	6/6/95	Bilat. CC	Left	mass	infiltrating ductal CA	no
37	6/12/95	Bilat. CC	Right	mass	fibroadenoma	no
38	12/5/94	Bilat. CC	Left	spic. mass with ?Ca++	intraductal CA	yes
39	12/27/94	Bilat. CC	Left	casting Ca++	fibrocystic changes	no
40	6/22/95	Bilat. CC	Right A108	2 masses	5mm and 9mm invas. duct. CA	yes
41	6/1/95	Bilat. MLO	Left A74	mass with Ca++	fibrocystic change	no
42	7/28/95	Bilat. CC	Left A116	Ca++	fibrocystic change Ca oxalate	no
43	3/6/95	Bilat. CC	Left	Ca++	hyperplasia	no
44	7/11/95	Bilat. MLO	Right A88	palpable mass not seen	no lesions indentified	no
45	7/13/95	Bilat. CC	Bilateral R A154	Ca++	R-early insitu L-fibrocyst.	yes-no
46	11/22/94	Bilat. CC	Right	Ca++	infiltrating CA	yes
47	1/13/95	Bilat. CC	Left	spic. mass with Ca++	ductal CA in situ	yes
48	8/17/95	Bilat. MLO	Left	hx nodular fasciitis (benign)	mass excised no bx report	no
49	12/15/94	Bilat. CC	Left	nodule	fibrocystic changes	no
50	11/17/95	Bilat. CC	Right	subtle Ca++ mid post.	intraductal CA comedo type	yes

4. COMPARATIVE READING

4.1. Preference Study of Hard Copy and Soft Copy of Screen Films

We performed a preference study to evaluate image quality of digital mammography derived from conventional screen films (i.e., hard copies) and a high resolution film digitizer (i.e., soft copies). We wanted to compare one view (either CC or MLO) of original mammograms with the digitized film at 100 micron spatial resolution. The SF hard copies were digitized and directly converted to Vicom format (see Section 3.3 Image Processing Protocol). Five board-certified radiologists, Dr. Wendelin Hayes (radiologist 1), Dr. Jaquelyn Hogge (radiologist 2), Dr. Matthew T. Freedman (radiologist 3), Dr. Curtis Green (radiologist 4), and Dr. Rebecca Zuurbier (radiologist 5), participated in this study of preference reading of soft and hard copies.

• Set up of environment for human readers

The reading environment for human readers was set up as follows:

- Low level ambient light.
- No time limit in the comparative reading of each case.
 - Radiologists can use trackball to control the window-and-level while viewing soft copies on the Vicom monitors. Note that, in this study, the window-and-level was controlled by the principal investigator while the radiologist requested the change of the intensity and contrast of the displayed images.
- Light box was used with emulsion side facing the reader.
- Human readers have to fill out the questionnaires (see Section 4.2) provided in this study
- Pairs of hard and soft copies of an SF will be displayed on the light box and two Vicom monitors, respectively (Figure 2). The human readers filled out the questionnaire.
- Hard copies and soft copies are both displayed in anatomically corrected views.
- Each hard copy was labeled as #1, #2, etc. The patients' demographic data on the screen films were covered up (on the dull side of the SF) by using black electrical tape.

• Comparative reading and reading-order effects

- Sample size (Table 1): 25 cases of proven biopsy cancer and 25 cases that appeared malignant but were proven by biopsy to be benign.
- In the first 25 cases, two of the human observers first read hard copies and then soft copies of the SF while the other three read the two modalities in the opposite order. In the second 25 cases, the two of the human observers read soft copies of the SF first and then hard copies while the other three read the two modalities in the opposite order. The purpose of this reading order arrangement was to reduce the *reading-order effects* [2]. The reading-order effects result in the biases that occur in a situation when two or more equivalent images of a particular patient are read in different order by the same observer. It is normally the case that the image read last will tend to be interpreted more accurately than the image read first if any relevant information is retained by the observer from a reading of one image of the patient to the next.
- During the comparative reading, the radiologist can always look back-and-forth to the two image modalities retrospectively and repeatedly.
- The questionnaire contained nine questions with check mark areas. If no lesions, such as microcalcifications (questions 1 and 2), masses (question 3 and 4), and asymmetric densities (questions 5 and 6), were found, the check mark areas were left blank. On the other hand, if both hard and soft copies of SF were equally well rated, both areas were marked.

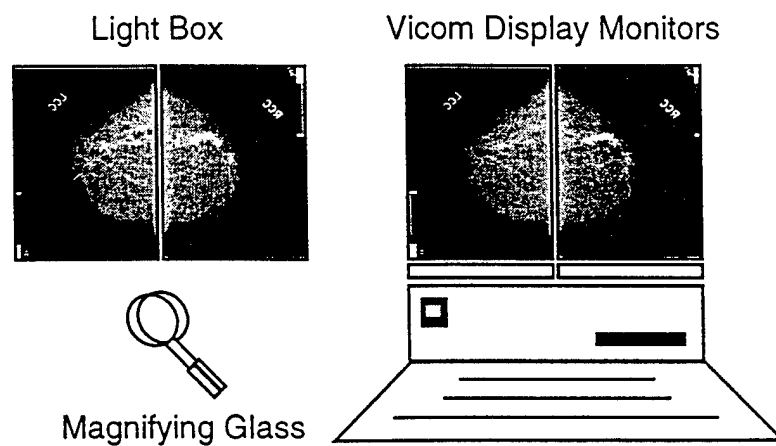


Fig. 2. Comparative Reading.

4.2. Questionnaire for Comparative Reading

Soft copy vs. Hard copy of SF

Viewer initial: _____ Date: _____

1. On which do you detect microcalcifications better?
Hard copy _____ or soft copy _____.
2. Which one better characterized the microcalcifications as to malignant vs. benign?
Hard copy _____ or soft copy _____.
3. On which do you detect masses better?
Hard copy _____ or soft copy _____.
4. Which one better characterized the masses as to malignant vs. benign?
Hard copy _____ or soft copy _____.
5. On which do you detect asymmetric density better?
Hard copy _____ or soft copy _____.
6. Which one better characterized the asymmetric density?
Hard copy _____ or soft copy _____.
7. Do you have a preference? Hard copy _____ or soft copy _____.
8. Do you see anything of clinical importance on the hard copy that you do not see on the soft copy? Yes _____ or No _____
If yes, what? _____
9. Do you see anything of clinical importance on the soft copy that you do not see on the hard copy? Yes _____ or No _____
If yes, what? _____
- *Comment: _____

5. QUESTIONNAIRE RESPONSES AND DATA ANALYSIS

5.1. Questionnaire Responses

From the questionnaire (a pair of soft vs. hard copies of SF), we obtained histogram bars of the number of soft and hard copies of SF which better characterize microcalcifications, masses, and asymmetric density as to malignant vs. benign. Moreover, we will have the number of preferred soft and hard copies of the five human readers. The five radiologists' responses to the specific questions in the questionnaire are shown in Table 2.1 - Table 2.5 and Figure 3.1 - Figure 3.5. In each table, the columns of Q1 - Q9 represent answers for questions 1 through 9 for the 50 cases, H represents hard copy, S represents soft copy, Y means yes, and N means no. The answer "both" or "H/S" means that there was no preference, i.e., both hard and soft copies were equally well accepted. The blank and the answer "neither" mean that the question was not applicable, i.e., the disease patterns were not seen. In each Figure 3.1 - 3.5, R1 - R9 represent responses to Q1 - Q9, respectively, H means hard copy and S means soft copy.

For all the questions 1 - 7, Dr. Freedman prefers hard copy to soft copy in all cases.
For all the questions 1 - 6, Dr. Zuurbier prefers hard copy to soft copy in all cases.

Responses to question 1:

- On which do you detect microcalcifications better?

Dr. Green prefers hard copy to soft copy in 38 calcification cases and no preference in 12 cases.

Dr. Hayes prefers hard copy to soft copy in all calcification cases.

Dr. Hogge prefers hard copy to soft copy in 47 calcification cases and no preference in 3 cases.

Responses to question 2:

Which one better characterized the microcalcifications as to malignant vs. benign?:

Dr. Green prefers hard copy to soft copy in all 37 calcification cases and no preference in 13 cases.

Dr. Hayes prefers hard copy to soft copy in calcification cases.

Dr. Hogge prefers hard copy to soft copy in 48 calcification cases and no preference in two cases.

Responses to question 3:

- On which do you detect masses better?

Dr. Green prefers hard copy to soft copy in 6 mass cases and no preference in 26 cases.

Dr. Hayes prefers hard copy to soft copy in 23 mass cases and soft copy to hard copy in 6 cases.

Dr. Hogge prefers hard copy to soft copy in 5 mass cases, soft copy to hard copy in 10 cases, and no preference in 35 cases.

Responses to question 4:

- Which one better characterized the masses as to malignant vs. benign?

Dr. Green prefers hard copy to soft copy in 11 mass cases and no preference in 21 cases.

Dr. Hayes prefers hard copy to soft copy in 28 mass cases and soft copy to hard copy in one case.

Dr. Hogge prefers hard copy to soft copy in 4 mass cases, soft copy to hard copy in 8 cases, and no preference in 38 cases.

Responses to question 5:

- On which do you detect asymmetric density better?

Dr. Green prefers hard copy to soft copy in 3 asymmetric density cases and no preference in 18 cases.

Dr. Hayes prefers hard copy to soft copy in 6 asymmetric density cases and soft copy to hard copy in 2 cases.

Dr. Hogge prefers hard copy to soft copy in 5 asymmetric density cases, soft copy to hard copy in 14 cases, and no preference in 31 cases.

Responses to question 6:

- Which one better characterized the asymmetric density?

Dr. Green prefers hard copy to soft copy in 8 asymmetric density cases and no preference in 14 cases.

Dr. Hayes prefers hard copy to soft copy in all asymmetric density cases.

Dr. Hogge prefers hard copy to soft copy in 4 asymmetric density cases, soft copy to hard copy in 9 cases, and no preference in 37 cases.

Responses to question 7:

- Do you have a preference?

Dr. Green prefers hard copy to soft copy in all 50 cases.

Dr. Hayes prefers hard copy to soft copy in 48 cases, soft copy to hard copy in one case, and no preference (i.e., equally well accepted) in one case.

Dr. Zuurbier prefers hard copy to soft copy in 45 cases and no preference in 5 cases.

Dr. Hogge prefers hard copy to soft copy in 36 cases, soft copy to hard copy in 8 cases, and no preference in 6 cases.

Responses to question 8:

- Do you see anything of clinical importance on the hard copy that you do not see on the soft copy?

The five radiologists saw microcalcifications on the hard copy that they did not see on the hard copy. Sometimes dramatic window-and-level adjustment may enhance the subtle microcalcifications, however, the other breast will become either too dark or too bright. Though both masses and architecture distortion can be seen on the soft copies, they seem to increase the perception of normal soft tissue structures as abnormal.

Responses to question 9:

- Do you see anything of clinical importance on the soft copy that you do not see on the hard copy?

The five radiologists saw microcalcification-like objects on the soft copy which are actually film defects on the hard copies. The five radiologists did not see anything else of clinical importance on the soft copy that they did not see on the hard copy.

Table 2.1 Questionnaire Responses of Radiologist 1

Radiologist 1	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
ID #	H/S	H/S	H/S	H/S	H/S	H/S	H/S	Y/N	Y/N
1	H	H					H	N	N
2	H	H	H	H			H	N	N
3	H	H	H	H			H	N	N
4			H	H			H	N	N
5	H	H	H	H	H	H	H	N	Y
6	H	H			H	H	H	N	N
7	H	H					H	Y	N
8	H	H					H	Y	N
9	H	H	H	H	H	H	H	Y	N
10	H	H					H	Y	N
11	H	H	H	H			H	Y	N
12			H	H			H	Y	N
13					H	H	H	N	N
14	H	H	S	H			H	Y	N
15	H	H	H	H			H	N	N
16			H	H			H	Y	N
17	H	H	H	H			H	Y	N
18	H	H	H	H			H	N	N
19					H	H	H	N	N
20	H	H					H	Y	N
21	H	H					H	Y	N
22	H	H	H	H			H	Y	N
23	H	H					H	Y	N
24			H	H			H	N	N
25	H	H	H	H			H	N	N
26			H	H			H	Y	N
27	H	H	H	H			H	Y	N
28			H	H			H	N	N
29			S	H			H	N	N
30			H	H			H	N	N
31	H	H	H	H	H	H	H	N	N
32			H	H			H	N	N
33			H	S			S	N	Y
34	H	H					H	N	N
35			S	H			H	N	N
36	H	H					H	N	N
37	H	H					H	Y	N
38	H	H	H	H			H	N	N
39					S	H	neither	N	N
40	H	H					H	N	N
41	H	H					H	N	N
42	H	H					H	N	N
43			S	H	H	H	H	N	N
44							H	N	N
45	H	H					H	N	N
46	H	H					H	N	N
47	H	H					H	N	N
48	H	H	S	H			H	N	N
49	H	H			S	H	H	N	N
50	H	H	S	H			H	Y	N

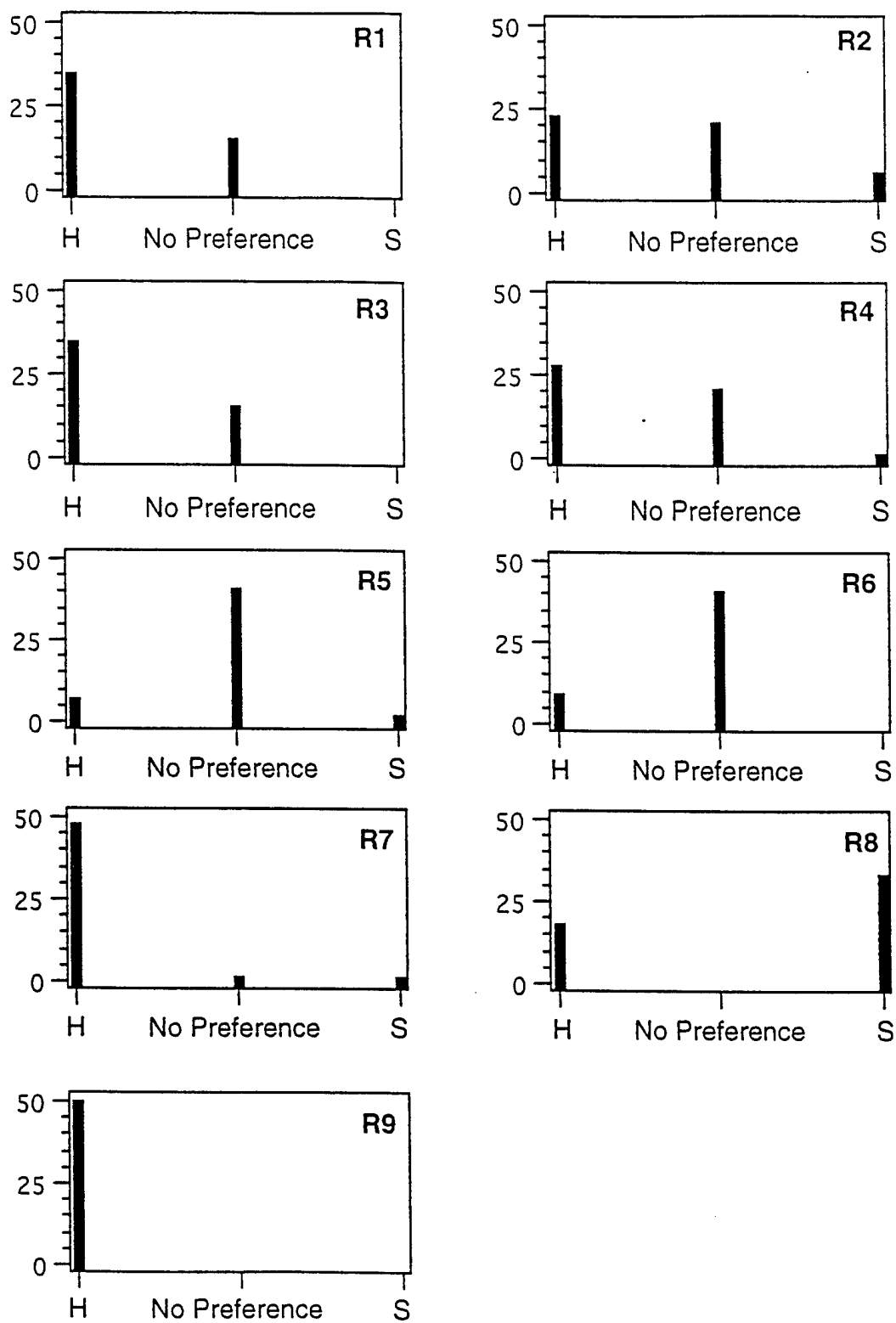


Figure 3.1 Questionnaire Response of Radiologist 1.

Table 2.2 Questionnaire Responses of Radiologist 1

Radiologist 2	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
ID #	H/S	H/S	H/S	H/S	H/S	H/S	H/S	Y/N	Y/N
1	H	H	H/S	H/S	H/S	H/S	H	Y	N
2	H	H	H/S	H/S	H/S	H/S	H	N	N
3	H	H	H/S	H/S	H/S	H/S	H	Y	N
4	H/S	H	H/S	neither	S	S	S	Y	N
5	H	H	S	S	S	S	S	N	N
6	H	H	H/S	neither	S	neither	neither	Y	N
7	H	H	H/S	neither	H/S	neither	H	Y	N
8	H	H	neither	neither	neither	neither	H	Y	N
9	H	H	S	neither	S	neither	H	Y	N
10	H	H	S	neither	neither	neither	H	Y	N
11	H	H	H/S	H/S	H/S	H/S	H	Y	N
12	H	H	H/S	neither	H/S	neither	neither	Y	N
13	H	H	S	S	S	S	S	Y	N
14	H	H	neither	neither	neither	neither	H	Y	N
15	H	H	H/S	neither	H/S	neither	H	Y	N
16	H	H	H	H	H	H	H	Y	N
17	H	H	H/S	neither	H/S	neither	H	Y	N
18	H	H	H/S	H/S	H/S	H/S	H	Y	N
19	H	H	neither	neither	S	neither	H	Y	N
20	H	H	neither	neither	neither	neither	H	Y	N
21	H	H	neither	neither	neither	neither	H	Y	N
22	H	H	H/S	neither	H/S	neither	H	Y	N
23	H	H	H/S	neither	S	neither	H	Y	N
24	neither	neither	H/S	neither	S	neither	H	N	N
25	H	H	H	neither	H	neither	H	Y	N
26	neither	neither	H/S	neither	H/S	neither	neither	N	N
27	H	H	H/S	neither	H/S	neither	H	Y	N
28	H	H	S	neither	S	neither	neither	Y	N
29	H	H	H/S	neither	H/S	neither	H	Y	N
30	H	H	H/S	H/S	H/S	H/S	H	Y	N
31	H	H	S	S	S	S	S	Y	N
32	H	H	H/S	neither	H/S	neither	H	Y	N
33	H	H	S	S	S	S	S	Y	N
34	H	H	S	neither	S	neither	no	Y	N
35	H	H	H/S	S	H/S	S	S	N	N
36	H	H	H	H	H	H	H	Y	N
37	H	H	neither	neither	neither	neither	H	Y	N
38	H	H	S	S	S	S	both	Y	N
39	H	H	H/S	neither	H/S	neither	H	Y	N
40	H	H	H/S	neither	H/S	neither	H	Y	N
41	H	H	H/S	S	H/S	S	S	Y	N
42	H	H	H	H	H	H	H	Y	N
43	H	H	H	H	H	H	H	Y	N
44	H	H	H/S	neither	H/S	neither	H	Y	N
45	H	H	H/S	neither	H/S	neither	H	Y	N
46	H	H	H/S	neither	H/S	neither	H	Y	N
47	H	H	neither	neither	neither	neither	H	Y	N
48	H	H	neither	neither	neither	neither	H	Y	N
49	H	H	S	S	S	S	S	Y	N
50	H	H	H/S	neither	H/S	neither	H	Y	N

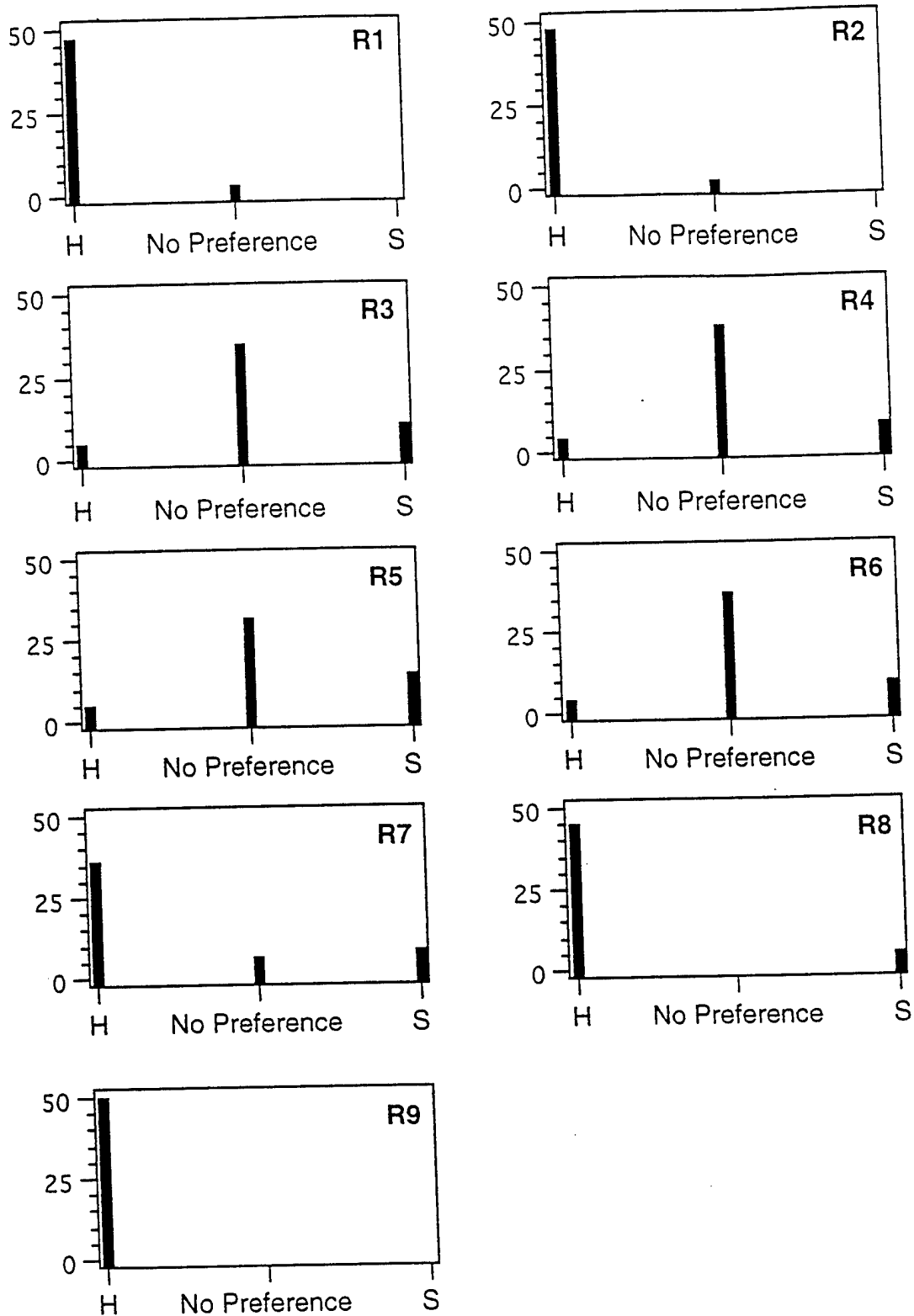


Figure 3.2 Questionnaire Response of Radiologist 2.

Table 2.3 Questionnaire Responses of Radiologist 3

Radiologist 3	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
ID #	H/S	H/S	H/S	H/S	H/S	H/S	H/S	Y/N	Y/N
1	H	H	H	H	H	H	H	Y	N
2	H	H	H	H	H	H	H	Y	N
3	H	H	H	H	H	H	H	N	N
4	H	H	H	H	H	H	H	Y	N
5	H	H	H	H	H	H	H	Y	N
6	H	H	H	H	H	H	H	N	N
7	H	H			H	H	H	Y	N
8	H	H			H	H	H	Y	N
9	H	H			H	H	H	Y	N
10	H	H	H	H			H	Y	N
11	H	H	H	H	H	H	H	Y	N
12			H	H	H	H	H	Y	N
13	H	H	H	H	H	H	H	Y	N
14	H	H	H	H	H	H	H	Y	N
15	H	H	H	H			H	Y	N
16	H	H			H	H	H	Y	N
17	H	H	H	H			H	Y	N
18	H	H	H	H	H	H	H	Y	N
19	H	H			H	H	H	Y	N
20	H	H			H	H	H	Y	N
21	H	H			H	H	H	Y	N
22	H	H			H	H	H	Y	N
23	H	H	H	H	H	H	H	Y	N
24	H	H	H	H			H	Y	N
25	H	H	H	H			H	N	N
26	H	H	H	H	H	H	H	N	N
27	H	H	H	H	H	H	H	N	N
28	H	H	H	H	H	H	H	Y	N
29			H	H			H	Y	N
30			H	H			H	N	N
31	H	H	H	H	H	H	H	Y	N
32	H	H	H	H			H	N	N
33	H	H	H	H	H	H	H	N	N
34	H	H					H	Y	N
35	H	H	H	H			H	N	N
36	H	H	H	H			H	Y	N
37	H	H			H	H	H	Y	N
38	H	H	H	H	H	H	H	Y	N
39	H	H			H	H	H	N	N
40	H	H	H	H			H	Y	N
41	H	H	H	H	H	H	H	Y	N
42	H	H	H	H	H	H	H	N	N
43	H	H	H	H	H	H	H	Y	N
44					H	H	H	N	N
45	H	H	H	H	H	H	H	Y	N
46	H	H	H	H	H	H	H	Y	N
47	H	H					H	Y	N
48	H	H			H	H	H	Y	N
49	H	H	H	H			H	N	N
50	H	H			H	H	H	N	N

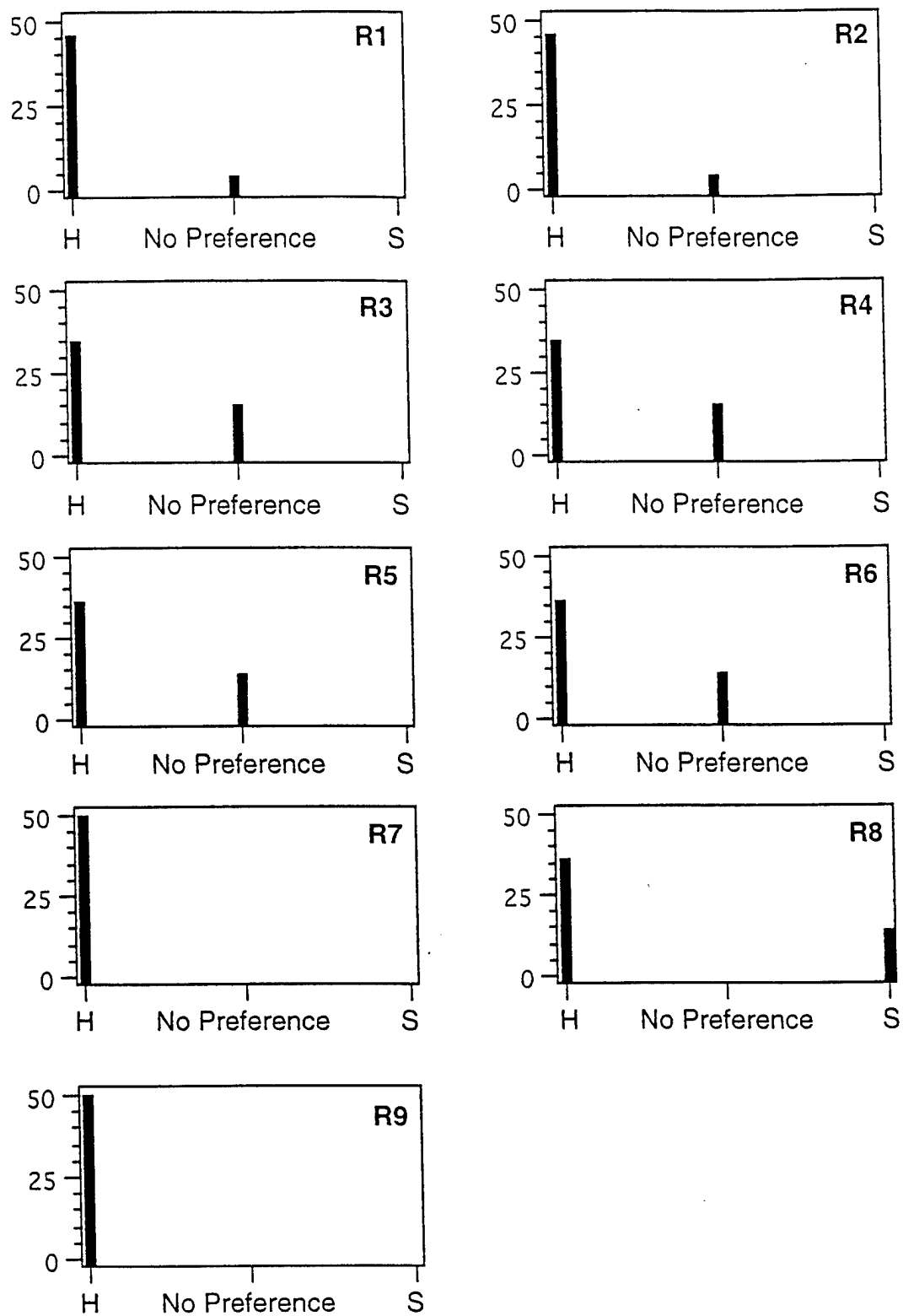


Figure 3.3 Questionnaire Response of Radiologist 3.

Table 2.4 Questionnaire Responses of Radiologist 4

Radiologist 4	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
ID #	H/S	H/S	H/S	H/S	H/S	H/S	H/S	Y/N	Y/N
1	H	H	H/S	H/S	H/S	H	H	Y	N
2	H	H	H	H	H/S	H/S	H	N	N
3	H/S	H/S	H/S	H			H	N	N
4	H	H	H/S	H/S			H	N	N
5	H/S	H/S	H/S	H/S			H	N	N
6	H/S	H/S			H/S	H/S	H	N	N
7	H/S	H/S			H/S	H/S	H	N	N
8	H/S	H/S			H/S	H/S	H	N	N
9	H	H	H	H	H	H	H	Y	N
10	H	H					H	N	N
11	H	H	H/S	H/S	H/S	H	H	Y	N
12	H	H	H	H			H	N	N
13			H/S	H			H	N	N
14	H	H	H/S	H			H	Y	N
15			H/S	H/S			H	N	N
16	H	H	H	H			H	Y	N
17	H	H					H	Y	N
18	H	H	H/S	H/S			H	Y	N
19	H	H			H/S	H	H	N	N
20	H/S	H					H	N	N
21	H	H			H/S	H/S	H	Y	N
22	H	H			H/S	H	H	Y	N
23	H	H					H	Y	N
24			H/S	H/S			H	N	N
25	H/S	H/S	H/S	H/S			H	N	N
26			H/S	H			H	Y	N
27	H	H	H/S	H			H	Y	N
28	H	H	H/S	H/S	H/S	H/S	H	Y	N
29			H/S	H/S			H	N	N
30	H	H	H/S	H/S			H	N	N
31	H/S	H/S	H/S	H/S	H/S	H/S	H	N	N
32			H/S	H/S			H	N	N
33			H	H			H	Y	N
34	H	H/S					H	N	N
35	H/S	H/S	H/S	H/S			H	N	N
36	H	H					H	N	N
37	H	H			H	H	H	Y	N
38	H/S	H/S	H/S	H/S	H/S	H/S	H	N	N
39	H	H			H	H	H	Y	N
40	H	H			H/S	H/S	H	Y	N
41	H	H	H/S	H	H/S	H	H	Y	N
42	H	H	H/S	H/S	H/S	H/S	H	Y	N
43	H	H/S	H/S	H/S	H/S	H/S	H	N	N
44							H	Y	N
45	H	H	H/S	H/S			H	Y	N
46	H/S	H/S	H/S	H/S			H	N	N
47	H/S	H/S					H	N	N
48	H	H	H	H	H/S	H/S	H	N	N
49	H	H			H/S	H/S	H	Y	N
50	H	H	H/S	H/S			H	Y	N

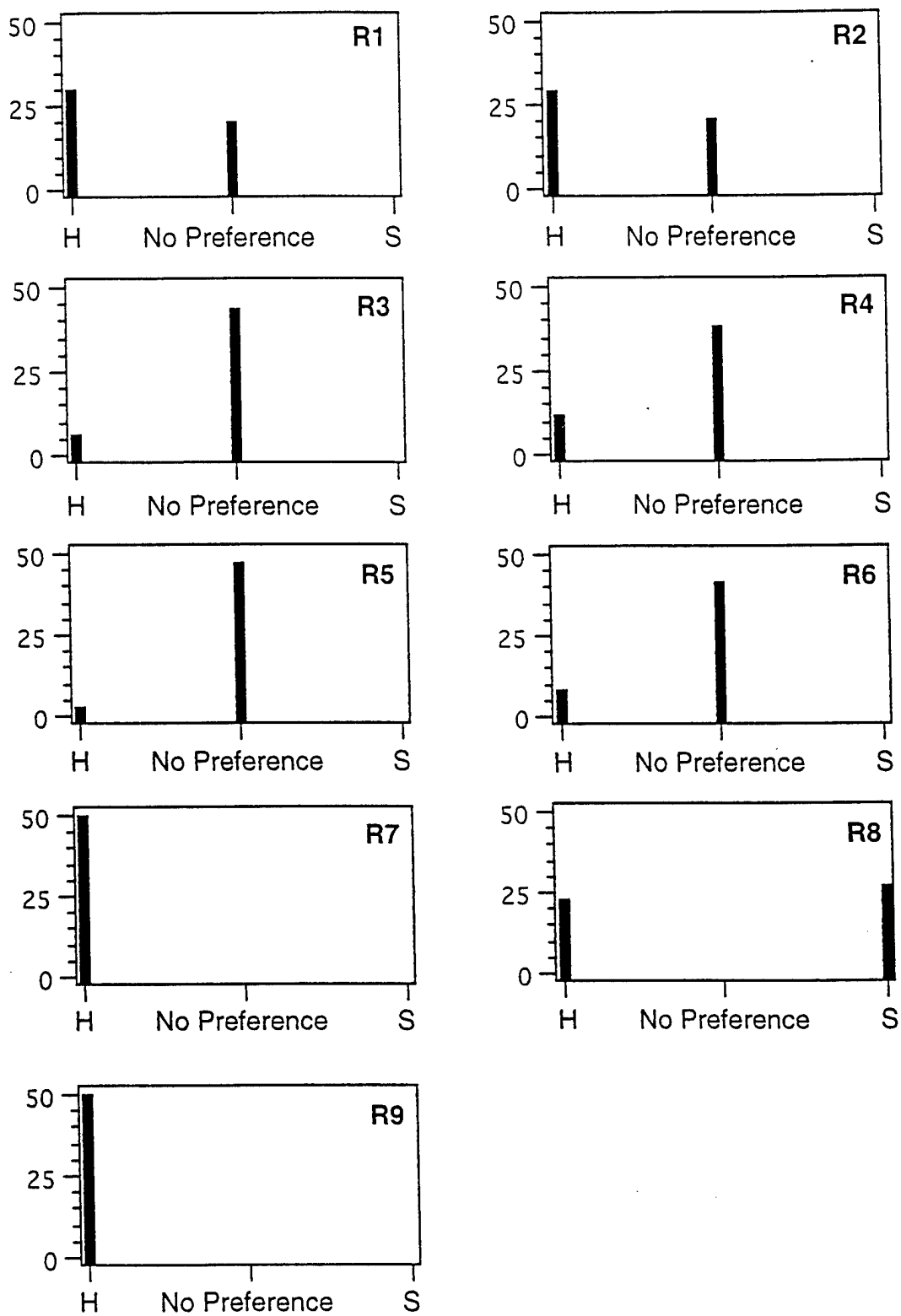


Figure 3.4 Questionnaire Response of Radiologist 4.

Table 2.5 Questionnaire Response of Radiologist 5

Radiologist 5	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
ID #	H/S	H/S	H/S	H/S	H/S	H/S	H/S	Y/N	Y/N
1	H	H	H	H	H	H	H	Y	N
2	H	H	H	H			H	N	N
3			H	H	H	H	H	N	N
4	H	H	H	H	H	H	H	N	N
5	H	H	H	H			H	N	N
6	H	H	H	H			H	N	N
7	H	H	H	H	H	H	H	Y	N
8	H	H					H	Y	N
9					H	H	H	N	N
10	H	H	H	H			H	N	N
11	H	H	H	H	H	H	H	N	N
12							H	N	N
13							H	N	N
14	H	H	H	H	H	H	H	Y	N
15	H	H					H	N	N
16	H	H	H	H			H	Y	N
17	H	H	H	H			H	N	N
18	H	H					H	Y	N
19	H	H					H	N	N
20	H	H	H	H	H	H	H	N	N
21	H	H	H	H	H	H	H	Y	N
22	H	H	H	H	H	H	H	Y	N
23	H	H	H	H	H	H	H	Y	N
24							H	N	N
25			H	H	H	H	H	N	N
26			H	H	H	H	H	N	N
27	H	H					H	Y	N
28	H	H					H	Y	N
29							neither	N	N
30							neither	N	N
31			H	H			H	Y	N
32			H	H			H	N	N
33							neither	N	N
34	H	H					H	N	N
35	H	H					H	Y	N
36	H	H					H	Y	N
37	H	H			H	H	H	Y	N
38	H	H	H	H			H	Y	N
39							neither	N	N
40	H	H					H	Y	N
41	H	H	H	H			H	N	N
42	H	H					H	Y	N
43							neither	N	N
44			H	H	H	H	H	N	N
45	H	H	H	H	H	H	H	N	N
46	H	H					H	N	N
47							H	N	N
48	H	H					H	Y	N
49	H	H	H	H			H	N	N
50	H	H					H	Y	N

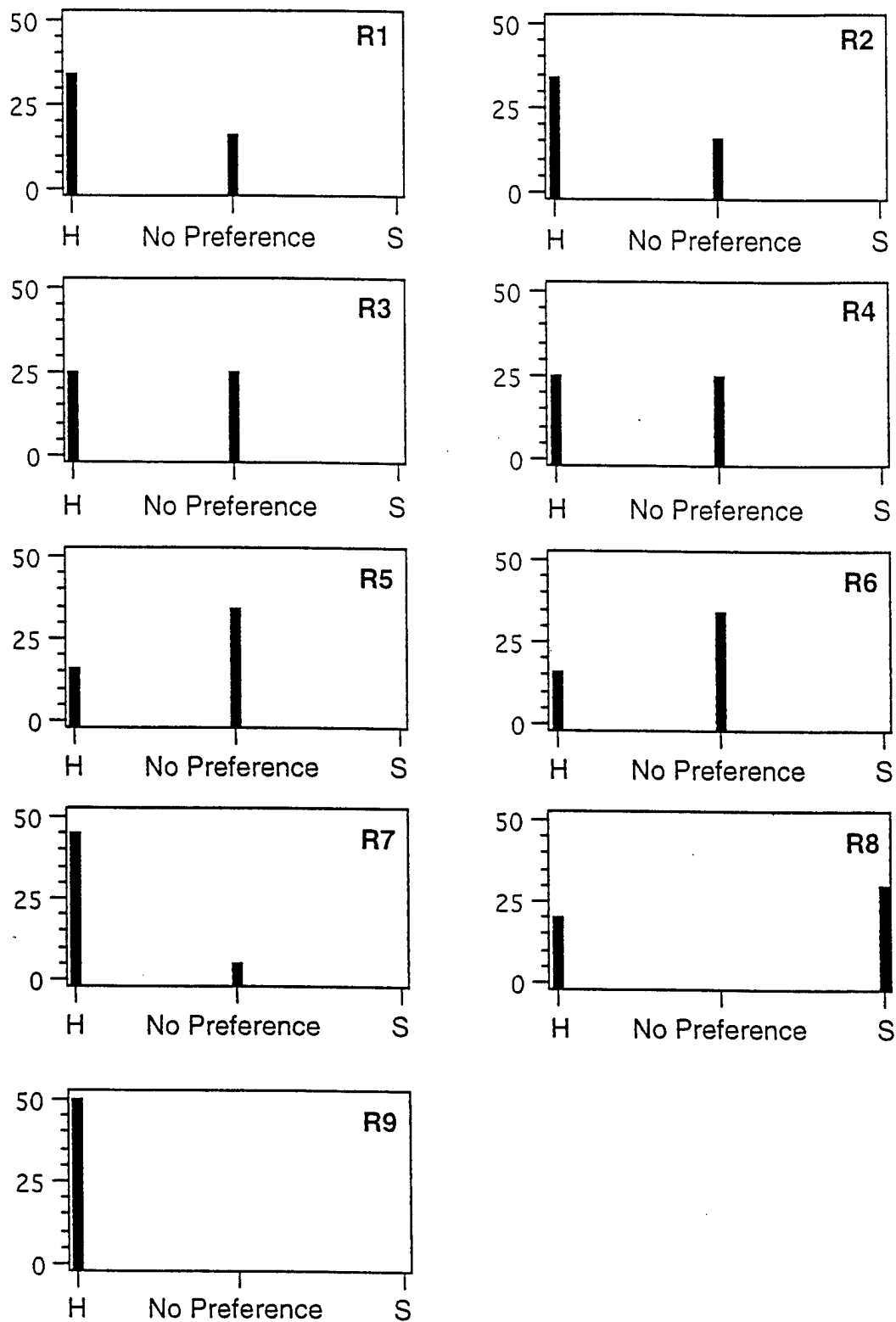


Figure 3.5 Questionnaire Response of Radiologist 5.

5.2. Comments of the Five Radiologists

Dr. Hayes's Comments:

- Prefer the hard copy to soft copy.
- Microcalcifications were difficult to see on soft copies unless the contrast was increased considerably, however, at that time the remainder of the breast is too dark.
- Microcalcifications can not only be identified more easily on hard copy (10 line pairs), they can also be better defined on hard copy as to benign vs. malignant.
- Due to the magnification with soft copies, masses sometimes are more easily detected and better characterized with soft copies.
- On soft copies it was difficult to evaluate the skin without changes in density.
- By changing density on soft copy skin retraction can be better detected than on hard copy without "hot light".
- Can better define number and sizes of calcifications and detect more scattered microcalcifications on hard copy.
- Can not determine adequate number of microcalcifications on soft copy. Microcalcifications are better delineated on hard copy. However, because of the lower resolution on soft copy, some microcalcifications are smudged and can not be seen on soft copy.
- Some easily identified artifacts on the hard copy almost look like microcalcifications on the soft copy. Artifacts are more clearly defined on hard copy.

Dr. Hogge's Comments:

- Calcifications are lost on soft copies in areas of dense tissue - this is a serious limitation.
- Masses are better depicted on soft copies in parts with fatty breasts, but not in dense breasts.

Dr. Freedman's Comments:

- Soft copy naturally makes it harder to see microcalcifications and sometimes they can not be seen at all.
- Masses and asymmetry density are lower central on soft copies and therefore harder to search.

Dr. Green's Comments:

- Manipulation of the window-and-level took time and diverted vision and attention away from the image being diagnosed.
- Some images, especially those containing dense tissues, needed to be enhanced dramatically in the contrast, however, the information in other breast areas is lost.

Dr. Zuurbier's Comments:

- In general, the soft copy images have a "gauze-y" veil-like effect (and resolution).
- Soft copy seems to increase the perception of normal soft tissue structures as abnormal.
- Soft copy loses faint but important microcalcifications in dense tissue.
- Hard copy provides better sense of the contrast of whole image, normal and abnormal soft tissues.
- See detail of dense soft tissue better on hard copy.

6. CONCLUSIONS

- **Image Spatial Resolution** - Our study results showed that, in general, the five radiologists preferred hard copy to soft copy screen films. It also showed that the disease patterns are better characterized on hard copy than on soft copy SF. Because of the spatial resolution of the soft copy, some subtle microcalcifications in the dense tissue area are not well characterized and become difficult to detect. Furthermore, some obvious film defects on the films become microcalcification-like on the soft copy where spatial resolution and contrast become lower than those on the hard copy. Specifically, the screen film hard copy has a spatial resolution of approximately 10 line pairs per millimeter while the digitized 100 μm SF (i.e., soft copy) has only about 5 line pairs per millimeter. The results indicate that high spatial resolution is important in the identification of some subtle and faint microcalcifications in the dense breast area. Direction digitization of screen film at 100 μm and without other image enhancement is not adequate to retain the important clinical information of the microcalcifications, especially those are small and subtle microcalcifications which lie in the dense breast areas. The radiologists experience and training in reading soft copy may have some effects on the study results.
- **Classification of Benign and Malignant Diseases** - The classification of benign and malignant disease patterns, such as microcalcifications and masses, presented a very challenging task for the radiologists. Our study results showed that the five radiologists achieved average 50% accuracy in the classification of benign and malignant disease patterns. The five radiologists found single or multiple disease patterns in all 50 cases (see also the "MAMMO FINDINGS" in Table 1), however, only 25 cases were biopsy proven cases. The 50% accuracy on cancer case was measured no matter what the disease - microcalcification, mass, or architecture distortion. The study of using artificial neural network (ANN) and computer-aided diagnosis (CADx) to classify benign and malignant microcalcifications is currently under intensive investigation at ISIS Center.
- **Display Workstation** - The Vicom display workstation which provides 8 bit (256 gray levels) display is not sufficient to the contrast information on the films as perceived by the radiologists. The contrast and intensity of the abnormal soft tissues as opposed to that of normal soft tissues are better perceived on the hard copy; the abnormal and normal soft tissues of the soft copy have similar contrast and intensity when displayed on the monitors. In some cases, masses and architecture asymmetry were easier to detect on the soft copy partly because of the magnification when displayed on the display monitors. However, the high contrast on the hard copy make the masses better characterized than on soft copy.
- **Periodic Maintenance and Quality Control** - Periodic maintenance of film digitizer and QC on Fuji CR9000 and display monitors are necessary in digital mammography of the MDIS environment. We are in the process of acquiring a multi-format pattern generator and a comprehensive display evaluation system for QC on high resolution display monitors.
- **On-going Researches** - The results of this study indicate that, based on our current experimental setting, soft copy display is not good enough to be clinically useful. However,

improvement and optimization of display equipment and advanced image processing, such as the unsharp masking technique provided by Fuji CR systems, may make soft copy display more clinically useful. Two possible research directions: one is to apply image processing to enhance microcalcifications especially in the dense breast areas of 100 micron images, the other is to use 50 mm images and a higher gray level (10 or 12 bit) instead of an 8 bit display monitor.

- The display of the full image data of a mammogram digitized at 50 micron resolution (4096 pixels $\times$ 5120 lines $\times$ 12 bits) may require roaming of the image on the 2048 pixels $\times$ 2560 lines display monitor. The potential effects, such as the diversion of vision and attention away from the region-of-interest being viewed, caused by roaming and other image manipulations need to be further investigated.

- We have developed a region-based image processing technique to enhance the visibility of subtle microcalcifications in the dense tissues. The comparative study of SF and the processed 100 micron SF images is currently under development.

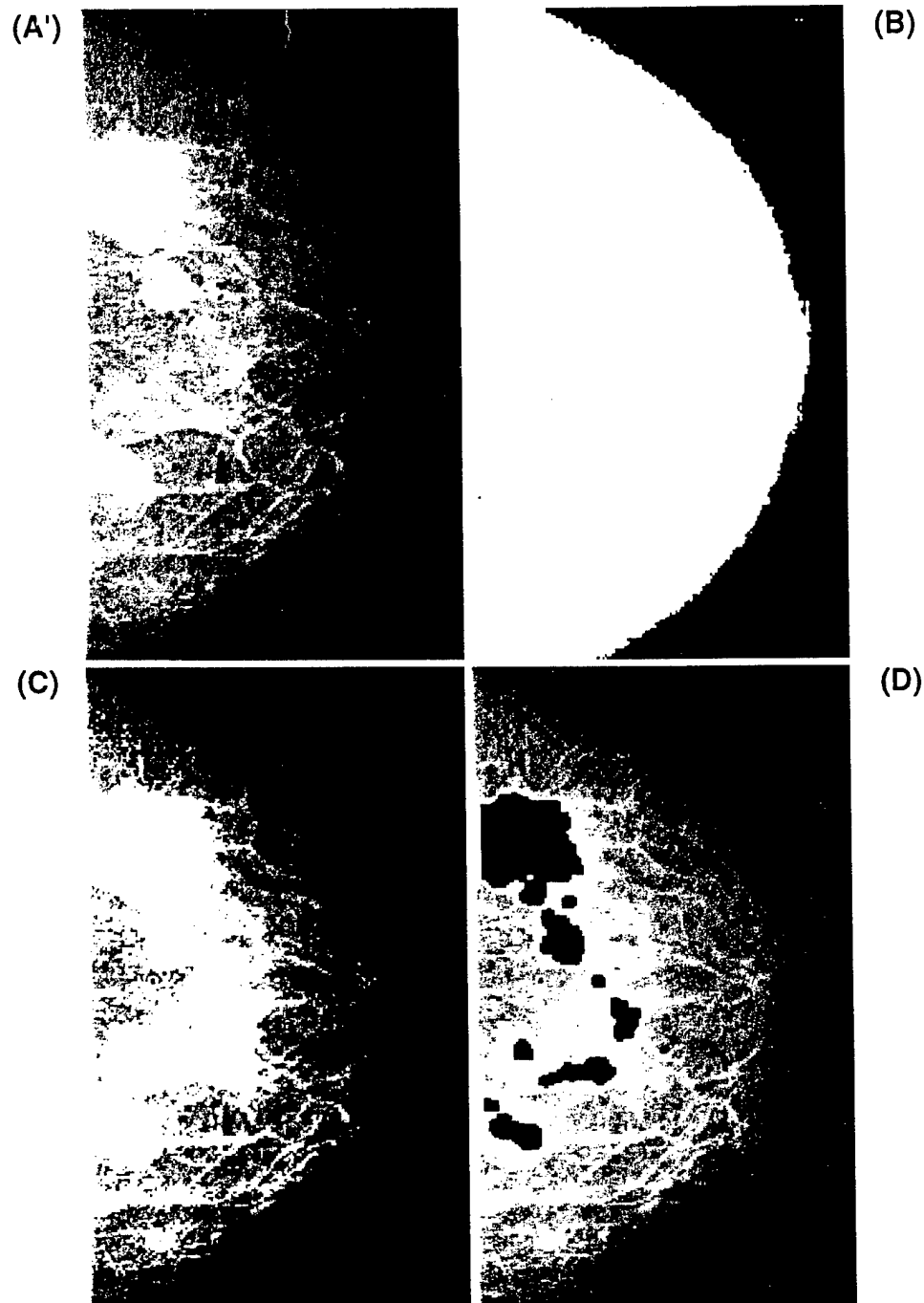
7. ADDENDA

7.1. Acronym / Symbol Definition

ACR -	American College of Radiology
ACR-NEMA -	ACR-National Electrical Manufacturers Association
CDMAM -	Contrast detail mammogram-phantom
CIRS -	Computerized Imaging Reference Systems
CR -	Computed radiography
DASM -	Data acquisition and system management
GUMC -	Georgetown University Medical Center
LCC -	Left cranio caudal
LMLO -	Left mediolateral oblique
MDIS -	Military diagnostic imaging systems
QC -	Quality control
RCC -	Right cranio caudal
RMLO -	Right mediolateral oblique
SCIB -	Second channel interface board
SCSI -	Small computer system interface
SF -	Screen film
SMPTE -	Society of motion picture and television engineers

7.2. References

- [1] Lumisys, *Lumiscan Service Manual: Model 100/150/200*. Sunnyvale, CA 94086, 1994.
- [2] C. E. Metz, "Some practical issues of experimental design and data analysis in radiological ROC studies," *Investigative Radiology*, vol. 24, pp. 234-245, 1989.



The result of detecting the regions of suspicious masses based on enhanced mammogram.

- (A') enhanced mammogram;
- (B) breast boundary;
- (C) enhancement by segmentation;
- (D) suspicious masses extraction.

Quality Control of Storage Phosphor Digital Radiography Systems

Matthew Freedman, Dot Steller, Hamid Jafroudi, and Seong Ki Mun

Quality control (QC) of storage phosphor devices is important in assuring that the image information entered into an Image management and communication (IMAC) system is sufficient for diagnosis. QC of storage phosphor digital radiography systems is complex because of the self-corrective nature of the image-processing software used in these machines. Currently, one must produce hard copy to perform adequate QC. Inspection of images with reject analysis and inspection of cassettes and imaging plates has helped us in our QC program. For those QC tests using control limits, the appropriate settings for these limits are unknown. Starting approximations are given. Recommended tests are described.

Copyright © 1995 by W.B. Saunders Company

KEY WORD: Digital radiography quality.

DIGITAL PROJECTION radiography is an important method for introducing radiographic information into image management and communication (IMAC) systems. There are four methods of digital radiography currently used in medicine and dentistry. These are based on storage phosphors, charge coupled devices, selenium receptors, and image intensifier tubes. This paper will discuss the quality control (QC) of devices using storage phosphor technology based on the author's experience with Fuji and Agfa systems. There have been several prior reports on QC of storage phosphor digital radiography machines.^{1,2} At this time, the authors use hard-copy output for performing QC of storage phosphor projection radiography devices. In this report, the term "computed radiography" (CR) is used to apply to devices using storage phosphor technology.

CR has introduced new requirements for quality control that are different from those needed with other imaging technologies. The balancing of optical density that occurs with CR systems means that it is more difficult for the radiologist to detect malfunction of the equipment by simple inspection of the image: a more complex system is necessary.

The image processing software of CR devices results in a robust system designed to internally correct errors in exposure. The mechanisms installed to correct errors in exposure also function quite effectively in minimizing the

effect on the final image quality of many internal errors of machine functioning that could occur. Therefore, in clinical practice one tends to see total failure of components rather than progressive deterioration of image quality. The manufacturers of the machines we have tested provide only limited supporting functions for doing QC of the internal functioning of the machine.

The QC program we use is derived from our 30 months of experience with an Agfa Diagnostic Center (ADC; Mortsels, Belgium) prototype, 49 months of experience with a Fuji (Tokyo, Japan) AC-1 (modified to an AC-1+ with later modification to their high-resolution system) and 10 months experience with a Fuji 9000. Approximately 100,000 Fuji CR images are now obtained annually. Information was also acquired from manuals provided by Agfa and Fuji and from attendance at the Fuji course used for training their service personnel.

The most frequent errors detected by our QC program are based on human factors rather than machine factors.

In this report, the authors will be describing simple QC procedures that can be used to assess drift from preselected standards of performance. These standards should be first measured after the machine has been initially calibrated by the manufacturer and undergone acceptance testing. A longer document on acceptance testing and QC is available from the authors.³ It is assumed that the readers are well acquainted with the meaning of the image-processing factors on the Agfa or Fuji machine they operate. Additional information on optimizing image appearance is available in the literature.^{4,5} This report first discusses the types of failure and then the method of detecting fail-

From the Department of Radiology, Division of Imaging Science and Information Systems, Georgetown University School of Medicine, Washington, DC.

Address reprint requests to Matthew Freedman, MD, Department of Radiology, Georgetown University Medical Center, 3800 Reservoir Rd, NW, Washington, DC 20007.

Copyright © 1995 by W.B. Saunders Company
0897-1889/95/0802-0001\$3.00/0

ures. Suggestions are included for machine design changes in CR devices that might either decrease the incidence of failures or improve their early detectability.

TYPES OF CR FAILURES

The purpose of a QC program is to find systematic or generalizable failures early enough so that their effect on overall performance is minimal. The process combines a search for minor deviations from standards that, if continued, would worsen and a review of specific process failures. The difficulties in QC analysis are in differentiating minor deviations that represent trends toward failure from random variations that are self correcting (The data for doing this are currently lacking) and the appropriate generalization from the specific failure to the process that caused the failure.

Patients vary and, therefore, failure of one image may be caused by patient variability or machine failure. Each component of a machine is designed to have a range of tolerance of performance that depends on engineering factors related to the cost of greater perfection. Therefore, a certain amount of variability is expected.

Failures can result from human failure or machine failure. Certain aspects of the design of the CR machine increase the chance of human failure. Certain human factors increase the chance of machine failure.

HUMAN FAILURE AND IMAGE QUALITY OF CR DEVICES

The most common cause of failed images, in our experience, is technologist error. The three main causes of technologist error are lack of sufficient training in CR, curiosity leading to experimentation, and machine complexity.

Training

CR is not a simple replacement device for screen-film radiography. CR introduces several new important concepts that technologists must learn. For example, in CR, resolution is dependent on image plate size; good image quality requires the correct selection of the body part on the identification (ID) station; two images on the same imaging plate (IP) require exposure

balancing different than in screen film radiography; and post processing is necessary for some images. Collimation methods can affect image quality on the Fuji AC-1 and Agfa ADC prototype, but, apart from very small field sizes, usually do not affect images on the Fuji 9000. Because the machine corrects for errors in exposure, wide variation in exposure can occur and patients may be exposed to excess radiation.⁶ Many failures result from lack of proper training in this new modality. A prior report⁶ shows the long-term beneficial effect of technologist training.

Curiosity Leading to Experimentation

Technologists are curious about the equipment they use and are interested in creating the best image possible. Given the opportunity, some will try to figure out how the machine works by trying different settings of the machine parameters, knowing they can usually save a bad image by post-processing on the workstation. Because of a lack of level of access control on the ID terminal, what was intended as a one-time experiment can be loaded into the ID terminal memory as a permanent change affecting many subsequent images. In our initial 3-month experience with the Fuji 9000, we identified unauthorized technologist-induced changes in image-processing settings 12 times. Because of this, we incorporated into our training program hands-on training on how to properly experiment with the system without causing damage. By enlisting the technologists into the search for better imaging with CR, we both decreased the incidence of unauthorized changes (one in the last 5 months) and received several suggestions for improved image quality that we have added to our routine.

In the ninth month, the Fuji 9000 shut down because the memory card was completely filled. The system adds each new image processing setting to its memory rather than overwriting the prior changes. When the memory is full, the system shuts down requiring reloading of software.

Software controls limiting changes in image processing settings to a few key operators would decrease the unauthorized resetting of the factors.

Machine Complexity

The operation of these machines is complex and can result in errors. For example, the technologist must enter the body part radiographed into the ID terminal. Technologists often do several examinations on the same patient stacking the cassettes and may not remember which cassette represents which body part. In our multiple image trauma studies, ~10% of imaging plates will have the body part misidentified. When used for bedside examinations, the lack of a clip on the cassette to hold the patient identification card results in a misidentification rate of ~1 in 80 images.

When CR is used to enter data into an IMAC system, orientation of the image on the affects whether the final image will be upright; selecting antero-posterior or postero-anterior projection will affect which side the heart is on, etc. The technologist who tries to get a better image through post processing or to correct the effects of misidentification of body parts to create a proper image is limited by lack of knowledge of a complex image-processing system. We have found that hands-on training in image processing allows the technologist to correct most of the errors and increases technologist satisfaction. Machine simplification would be important in decreasing the incidence of technologist error.

MACHINE FAILURE AND DECREASED IMAGE QUALITY IN CR DEVICES

There are five components of the machine that can result in decreased CR image quality in an IMAC system: the cassette, imaging plate, image plate reader, image processing software, and the image display device. We will not discuss the display device.

Self-Correcting Features of CR Image Quality

The designers of CR systems have made the systems largely self-corrective to factors that might result in image-quality degradation. The software that corrects for variation in IP exposure by measuring the exposure and then setting a sensitivity or "S" value to control the final image density will also correct variations among imaging plate x-ray absorption, IP light emission, laser power used in the plate reader,

deterioration of the light guide and photomultiplier device. It will also correct for certain deterioration in the system electronics. Detecting variations in imaging plates or machine function by monitoring the S value in clinical cases is likely to be masked by variability in exposure settings and patient size. In prior work, we found that well trained and supervised technologists still showed a fivefold variability of S numbers on bedside chest radiographs.⁶

A failing CR system that adjusts its sensitivity to correct for gradual component degradation would introduce more noise into the image (the image would have a lower signal-to-noise ratio). However, the self-corrective nature of the CR machine obscures the visibility of the noise through filtering (on the Fuji systems this is the RT factor). The Agfa ADC prototype provides software for measuring the signal-to-noise ratio; the Fuji software does not.

The self-correcting features for image appearance built into the CR machines to make them robust for variations in exposure make it quite difficult to detect gradual machine or IP deterioration based on clinical image output. Therefore, tests of machine function must disable these features before the test.

Cassettes

The cassettes are one of the weak links in CR systems. They need frequent inspection to check on the hinges, ID windows, latches, and lead backing (if lead backed). Failures of each of these have resulted in failures of our Fuji system. One third of our 14- × 17-inch Fuji cassettes placed in the Fuji 9000 failed in the first month because of broken ID windows jamming the machine. The machine jammed once because of detachment of the lead of the lead-backed cassettes. Fuji has redesigned the cassettes. We are testing the recently introduced new model, so far without failures. The Agfa cassettes have not broken.

Imaging Plates

The IPs develop an electrostatic charge over time that attracts dust. Dust can result in white or black spots. Intense white spots on an image result from reflection of the laser beam more directly into the light guide. Black spots in the

image result from local failures in the erasure/activation of the IP: the nonactivated portion of the IP will not record the exposure information on the next exposure. White dots can be confused with microcalcifications in digital mammography and foreign matter in extremity radiographs. (Digital mammography using Fuji CR is in clinical use in Europe and Japan and is in clinical trials in the United States.)

IPs can develop fissures and scratches on their surface. These can be detected on images made from these plates and on direct inspection of the IP surface. These defects cannot be repaired and, if they interfere with diagnosis, the IP should be replaced. On the Fuji system, the ID number of the IP is recorded on the hard-copy image. On the Agfa system, the IP number is matched with the image in the ID station data base. We have had to replace three imaging plates that were damaged in machine jams in the first 3 months of use of the Fuji 9000 and 4 IPs damaged in jams of the AC-1+. Initial cracks in the IPs used interchangeably in the Fuji 9000 and Fuji AC-1 started to appear after 5 months of use. We have replaced 14 of the 14- × 17-inch IPs (two thirds of our IPs of this size), two 10- × 12-inch IPs, and two 18- × 24-cm IPs for cracks.

We have recently received redesigned IPs for use in our Fuji units. The new design should provide greater mechanical strength.

IP Reader

The IP reader contains mechanical transport mechanisms, a laser, a light guide, a photomultiplier device, and IP erasure device. Each of these systems can result in image degradation. The transport mechanism can jam and shut down the machine, dust can interfere with the light guide, and the laser, photomultiplier, and erasure light can slowly deteriorate or rapidly fail. Slow deterioration is difficult to detect because of the image self-correction feature of the software. Dust on the light guide can result in streaks or lines on the final image. Laser-scan position accuracy degradation can result in uneven edges of a sharply defined object.

After the initial 3 months of use of the Fuji 9000 in which jams were frequent and usually caused by the cassette problems described above, we now experience approximately one jam every

2 weeks. The machine can jam from IP or cassette transport errors. Light-guide malalignment was detected once on the Fuji AC-1. The erasure light on the Fuji 9000 has failed once resulting in automatic shutdown of the machine.

The Image-Processing Software

The image-processing software on CR devices is quite complex. Once a machine is calibrated, we have not detected any spontaneous deterioration in it. It is necessary to check that people using the ID terminal have not changed the settings because it is possible to do this inadvertently or as a prank. Three times, our Fuji machine has reverted from our preferred image-processing settings to Fuji's original settings during machine repairs and power outages. We have not had similar problems with the Agfa processing software. A list of the preferred image-processing settings for each body part should be maintained so that if the settings revert, the optimized settings can be reentered. Because this is a particular problem of the Fuji system, it would be helpful if Fuji could provide a memory card to users so they could store their preferred settings and reload them as needed.

ENVIRONMENTAL CAUSES OF DECREASED IMAGE QUALITY

Dust is one of the main causes of image-quality degradation in CR devices. The IPs, through use, become static charged and avidly attract dust. Control of environmental dust is of high priority.

Users of Fuji CR devices have to be trained to always close the door of the unit between use. It would be helpful if this feature were automatic as it is on the Agfa machine.

The light guide on the Fuji 9000 has required cleaning for dust three times in 9 months; however, it is usually cleaned when the machine case is opened for any purpose. Dust is identified in images requiring special cleaning of IPs approximately three times a week. Because the CR system is a closed system, the cassette is opened only inside the machine and thus the dust accumulation on the IPs should occur only within the machine. Because of the amount of dust accumulating inside our Fuji 9000, Fuji service has recommended that their yearly pre-

ventive maintenance be performed at 10 months to allow full cleaning of the machine. The air vents on the machine should be cleaned weekly. We usually remove 2 to 4 cc of dust weekly.

Liquids (especially blood and other proteinaceous body fluids) can permanently damage the imaging plates and possibly damage the light guide resulting in expensive repairs. We train our technologists to enclose the cassette in a cover when it might be contaminated with body fluids.

QUALITY CONTROL TEST RECOMMENDATIONS

There are three methods appropriate for quality control of CR devices: component inspection, reject analysis, and control chart methods. All three methods are important.

Component Inspection

Cassettes and IPs are both subject to failure. IPs should be cleaned weekly and whenever dirt artifacts are found on an image. This is an appropriate time to inspect the cassettes for partial failures of hinges, ID windows, lead backing, and clasps, and to check the imaging plates for scratches, stains, fissures, and cracks. Bad components should be replaced.

Reject Analysis

Reject analysis or repeat analysis collects all images that have been replaced with repeat images and attempts to analyze the nature and cause of the failure. Failures may result from unique problems or generic problems. In CR, an analysis of images resulting from repeat exposures or from post processing should both be reviewed. When hard copy is used, the bad images can be kept for later review. When only soft copy is used, a paper or electronic record should be kept of images requiring repeats. Placing the images in an electronic QC folder would simplify their analysis.

Action should be taken based on the reject analysis appropriate to the problem found. The most frequent problems identified by repeat analysis are related to technologist error: improper positioning, improper exposure factors, incorrect selection of body part on the ID station, incorrect selection of imaging plate size or resolution, and incorrect matching of high-

resolution IPs with appropriate exposure. The action taken in these cases is technologist education.

Reject analysis resulted in our changing some of our preset image processing to accommodate a wider variety of patient sizes.

Control Limit Tests

Quality control limit tests are set to measure a factor repeatedly for possible deviations from a standard. Control limits (or action limits) are set so that action can be taken when the measurement deviates beyond a certain amount from the standard. These control limits can be derived in two ways: first, from recording a large number of readings on machines that are functioning appropriately and using two or three standard deviations from the mean as the control limits or, second, by determining the level at which the deviation from the standard results in some measurable important change in function or output or becomes predictive of impending component failure.

Appropriate factors to measure using control limits include S numbers, image optical density, and signal-to-noise ratio. However, currently, setting clinically relevant control limits on these factors is mainly guess work. We have not detected any incipient machine failures that showed variations in these numbers, nor have we seen failures of a type expected to result in variations in these numbers.

Monitoring of the S Number

The S number is used on the Fuji systems to indicate the amplification used to correct for differences in exposure. There is no Agfa equivalent.

Monitoring of the S number in clinical cases may indicate problems caused by x-ray machine failures or technologist errors. In clinical cases, S variability is much more likely to be caused by exposure and patient-specific differences rather than machine failure.⁶

Monitoring of S numbers resulting from a standard exposure would detect poor-quality imaging plates, deterioration of imaging plates, decreased laser power, photomultiplier deterioration, malalignment of the light guide, and partial or complete failure of the erasure process. Each of these failures would result in less

signal reaching the computer for analysis. The computer would correct for these deviations by increasing the S number to create an image of appropriate optical density for the body part.

There are two complementary methods of testing these image acquisition components: use a standard flat-field exposure in auto mode and follow the S numbers graphically over time, or set a standard S number in the fixed mode and measure the resulting optical densities of the serial images. On the Agfa system, one tests the system by setting the machine in fixed mode and measuring the optical density of the output image on serial images.

Setting the Control Limit on S Numbers

If one chooses to obtain a standard exposure using standard parameters, one should see a variability of S numbers with SD of about 10% because that is the manufacturing limit set for the internal control system of the manufacturer. Thus, one can choose to use two or three SDs as the control limit. In making this decision, one would want knowledge of the clinical importance of variations of S numbers of this magnitude and the predictive value of machine failure of S number variability that exceeds that limit. If one is to accept a 60-point spread from an S of 200, one would want to know whether that magnitude of difference has any clinical significance in diagnosis or any effect on the early detection of machine failure.

Currently we have no information of the predictive value of variations in the S number as an indicator of incipient machine failure. For research purposes, it would be helpful to compile this data from multiple sites and evaluate its predictive value. We are currently compiling this data on our two Fuji machines.

Is it clinically important to limit variation to 60 units from a mean S of 200? Workers in this area would agree that there is some level of exposure below which information is lost.⁷ For chest radiographs (which would normally have an S of 200), an S greater than 500 may result in some loss of information about the locations of tubes in the mediastinum.⁸ In our experience, there is a loss of information about tubes and lines in the mediastinum that occurs between an IP exposure of 0.13 and 0.06 mR. It is likely that the S number required for full information will

differ with different diagnostic tasks. The required IP exposures for most clinical diagnostic tasks are not known. Loss of information would tend to occur in focal areas of the images receiving the least information. Because the "S" number reflects an average exposure over all regions of the IP and information loss is likely to occur only in those regions of the IP receiving the least exposure, it is difficult to determine the relevance of a change in the global S number to diagnostic accuracy.

At the moment, using control limits to an S of 200 of ± 60 units seems reasonable, but without much supporting evidence that it is meaningful.

Setting Control Limits for Optical Density (OD)

As with the S number, there is no evidence that variability of OD in serial images taken with a standard exposure has any predictive value of potential machine failure or serves as an indicator of component degradation.

Variations in (OD) that cause the image to fall more in the toe or shoulder of the contrast curve can affect diagnosis. In the more usual CR operating modes of auto or semi mode, the image processing will automatically correct the optical density so the main effect in most clinical uses of CR will be that differences in exposure affect the noise in the image, rather than its OD. Thus, the clinical meaning of variations in OD occurring during QC studies is unknown.

Measuring Signal-to-Noise Ratio

Signal-to-noise measurements can be made on the Agfa workstation, but not on the Fuji machines. Tracking of this measurement would track performance of multiple components: the IP and components of the plate reader. We are unaware of any data on the predictive value of this for predicting failure of components. Values of SNR affecting diagnosis would vary with the image features of the disease in question. We are not aware of current data that would suggest appropriate control limits for this factor.

QC RECOMMENDATIONS

Initial Control Images

The machine should be properly calibrated by the manufacturer and acceptance testing completed. Then images should be obtained using available phantoms. One should use a

well-calibrated x-ray machine that is expected to be available for many years. Variability will be reduced if the same IP is used for both the initial and follow-up tests.

Images of phantoms. For each available phantom (chest, hand, skull, or whatever is available), obtain two images processed with your standard algorithm settings for that body part. These images should be stored for future comparison with new images obtained with the same exposure and image-processing factors. Although we do not recommend routinely repeating these images, they may prove helpful in confirming that a change in image quality detected on the step wedge test (test B) has occurred.

Images of a step wedge. Images of a step wedge should be obtained at 80 kilovolt peak. On the Fuji machine, the image should be obtained on a 14- × 17-inch cassette set in fixed mode, S, 200; processed with GA, 1; GT, A; GS, 0; GC, 1.0; and RE, 0. RN and RT can be any value. Adjust exposure (mAs) until the step wedge is well exposed with most levels visible. Record exposure factors (kVp, mAs, focal-spot film distance, focal-spot size). Make two such exposures. Print final images. Change to GA, 2; reprint new images. Reset image processing to GA, 1; RN, 5; RT, S; and RE, 10. Reprint new images. Store for comparison with later images.

Scan accuracy image. A scan accuracy image should be obtained. Place a steel ruler in the center of the imaging plate with edges parallel to the cassette hinges (perpendicular to the scan lines). Expose at 80 kVp, 1 mR approximate exposure. Record factors used. Process as *organ name: sensitivity, organ code: 0900, semi mode, density: 1.0*. Wait 10 minutes before processing IP. Store film for reference. Interpretation of this image is done by using a high-power magnifier (8×) to examine the edge of the steel ruler. The edge should appear straight.

Spatial resolution. Using a standard resolution phantom oriented obliquely to the edge of the cassette, obtain an image on a 14- × 17-inch standard imaging plate at 80 kVp at approximately 1 mR. Record exposure factors used. Process as *organ name: sharpness, organ code: 0901, semi mode, density: 1.0*. Store image for future reference. Repeat with a high-resolution imaging plate if used at your site.

S number monitoring. Select a 14- × 17-inch imaging plate. Erase the plate. Use a flat-field exposure at 80 kVp and approximately 1 mR. This should result in an S of approximately 200. This should be monitored weekly using the same IP and same exposure.

ROUTINE QC PROCEDURES

Daily/Weekly

Reject analysis. All rejected images should be inspected by the QC technologist as they are rejected (if possible) or at least daily. A proposed explanation for the image failure should be recorded for each image. The accumulated images over the preceding 4 weeks should be reviewed weekly looking for trends. Action should be taken as trends are detected.

Weekly

Check that the settings of image-processing algorithms have not changed. The commonly used algorithms should be checked weekly. The uncommonly used settings should be checked every 4 weeks (one-fourth done each week). Alternatively, if these factors could be stored on a memory card, one would simply reload the whole set of optimized factors each week without having to check for changes.

Using a standardized flat-field exposure, check the S number as described in "S number monitoring." Graph sequential results. Use control limits of ± 60 units.

Repeat the standard step wedge image as described above and process it with the three indicated image-processing settings. Visually compare the new image with the original. If there is any visible change, measure to confirm the change. If greater than 0.25 OD units in any cell, investigate to find the cause.

Clean and inspect cassettes and imaging plates.

Monthly

Repeat the scan accuracy test as described above. Inspect both edges of ruler with 8× to 20× magnifier to look for unevenness of scan lines.

Repeat resolution test as described above in "spatial resolution." Compare resolution on the original and current images.

REFERENCES

1. Seibert JA: Photostimulable phosphor system acceptance testing. AAPM Summer School, 1991, pp 899-929
2. Freedman M, Mun SK, Pe E, et al: Quality control of storage phosphor imaging devices, in Lemke HU, Inamura K, Jaffe CC, Felix R (eds): *Computer Assisted Radiology*. Berlin, Germany, Springer-Verlag, 1993, pp 456-460
3. Jafroudi H, Steller D, Freedman M: *Computed Radiography Quality Control*. Georgetown University Internal Procedure Document. Washington, DC, Georgetown University Medical School,
4. Freedman M, Mun SK, Pe E, et al: Image Optimization on the Fuji AC-1. SPIE Medical Imaging 1897:480-502, 1993
5. Nelson M, Pe E, Freedman M, et al: Image processing in musculoskeletal digital radiography: A comparison of the AGFA and FUJI bone algorithms. SPIE Medical Imaging 2167:854-867, 1994
6. Freedman M, Pe E, Lo S-CB, et al: Risk of unnecessary radiation overexposure of patients with a storage phosphor plate imaging system. SPIE Medical Imaging 1897:472-479, 1993
7. Busch HP, Georgi M (eds): *Digital Radiography Workshop: Quality Assurance and Radiation Protection*. Mannheim, Germany, May 7-9, 1992, pp 95-96
8. Busch HP, Georgi M (eds): *Digital Radiography Workshop: Quality Assurance and Radiation Protection*. Mannheim, Germany, May 7-9, 1992

Quality Control of Storage Phosphor Radiography Machines

M. Freedman M.D., M.B.A, Seong Ki Mun PhD, Einar Pe RTR, John Weiser PhD,
John Romlein, S.-C. Benedict Lo PhD, Martha Nelson M.D.
Georgetown University, Washington, D.C., U.S.A.
MDIS Project Office, Ft. Detrick, MD., USA

INTRODUCTION:

Procedures for assuring the quality of image generation by Storage Phosphor Radiography (SR) machines have not been defined by the manufacturer and do not appear to have been described elsewhere. The information given in this article is based on experiences with several Fuji AC-1, several Fuji 7000, and one AGFA ADC. AGFA has designed a test phantom for SR that will be available in the future. The phantom tests geometric consistency and gray scale features of SR.

To assure that all components of the machine are functioning correctly global methods for testing the machine need to be devised, recommended time intervals suggested, and methods for testing individual components need to be developed. Based on our combined experience, the following methods are recommended.

TESTING THE MACHINE FOR PROPER INITIAL CALIBRATION:

The Fuji machines, as delivered and released for clinical use were found not to be intercalibrated. Thus images obtained with the same exposure and same image processing routines had visibly different appearances. The following procedure is recommended for initial checks of calibration:

- a. take two imaging plates for each machine you wish to test.
- b. erase each plate
- c. If available radiograph the AGFA SR phantom, if not place a standard step wedge on each plate. Expose each plate to a standardized, approximately 200 speed exposure. Do not collimate. Be certain that the orientation of the plate and the stepwedge is the same for all exposures. If not using a grid, use the longest possible tube-film distance to minimize the heel effect.
- d. Process two plates on each machine. Use the 2 algorithm settings that you expect to use most often (e.g. PA Chest, foot). Use algorithms that are set in the Fuji Auto mode. Use standard processing on the AGFA machine rather than DAM processing
- e. Measure the optical density (OD) of each step on the plate. Graph these findings.

- f. For each body part selected, the curves of the step wedge should be identical. A variation of up to 5 % may be acceptable.

This test procedure tests the proper function of the image plate reader, the auto mode PRIEF for the body part, the consistency of the Fuji G factors across machines, the settings of the laser camera, and the consistency of function of the film processor.

If the machines pass the tests they are appropriately cross calibrated for the two algorithm settings tested. One should test for cross calibration on all algorithms commonly used.

If the machines fail the test, then tests of individual components will be necessary. These are described below.

GLOBAL TEST OF THE CALIBRATION OF THE FUJI G FACTORS:

These tests can be used to test the calibration of the image processing G factors, testing either the machine calibration or the calibration of a workstation that controls a laser print image.

These tests may or may not be necessary. Because they are time consuming, they are probably not necessary if the first global test demonstrates correct cross calibration, but may be necessary if clinical use suggests that there are problems with cross calibration.

TEST OF THE GS FACTOR: The GS factor is a factor that shifts the gradient curve up or down by the number of OD units represented by the value of the GS factor.

- a. set the machine to process the image with a GT of A and a GS of 0.
- b. Take a standardized image of a stepwedge or of the AGFA phantom.
- c. Process the image
- d. Measure and graph the step wedge OD values
- e. Reset the machine with all factors remaining the same, except that the GS factor is set at 0.5. Repeat the above process.
- f. a graph of both groups of stepwedge values should show that they are 0.5 OD units apart.

TEST OF THE GA FACTOR: The GA factor is the slope of the characteristic curve of the system

- a. Set the machine to process the image with a GS of 0, a GT of A and a GA of 1, GC of 1
- b. Obtain a standardized image of a stepwedge or of the AGFA phantom.
- c. Process the image

- d. Repeat with GA factors of 2, and then with GA factor of 3.
- e. Measure and graph the OD measurements of the image of the stepwedge.
- f. Measure the slope of the mid portion of the characteristic curve.
- g. The relative slopes of the three curves should be that the curve with a GA of 2 is twice as steep as that with a GA curve of 1. The slope of the curve with a GA of 3 should be 3 times as steep as that with a GA of 1.

TEST OF THE GC VALUE: The GC value is the rotation point around which the GA factor rotates the gradient curve.

- a. Use the same initial data set as use for the determination of GA calibration given above.
- b. Set the machine to a GC value of 1.5, a GA of 1, a GT of A and a GS of 0.
- c. Expose the stepwedge or AGFA phantom to the standard exposure
- d. Reset the machine so that the GC value is 1.5 and a GA of 2, with all other values remaining the same.
- e. Take an exposure of the standard stepwedge with the standard exposure.
- f. Measure and graph the values.
- g. The three different settings of GA set above (with a GC of 1) should cross approximately at an OD of 1. The two settings with the GC of 1.5 and GAs of 1 and 2 should cross at approximately OD 1.5.

TESTS OF CALIBRATION OF THE R FACTORS:

Tests of the calibration of the R factors require the calculations of the modulation transfer function (MTF) or square wave function and a comparison of these values for several different settings of the R factors. Currently, no standardized numbers for the MTF values are available. Tests of R factor intercalibration are not yet devised.

TESTS OF INDIVIDUAL COMPONENTS:

TESTS OF THE IMAGING PLATES:

The imaging plates should be checked for surface defects and for uniformity.

Checks for surface defects. The imaging plates should be examined under white and ultraviolet light for any scratches or other surface unevenness. In some cases the ultraviolet light will demonstrate defects that are not detected in white light.

Checking for uniformity. This is performed by looking at the light emission that occurs from an imaging plate when exposed to radiation. The ST III Fuji imaging plate corresponds to a light output similar to a single 400 speed screen. The uniformity of light output can be tested by placing a single sheet of 400 speed film

and a sheet of foam in a cassette against the imaging plate and exposing the plate to a 200 speed exposure. The film is then processed and visual inspection and measurements should show a uniform pattern of optical density apart from any heel effect that may be present.

TESTS OF IMAGING PLATE ERASURE:

Following a standard (50 to 400 speed) exposure of a stepwedge or other phantom, the imaging plate should undergo Fuji secondary erasure (or standard erasure on the AGFA unit). A second 400 speed flat field exposure should be made and the plate processed. Any residual image that appears on a subsequent image would indicate incomplete erasure. If the imaging plate is processed the second time with no additional flat field exposure, then the machine would amplify any small amount of residual information and still create an image whether or not the erasure system is working.

TESTS OF LASER CAMERA:

This is described in the Fuji applications manual, and this description is not repeated here.

TESTS OF THE FILM PROCESSOR:

Testing the film processor is normally done by using a test strip produced by the laser camera. The Fuji recommendation for correcting problems in OD on this computer generated sensitometry strip is to adjust the power of the laser camera. Up to a point, this is acceptable. Beyond some point this procedure becomes unacceptable because, even if you correct processor problems in OD by overexposing or underexposing the film, the noise level on the film will be increased, either from too high a processor temperature or by bad processor chemistry. For this reason, when more than minimal deviations from the OD are found on the computer generated laser printed sensitometry strip, it is suggested that a film sensitometry strip be run through the processor.

The film sensitometry strip: A film sensitometry strip can be made by exposing the laser film to 20 flashes of a blue or green sensitometer. The film is then placed as the top film in the film holder. If a computer generated sensitometry strip is then run, the standardized light sensitometry strip will appear on its edge. This can then be compared to a standard strip obtained when it is known that the processor chemistry and temperature are correct. If one uses the same box of film for sensitometry, then this procedure would also detect problems in film consistency.

CHECKING THE ALGORITHM SETTINGS:

One of the most common problems encountered in the use of the Fuji machines is the tendency for the technologists to experiment with the image processing settings. It is important to check these settings at fairly frequent intervals to be certain they have not been changed.

A RECOMMENDED INTERVAL FOR CHECKING THE STORAGE PHOSPHOR IMAGING MACHINES:

Initial checks of machine intercalibration should be done when the machines are first delivered and also whenever major repairs are made. Checks of the imaging plates should also be done for each new imaging plate. Once the machines are in operation, the incidence of machine problems has been low. In approximate order of frequency the following types of problems have been detected: 1. Changes in image processing settings., 2. Dirt or scratches of the imaging plate. 3. Problems related to the film processor. 4. Incomplete erasure caused by inadvertent overexposure of the image plate. 5. General system failure.

RECOMMENDED INTERVAL FOR QUALITY CONTROL PROCEDURES:

- CONTINUOUS: Inspection of images as made for obvious machine failure.
- DAILY: Run computer generated strip to check on film processor.
- WEEKLY: Check image processing settings for all commonly used body parts and 1/4 of the less commonly body parts.
Clean exteriors of cassettes.
- MONTHLY: Clean and check image plates.
Do global check of stepwedge or AGFA phantom image and compare to baseline image.
- NEW FILM LOT: Run sensitometry strip with old and new film.
- MAJOR REPAIRS: Run global test for machine intercalibration as described above.

SUMMARY:

Quality control problems of storage phosphor imaging systems have been mainly related to lack of intercalibration of the machines at the time of initial delivery. Once intercalibration is accomplished, the major problems have been related to unauthorized changes in the imaging processing settings, scratches and dirt on the imaging plates, and problems with the film processor. Routine quality control checks should be performed at various intervals as listed above. Continuous inspection of images should be made by a supervising technologist for the sporadic failures that can occur.

Automated Detection of Skin Thickening in Mammograms

U. BICK, M. L. GIGER, Z. HUO, R. A. SCHMIDT, K. DOI, R. M. NISHIKAWA, C. J. VYBORNY

Kurt Rossmann Laboratories for Radiologic Image Research,
Department of Radiology, The University of Chicago
Chicago, IL 60637, USA

Summary

As part of a comprehensive computer-aided diagnosis scheme for detection of breast abnormalities in mammograms, we are developing a computer algorithm to detect and measure skin thickening in mammograms. The outside breast contour is located based on local gradient operations and adaptive gray-value thresholding. Following the determination of the internal skin line as a local gradient minimum, the skin thickness is measured perpendicular to the skin surface. The software was written in C and Fortran and is implemented both on a DEC VAXstation 3500 and a IBM Powerstation (RISC 6000 Series 560). As input, digitized mammograms 2k x 2k in size with a pixel size of 0.1 x 0.1 mm² and 10 bit grayscale were used.

The developed computer algorithm was shown to be able to automatically measure breast skin thickness in digitized mammograms. Both focal areas of skin thickening as well as generalized skin thickening were detected. Adequate digitization proved to be an important prerequisite for successful operation, since an insufficient dynamic range of digitization may result in loss of the external skin surface.

The described algorithm for detection of skin thickening in digital mammograms shows potential for use in our "intelligent" workstation for the computer-assisted analysis of mammograms, which currently includes the detection of mass lesions and clustered microcalcifications.

We are developing computer-aided diagnosis schemes to assist radiologists in breast cancer diagnosis on mammograms [1,2,7,8]. Although breast skin thickening may occur in a variety of benign disorders like edema, inflammation or scarring, it can also indicate underlying malignant disease and may be the only mammographic sign of an inflammatory carcinoma [4-6]. The purpose of our study is to develop a computer algorithm to automatically detect and measure skin thickening in digital mammograms.

Material and Methods

As input to the skin detection algorithm, conventional screen-film mammograms were digitized (Konica laser film digitizer LD 4500) with 10 bit quantization and a pixel size of 0.1 x 0.1 mm² yielding a matrix size of 2048x2580 for a 8" x 10" film. The dynamic

range for digitization was set to include 0 - 3.5 optical density. The software was written in C and FORTRAN and can operate both under VMS (DEC VAX 3500 workstation) and UNIX (IBM Powerstation, RISC 6000 Series 560).

The skin detection algorithm consists of the following steps:

- (1) Initial coarse segmentation of the image is based on a global histogram analysis establishing mean values for dark (direct exposure) and white (no exposure) image areas.
- (2) Determination of the breast region is achieved using adaptive gray-level thresholding based on the local pixel intensity of the direct exposure background.
- (3) Potential internal skinline points are identified as a local gradient minimum within a certain distance from the outside breast contour (Fig. 1).
- (4) An optimal track through these points is found using an energy function based on connectivity and distance from outside breast contour.
- (5) Skin thickness is measured perpendicular to the outside breast contour.

As a preliminary evaluation, two expert mammographers (R.A.S., C.J.V.) were asked to mark the external and internal skin border in five mammograms with skin thickening ranging between 4 mm and 2.2 cm. To assess the accuracy of the computer algorithm, the distance between each point marked by the radiologists and the computer output was calculated.

Results

The program was shown to be able to automatically detect and measure skin thickening in digitized mammograms. Both focal as well as generalized thickening was detected by the program. In the preliminary evaluation of five cases, there was good correlation between the computer results and the points marked by the radiologists (Fig. 2). The mean distance between the markings by the radiologists and the computer output was less than 1 mm in all cases (Table 1). The accuracy was higher in patients with only mild skin thickening of less than 1 cm (case #1, #2, #3). In these patients the maximum difference between radiologists markings and computer output was less than 2 mm. In areas with skin thickening of 1 cm or more (case #4, #5), differences of up to 4 mm for the internal skin line were found.

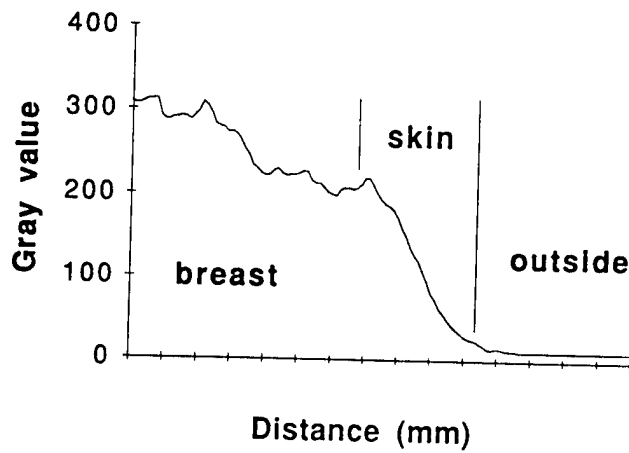


Fig. 1: Gray value profile of the breast perpendicular to the outside breast border. The internal skin contour is identified as a local gradient minimum. Skin thickness in this example (case #1) measures approximately 3 mm.

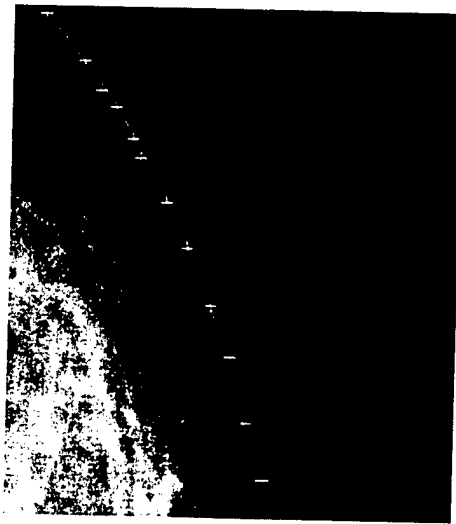


Fig. 2: Magnified view of computer generated internal (black line) and external skin border (white line) in case #2. For comparison, points marked by the radiologists are shown as black (x) along the internal skinline and as white (+) along the external skin contour.

Table 1: Mean and maximum distance between computer-generated border and radiologists markings for both the external and internal skinline. Results from the two radiologists were combined.

Case	Skin Thickness ¹	Distance ² between Computer Border and Radiologists Markings			
		number of points	mean distance	maximum	skinline
#1	5 mm	n = 146	0.23 ± 0.17 mm	0.7 mm	external
		n = 215	0.21 ± 0.17 mm	1.5 mm	internal
#2	6 mm	n = 105	0.29 ± 0.24 mm	1.5 mm	external
		n = 136	0.29 ± 0.27 mm	1.4 mm	internal
#3	7 mm	n = 161	0.26 ± 0.20 mm	1.0 mm	external
		n = 232	0.22 ± 0.20 mm	1.5 mm	internal
#4	15 mm	n = 144	0.50 ± 0.32 mm	1.6 mm	external
		n = 151	0.81 ± 0.67 mm	2.7 mm	internal
#5	22 mm	n = 145	0.30 ± 0.20 mm	1.5 mm	external
		n = 181	0.74 ± 0.84 mm	3.9 mm	internal

¹ values represent the maximum skin thickness found for each case.

² for each point marked by the radiologists, the shortest Euclidean distance to the corresponding computer border was calculated.

Discussion

Breast skin thickening can occur in a variety of benign disorders like edema, inflammation or scarring, however, it may also indicate underlying malignant disease [4-6]. Evaluation of the skin in mammograms is often difficult and may require the use of a hotlight in conventional mammography or special window settings in digital mammography. Thus, skin abnormalities may be easily overlooked and we have seen a case, in which the diagnosis of cancer was significantly delayed by this oversight. As part of a computer-aided diagnosis scheme for detection of breast abnormalities in mamograms, we are developing a computer algorithm to detect and measure skin thickening in mammograms.

Skin thickness in normal mammograms may vary between 0.5 and 3 mm and is usually greater in the inferior and medial portions of the breast [5,6]. Poor radiographic technique may impair visualization of the external skin surface and lead to underestimation of skin thickness [6]. For the same reason adequate digitization with a wide dynamic range is crucial when analyzing the skin in digitized mammograms [3]. Preliminary evaluation of our skin detection algorithm shows good correlation between computer-reported and radiologists-determined skin thickness for five clinical

mammograms. The mean difference between radiologists markings and computer results was less than 1 mm in all cases. Only in a few areas with marked skin thickening of more than 1 cm, points with differences of 2 - 4 mm were found.

By outlining areas of skin thickening, the computer results can be used to assist radiologists in detecting skin abnormalities on mammograms. The described segmentation of mammograms with reliable and accurate delineation of the outside breast contour may also serve as an entry point to our other automated analysis schemes in mammography including those for the detection of masses and clustered microcalcifications, which have been developed in our laboratory [1,2,7,8].

References

- 1 Chan H.-P.; Doi, K.; Galhotra, S.; Vyborny, C.J.; MacMahon, H.; Jokich, P.M.: Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography. *Med. Phys.* 14 (1987) 538-548.
- 2 Doi, K.; Giger, M.L.; MacMahon, H.; Hoffmann, K.R.; Nishikawa R.M.; Schmidt, R.A.; Chua K.-G.; Katsuragawa S.; Nakamori, N.; Sanada, S.; Yoshimura, H.; Metz, C.E.; Montner, S.M.; Matsumoto, T.; Chen, X.: Computer-aided diagnosis: development of automated schemes for quantitative analysis of radiographic images. *Sem. Ultrasound CT MRI* 13 (1992) 140-152.
- 3 Fredfeldt, K.E.; Christensen, E.; Conradsen, K.; Ersboll, B.; Stedstrup, S.: Automatic screening of plain film mammography. *Sem. Ultrasound CT MRI* 13 (1992) 135-139.
- 4 Leivonen, M.K.: Mammary skin thickening as a prognostic sign in breast cancer. *Ann. Chir. Gynaecol.* 76 (1987) 209-211.
- 5 Pope, T.L.; Read, M.E.; Medsker, T.; Buschi, A.J.; Brenbridge, A.N.: Breast skin thickness: normal range and causes of thickening shown on film-screen mammography. *J. Can. Assoc. Radiol.* 35 (1984) 365-368.
- 6 Willson, S.A.; Adam, E.J.; Tucker, A.K.: Patterns of breast skin thickness in normal mammograms. *Clin. Radiol.* 33 (1982) 691-693.
- 7 Wu, Y.; Doi, K.; Giger M.L.; Nishikawa, R.M.: Computerized detection of clustered microcalcifications in digital mammograms: applications of artificial neural networks. *Med. Phys.* (1992) 555-560.
- 8 Yin, F.-F.; Giger, M.L.; Doi, K.; Metz, C.E.; Vyborny, C.J.; Schmidt, R.A.: Computerized detection of masses in digital mammograms: analysis of bilateral subtraction images. *Med. Phys.* 18 (1991) 955-963.

Detection of Clustered Microcalcifications Using Fuzzy Modeling and Convolution Neural Network

Shih-Chung B. Lo, Huai Li, Jyh-Shyan Lin, Akira Hasegawa, Osamu Tsujii,
Matthew T. Freedman, and Seong K. Mun

Center of Imaging Science and Information Systems, Radiology Department,
Georgetown University Medical Center, Washington, DC 20007
e-mail: lo@isis.imac.georgetown.edu

Abstract

This paper describes an automatic computer searching system for detecting clustered microcalcifications. A fuzzy classification modeling was employed to extract each suspected microcalcification possessing similar physical parameters. Therefore, only those possible classes were evaluated using a sophisticated convolution neural network which requires a great deal of computation and serves as a discriminator. Based on the detected spots, many of them are true microcalcifications, the computer can easily make a determination when 5 spots are located within a defined region. However, when a cluster consists of only two to four suspicious spots a fuzzy function was used to determine the inclusion of other spots near the cluster. This can be very important for the detection of subtle cases. The membership of the latter fuzzy function was composed of the distance between the suspected spots as well as the output values of the convolution neural network.

We have tested the improved algorithms on our research database consisting of 45 mammograms. The results indicated that the fuzzy classification modeling decreased the number of false-positives from 2,874 to 1,067 suspected spots per image without increasing any false-negative detection. The over-all performance in the detection of clustered microcalcifications through the updated algorithms was 90% sensitivity at 0.5 false-positive per image. The computation time using a DEC-Alpha workstation was decreased from 5 minutes to about 3 minutes per image.

1. Introduction

Although mammography has a high sensitivity for examination of breast cancers when compared to other imaging modalities, studies indicate that radiologists do not detect all cancers that are visible on retrospective analyses of the mammograms¹⁻⁸. These missed detections are often a result of the very subtle nature of the radiographic findings. However, many missed diagnoses factors can be attributed to human errors including subjective or varying decision criteria, distraction by other image features, or simple oversight¹⁻². These problems may be reduced by an appropriate altering system which can assist in detecting and/or analyzing the mammographic lesions consistently and reproducibly. A well-trained computer system, that can be one of such a device, may draw the radiologist's attention to subtle image features that require more careful evaluation, or provide some quantitative measure of the likelihood of malignancy by comparing the image features of the lesions with those in its database.

Several investigators have been developing computer programs for automated detection of microcalcifications on mammograms⁹⁻¹⁴. It is difficult to compare the performance of these programs because the reported performance strongly depends on the degree of subtlety of microcalcifications on the test samples used. Our goal has been to detect subtle cases such as faint microcalcifications superimposed on dense breast regions, fibroglandular lines, or ducts. In this paper, we describe an

improved computer search system for the detection of clustered microcalcifications. Our methods include (a) fuzzified rules to enhance the extraction of suspected spots, (b) an artificial convolution neural network (CNN) to distinguish microcalcifications from background spots (noise), and (c) a fuzzy clustering method to group the detected calcifications.

2. Detection Methods

2.1. Wavelet Filtering for Background Reduction and Extraction of Suspected Microcalcifications

We used Daubechies 8-tap wavelet kernel to perform 3-level wavelet transform; only the lowest frequency compartment was eliminated for high-pass filtering¹⁵. We did not selectively enhance the isolated coefficients on the other high or medium frequency compartments because the reconstructed suspected areas were used for further evaluation in the artificial vision neural network to be discussed. Many microcalcifications were easily seen on the reconstructed mammogram after the wavelet filtering. In fact, a simple thresholding technique can extract obvious microcalcifications at this stage¹⁶⁻¹⁷. To detect those subtle microcalcifications, namely microcalcifications of 1-2 pixels in size, near the noise level in the local region, we need to use very low level threshold to extract all the suspect spots. In practice, several threshold from high to low level intensity were used to extract various contrast spots. Note that the filtered image has a mean value of 0.

2.2. Pre-Screening of Suspected Microcalcifications Using Gaussian Modeling

At this point, more than 10,000 spots per image may be selected if a threshold value near the standard deviation of the local region is used. Each local standard deviation, calculated excluding the high intensity pixels in the region, is considered the background fluctuation. The background noise is primarily attributed to the breast parenchyma and x-ray quantum mottle (Poisson noise). Typically, the denser the area the higher the noise. As far as calcifications are concerned, we expect that large calcifications (> 500 μm) possess high contrast in intensity and vice versa.

When a suspected spot is extracted, five parameters at the local area of the original image were computed, namely (i) the contrast of the peak (i.e., the peak intensity against the mean of the background) (C), (ii) the mean value of the local background around the peak (B), (iii) the standard deviation of the local background (SD), (iv) the size of the spot in pixel (SZ), and (v) the gradient of the spot from the center to peripheral (G).

For each parameter, $p \in \{C, B, C/SD, SZ, G\}$, the membership function is modeled by Gaussian distribution. The modeling is based on the true microcalcifications of the training mammograms.

$$u_{j(p)} = \exp\left\{-\left(p_j - p_o\right)^2 / \alpha(p)\right\} \quad \dots(1)$$

where j represents all the true microcalcifications in the training set, N is the total number of microcalcifications, $\alpha(p) = \frac{1}{N-1} \sum_{j=1}^N (p_j - p_o)^2$, and $p_o = \frac{1}{N} \sum_{j=1}^N p_j$

A set of mixed fuzzy and crispy rules are given below¹⁸⁻¹⁹

- Rule 1. If SZ is very large, then reject.
- Rule 2. If B is large and C/SD is small, then reject.
- Rule 3. If B is small and C/SD is very small, then reject.
- Rule 4. If C is large and SZ is very small, then reject.
- Rule 5. If C and SZ are large and G is very small, then reject.
- Rule 6. If rules 1-5 do not apply, submit for further evaluation by the CNN.

Membership functions for the linguistic labels (i.e., very small, small, large, very large) are defined by the domain of each parameter normalized by the mean value of the parameter obtained in the training set.

2.3. Convolution Neural Network as an Artificial Vision System^{13-14, 20}

2.3.1. The structure of the convolution neural network (CNN)

The CNN is a simplified version of the neocognitron²¹. We used a convolution constrained neural network with the well known backpropagation method for training. Readers are encouraged to see the detailed description and the structure of the CNN in the references.

Basically, the computer locates the bright spot at the center of the patch, which was extracted through the pre-scan procedure, for the CNN evaluation. The operation between each kernel and the image block follows a standard 2-D convolution calculation. For the forward signal propagation, the resultant values of the weighting factors of the kernel convoluting the element values of the front layer are collected into the corresponding matrix elements on the receiving layer. The collected value at each element is further operated with a Gaussian-like or sigmoid function in the forward propagation.

Each suspected image block of $N \times N$ pixels indicated in the pre-scan program is extracted as an object for CNN classification. The first hidden layer consists of 12 groups. Each group has 12×12 pixels formatted in a square array. The second hidden layer also consists of 12 groups. Each group has 8×8 pixels. In our experiment, both one and two hidden layer systems were tested for the evaluation of microcalcifications. The output layer has only one node which is fully connected to the second hidden layer.

It is important to realize that the total number of nodes needed in the hidden layers somewhat depends on the total number of training samples. Since we plan to expand our database and the use of rotated versions of an input matrix, we expect that our training samples will be very large in the future. The number of layers used should depend upon the sophistication of the features that the neural network is intended to perceive. The more complicated the disease patterns, the more layers are required to distinguish high order information of image structures. Since microcalcifications are considered relatively simple in terms of image patterns, two hidden layers and one output node should be able to handle various cases based on our experience.

2.3.2. Backpropagation training with Gaussian-like and sigmoid activation functions

The main difference between conventional weights and kernel weights is that conventional weights are independent and kernel weights are constrained by grouping. We believe that the latter is more powerful than the former method for direct image pattern recognition. Training requires many iterations for the network to obtain solutions for all weights applied to the propagation while the error function reaches a minimum value.

By looking at the CNN processing, one may find that signals are filtered and modulated as in a complicated circuit system. Signal propagation from one layer to the next is composed of a two-step calculation: (a) adaptive convolution combiner and (b) an activation function (Gaussian-like and sigmoid function are used in this study). To be exact, the Gaussian-like activation function is used at the first hidden layer and the sigmoid function is applied at the second hidden and output layers given by Eqs. (2) and (3), respectively.

$$G_x((i,j);n) = \frac{2 \times \exp \left\{ - \sum_{u,v,m \in n} [k_x((u,v);n,m) \times S_{x-1}((i-u,j-v);m)] \right\}}{\left(1 + \exp \left\{ - \sum_{u,v,m \in n} [k_x((u,v);n,m) \times S_{x-1}((i-u,j-v);m)] \right\} \right)^2} \quad \dots(2)$$

$$S_x((i,j);n) = \frac{1}{1 + \exp\left\{-\sum_{u,v,m \in n} [k_x((u,v);n,m) \times S_{x-1}((i-u,j-v);m)]\right\}} \quad \dots(3)$$

where $G_x((i,j);n)$ or $S_x((i,j);n)$ represents the signal at node (i,j) , n^{th} group, and x layer. $k_x((u,v);n,m)$ denotes the weighting factor value of net (u,v) in n^{th} group of $x-1$ layer which connects m^{th} group of x layer. $m \in n$ represents the connection between the two groups.

The error function which is expected to reach a local minimum through the error backpropagation training can be given as

$$E = \frac{1}{2} \sum_{n_o=1}^T [y(n_o) - S_o(n_o)]^2 \quad \dots(4)$$

where $y(n_o)$ and $S_o(n_o)$ are the target output and calculated output signals for output node n_o , respectively. T is the total number of output nodes. Based on Eqs. (2), (3) and (4), the iterative version of kernel weights for sigmoid function derived by the generalized delta rule is given as

$$k_x((u,v);n,m)[t+1] = k_x((u,v);n,m)[t] + \eta \sum_{i,j} \delta_x((i,j);n) S_{x-1}((i-u,j-v);m) + \alpha \Delta k_x((u,v);n)[t] \quad \dots(5)$$

where t is the iteration number during the training, η is the gain for the current weight changes, α is the gain for the momentum term received in the last learning loop, and δ is the weight-update function which is given as

$$\delta_x((i,j);n) = S_x((i,j);n) [1 - S_x((i,j);n)] Q_x((i,j);n) \quad \dots(6)$$

$$\text{and } Q_x((i,j);n) = \sum_{u,v,n \in m} k_{x+1}((u,v);n,m) \times \delta_{x+1}((i+u,j+v);m).$$

Similarly, the iterative version of kernel weights for Gaussian-like activation function derived by the generalized delta rule is given as:

$$k_x((u,v);n,m)[t+1] = k_x((u,v);n,m)[t] + \eta \sum_{i,j} \delta_x((i,j);n) S_{x-1}((i-u,j-v);m) + \alpha \Delta k_x((u,v);n)[t]. \quad \dots(7)$$

In this study, all weighting factors including the kernels were initially given a normalized random number. The normalization is based on the number of nets connecting to a destination node in the next layer.

2.3.3. Output value of the CNN

After the training, all weights in the neural network are fixed. Since we only use one node at the output layer, the output value of 1 or close to 1 should be considered as a determination of a positive determination by the CNN. Otherwise, the low output value is a negative determination. It is very difficult to use 0.5 as a cut-off point using the trained neural network. The cut-off point may be

shifted by the inevitable bias in the training cases. In practice the cut-off point is established by many clinical cases in a rigorous evaluation study.

2.4. Fuzzy Clustering Based on CNN Outputs and Distance between Suspected Microcalcifications

In general, clustered microcalcifications of more than 5 within a 1 cm² area are a clinically significant sign of potential breast cancer. However, clusters containing only 3 or 4 microcalcifications are sometimes seen in subtle cases. The distance between suspected microcalcifications, which receive a relatively high output value from the CNN, were calculated. Once a short distance is found between two spots (less than 1.5 cm), then a pre-cluster status is engaged. The remaining suspects were also evaluated to test the possible clustering with the group. Two main parameters, namely, the distance (d_{ij}) from the spot, i , to the center of the suspected cluster j and the CNN output value of the spot O_i , are evaluated for the possibility of engaging a cluster. Only if d_{ij} is shorter than a predefined distance d_{r2} and O_i is greater than a low bound of CNN output value O_l , the likelihood of being a member of a cluster be considered

$$LM_{ij}(O_i, d_{ij}) = \begin{cases} O_i & \text{for } d_{ij} < d_{r1} \text{ and } O_i \geq O_l \\ O_i \times [(d_{r2} - d_{ij}) / (d_{r2} - d_{r1})] & \text{for } d_{r2} > d_{ij} > d_{r1} \text{ and } LM_{ij} \geq O_l \end{cases} \quad \dots(8)$$

where d_{r1} denotes the near cluster distance. The first situation includes those spots possessing relatively high output values and are close to the cluster. The second situation represents those spots very high CNN output value with a reasonably short distance from the existing cluster. A fuzzified factor, $[(d_{r2} - d_{ij}) / (d_{r2} - d_{r1})]$, is used to decrease its likelihood of being a member in the cluster as the distance is increased. In either case, the resultant value of its likelihood membership should be higher than O_l ; otherwise its application of being a member in the cluster will be rejected at this stage. Both situations will be considered as members in the cluster. The updated likelihood of a cluster, if the new member is qualified, is given by

$$LC'_j = \begin{cases} LC_j & \text{for } N_j > 3 \text{ and } LM_h > LM_{ij} > LM_l \\ (LC_j \times N_j + LM_{ij}) / (N_j + 1) & \text{for } LM_{ij} > LM_h \end{cases} \quad \dots(9)$$

where N_j denotes the number of existing members in the cluster. LM_h and LM_l represent the high and low bounds of membership qualification, respectively, for being included in the cluster. The number of members in the cluster is also updated:

$$N'_j = N_j + 1 \quad \text{for } (N_j > 3 \text{ and } LM_h > LM_{ij} > LM_l) \text{ or } (LM_{ij} > LM_h). \quad \dots(10)$$

3. Results

A total of 65 mammograms were digitized by a LumyScan film digitizer with a spot size of 100 μ m. Each mammogram takes 2048 $\times$ 2430 $\times$ 12 bits of computer space. Of these mammograms, 20 were used as the training set. In this training set, each mammogram contains a biopsy proven clustered microcalcifications. A total of 159 microcalcifications are embedded in 20 clusters. The remaining 45 mammograms were used for the test set. Of these mammograms, 38 contain biopsy proven clustered microcalcifications. A total of 220 microcalcifications are embedded in 41 clusters.

3.1. Results of the Pre-Scan

In the prescan study, the average initial detection of suspected spots was 82,568 spots per image. Using the conventional contrast discrimination method, we could narrow this number down to

2,874 spots per image with 5 false-negatives out of 220 microcalcifications. However, the fuzzy modeling further reduced the suspected areas to 1,067 spots per image with 2 false-negative findings. These two microcalcifications are overlapped with fibroglandular tissues. Further testing indicated that these two spots could not be detected by the convolution neural network either.

3.2. Results of CNN and Fuzzy Clustering

Of 41 true clustered microcalcifications, only 3 cases possess less than 5 microcalcifications in each cluster. The fuzzy clustering was able to include these three cases for final group evaluation. However, total of 15 and 8 false-positive clusters consisting of 3-4 suspected spots were also included in the final fuzzy clustering evaluation for the 1- and 2-hidden layer CNN, respectively. Using 1-hidden layer CNN, 5 out of 15 false-positive clusters contributed significant LC values which degraded the performance of the entire system. None of the 8 false-positive clusters in 2-hidden layer system produced a high LC value and did not contribute negative impact on the overall Az performance. Figure 1 shows the Az's of the two systems with various CNN training errors. Syntax Sn/Hm/Kt indicates the CNN parameters: image block size $n \times n$, m hidden layers, and kernel size $t \times t$.

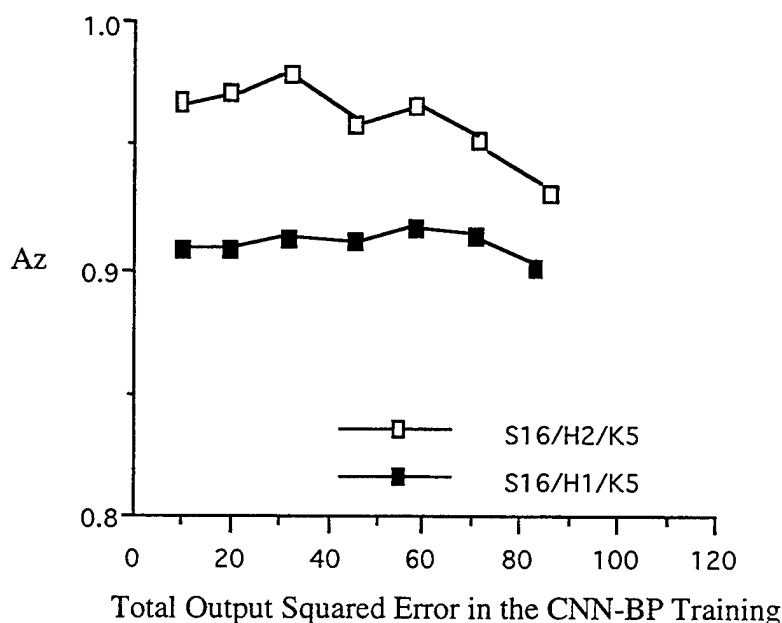


Figure 1. Final system performance using 1-hidden and 2-hidden layer CNN.

4. Discussion and Conclusions

The purpose of the pre-scan is to reduce the number of false positives without losing the suspect spots. Since our goal is to detect subtle microcalcifications, we should not tighten the screening parameters. Because tightened parameters will force fuzzy logic to exclude the subtle microcalcifications. Although over 170,000 spots on 20 mammograms were evaluated in the fuzzy model training study, most of them are correlated and can not represent the diversity in microcalcification patterns. In this study, loosened parameters were used in the fuzzy modeling to discriminate the noises of the mammographic background. We found that a significant number of false-positives were reduced when the fuzzification of the measured parameters was employed. This result indicated that the fuzzy system is able to find a large common area (or areas) in the multidimensional space containing many false-positives and without producing additional false-negatives.

These results would greatly reduce the load of CNN which is the computational bottleneck of this computer search program. Although the pre-scan has eliminated a large number of suspected microcalcifications, the majority of the remaining suspects are still false-positives. The convolution neural network is then used to discern the true and false-positives. The pre-scan failed to detect on the same clustered, hence the cluster was not detected. We also used the CNN to evaluate the two specific areas, the out values of CNN were 0.003 and 0.14. This meant that even if the pre-screening had included these two true microcalcifications, the CNN would have still missed them.

We found that the fuzzy clustering involving CNN output values and distance between spots is an effective method to determine microcalcification cluster. Particularly in the 2-hidden layer system, this coupling method can detect subtle microcalcification clusters and at the same time exclude false-positives.

5. Acknowledgments

This work was supported by a US Army Research Grant No. DAMD17-93-J-3007. The content of this paper does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred. The authors are grateful to Ms. Susan Kirby for her editorial assistance.

6. References

1. Martin JE, Moskowitz M, Milbrath JR: Breast cancer missed by mammography. *AJR* 1979;132:737.
2. Kalisher L: Factors influencing false negative rates in xeromammography. *Radiology* 1979;133:297.
3. Moskowitz M. Screening for breast cancer: how effective are our tests? A critical review. *Ca* 1983;33:26-39.
4. Baines CJ, Miller AB, Wall C, et al: Sensitivity and specificity of first screen mammography in the Canadian National Breast Screening Study: A preliminary report from five centers. *Radiology* 1986;160:295.
5. Haug PJ, Tocino IM, Clayton PD, Bair TL: Automated management of screening and diagnostic mammography. *Radiology* 1987;164:747.
6. Bassett LW, Bunnell DH, Jahanshahi R, Gold RH, Arndt RD, Linsman J: Breast cancer detection: One versus two views. *Radiology* 1987;165:95.
7. Hillman BJ, Fajardo LL, Hunter TB, et al: Mammogram interpretation by physician assistants. *AJR* 1987;149:907.
8. Hermann G, Janus C, Schwartz I, Papatestas A, Hermann DG, Rabinowitz JG: Occult malignant breast lesions in 114 patients: relationship to age and the presence of microcalcifications. *Radiology* 1988;169:321.
9. Chan HP, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM: Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography", *Med. Phys.*, 1987;14:538.
10. Chan HP, Doi K., Vyborny CJ, et al.: Improvement in radiologists' detection of clustered microcalcifications on mammograms: The potential of computer-aided diagnosis," *Invest. Radio.* 1990; 25:1102-1110.
11. Davies DH, Dance DR: Automatic computer detection of clustered calcifications in digital mammograms. *Phys Med Biol* 1990; 35:1111.
12. Fam BW, Olson SL, Winter PF, Scholz FJ: Algorithm for the detection of fine clustered calcifications on film mammograms. *Radiology* 1988;169:333.
13. Chan HP, Lo SC: Computer-aided detection of mammographic microcalcifications, *Medical Physics*, 1995; 22(10):1555-1567.

14. Lo SC, Chan HP, Lin JS, Li H, Freedman MT, and Mun SK: Artificial convolution neural network for medical image pattern recognition, *Neural Network*, 1995; 8(7):1201-1214.
15. Daubechies I: Orthonormal based of compactly supported wavelets, *Comm. on Pure and Appl. Math.*, 1988; XLI:909-996.
16. Qian W, Clark LP, Kallergi, and Clark RA: Tree-structured nonlinear filters in digital mammography, *IEEE Trans. on Med. Img*, 1994;13(1):25-36.
17. Liang AF, Schuler S, Fan J, and Huda W: Mammographic feature enhancement by multiscale analysis, *IEEE Trans. on Med. Img*, 1994; 13(4):725-740.
18. Zedeh LA, Probability measures of fuzzy events, *J. Math. Anal and Applicat*, 1968; 23:421-427.
19. Bezdek J and Pal SK, Fuzzy Models for Pattern Recognition, New York, IEEE press, 1992.
20. Lo SC, Lou SL, Lin JS, Freedman MT, and Mun SK: Artificial Convolution Neural Network techniques and applications for lung nodule detection, *IEEE Trans. on Med. Img*, 1995; 14(4):711-718.
21. Fukushima K, Miyake S, and Ito T: Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 1983; 13(5):826-834.