

AD _____

Award Number: DAMD17-01-1-0326

TITLE: Automated Method for Analysis of Mammographic Breast Density – A
Technique for Breast Cancer Risk Estimation

PRINCIPAL INVESTIGATOR: Heang-Ping Chan, Ph.D.

CONTRACTING ORGANIZATION: The University of Michigan
Ann Arbor, MI 48109-1274

REPORT DATE: July 2006

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, 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.

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> 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 this 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 Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 01/07/06		2. REPORT TYPE Final		3. DATES COVERED (From - To) 01 July 01 – 30 Jun 06	
4. TITLE AND SUBTITLE Automated Method for Analysis of Mammographic Breast Density – A Technique for Breast Cancer Risk Estimation				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER DAMD17-01-1-0326	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Heang-Ping Chan, Ph.D. E-Mail: chanhp@umich.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Michigan Ann Arbor, MI 48109-1274				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The goal of this proposed project is to develop an automated technique to assist radiologists in estimating mammographic breast density. During the project years, we developed an automated mammographic density segmentation system, referred to as Mammographic Density ESTimator (MDEST), for both DMs and DFMs. Our studies showed that the automated MDEST system can provide percent dense area estimates that are highly correlated with radiologists' interactive thresholding results and the percent volumetric fibroglandular tissue estimates from MR breast images. The quantitative estimates are superior to the radiologists' qualitative BI-RADS density assessment. The MDEST system can provide a consistent and reproducible estimation of percent dense area on routine clinical mammograms. This will facilitate studies of various factors associated with breast cancer risk and mammographic sensitivity, and monitoring the effects of interventional or preventive strategies. The image analysis tool will therefore contribute to the understanding of the relationship of density to breast cancer risk, detection, prognosis, and to the prevention and treatment of breast cancers.					
15. SUBJECT TERMS Breast cancer Mammography, breast density, computer-assisted analysis, automated segmentation, risk monitoring					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			USAMRMC
			UU	71	19b. TELEPHONE NUMBER (include area code)

(3) Table of Contents

Introduction.....	4
Body	5
(A) Collection of a Database of Full Field Digital Mammograms (DMs) and Digitized Mammograms (DFM)	
(B) Automated Breast Density Analysis System for Digitized Mammograms	
(C) Correlation between Percent Dense Area and Percent Volumetric Fibroglandular Tissue	
(D) Correlation between Breast Density estimated on Digitized Screen-Film Mammograms and Full Field Digital Mammograms	
(E) Automated Breast Density Analysis System for Digital Mammograms	
(F) Breast Segmentation: Breast Boundary Detection and Pectoral Muscle Trimming	
(G) Comparison of Mammographic Density by MDEST Radiologists' Estimates and BI-RADS Categories	
Key Research Accomplishments	7
Reportable Outcomes.....	8
Conclusions.....	10
Personnel.....	10
References.....	10
Appendix.....	10

(4) Introduction

Previous studies have found that there is a strong correlation between mammographic breast density and the risk of breast cancer. Mammographic breast density has been used by researchers in many studies to estimate breast cancer risk of epidemiological factors, monitor the effects of preventive treatments such as tamoxifen or dietary interventions, monitor the breast cancer risk of hormone replacement therapy, and investigate factors affecting mammographic sensitivity and cancer prognosis. However, most studies used Breast Imaging Reporting and Data System (BI-RADS) density rating as a measure of mammographic breast density, which contributes large inter- and intraobserver variations and may reduce the sensitivity of the analysis.

The goal of this proposed project is to develop a fully automated technique to assist radiologists in estimating mammographic breast density. We hypothesize that the computerized technique can accurately and efficiently segment the dense area on digitized or digital mammograms, thereby eliminating inter- and intra-observer variations. The dense area as a percentage of total breast area thus estimated will be more consistent and reproducible than radiologists' subjective BI-RADS rating. To accomplish this goal, we will (1) collect a large database of mammograms, including digitized film mammograms and digital mammograms, for training and testing the dense area segmentation program; (2) evaluate the correlation between the radiologists' breast density classification based on BI-RADS lexicon and the percent breast dense area; (3) study the correlation of percent breast dense area between different views of the same breast and between the same view of the two breasts; and (4) investigate the correlation between the percent breast dense area estimated from mammograms and the volumetric dense breast tissue estimated from a data set of magnetic resonance (MR) breast images. These comparisons will provide important information regarding the consistency of the BI-RADS rating with the measured percent breast dense area, the appropriate measure of % dense area from different mammographic views, and the usefulness of using the percent breast dense area on mammograms as an indicator of volumetric breast tissue density.

It is expected that this project will produce a fully automated and effective tool for analysis of mammographic breast density, which can be applied to routinely acquired mammograms without special calibrations. This will facilitate studies of various factors associated with breast cancer risk and mammographic sensitivity, and monitoring the effects of interventional or preventive strategies. The image analysis tool will therefore contribute to the understanding of the relationship of density to breast cancer risk, detection, prognosis, and to the prevention and treatment of breast cancer.

(5) Body

This is the final report of the project. We have described in detail the results of our studies in the past annual progress reports. The investigations conducted in this project are summarized in the following.

(A) Collection of a Database of Full Field Digital Mammograms (DMs) and Digitized Mammograms (DFM)

With IRB approval, we have collected a database of full field direct digital mammograms (DM) and a database of screen-film mammograms (DFM). The database of DMs contained about 290 cases with over 580 pairs of DMs. The database of DFMs included about 380 cases with over 760 DFMs. The data sets were used for the development of the breast density analysis tools for DMs and DFMs.

(B) Automated Breast Density Analysis System for Digitized Mammograms

We developed an automated image analysis tool Mammographic Density ESTimator (MDEST), to assist radiologists in estimating breast density in terms of percent dense area on mammograms. MDEST performs dynamic range compression, breast boundary tracking, pectoral muscle segmentation for the MLO view, automatic thresholding based on gray level histogram analysis, and calculates the percent dense area on a mammogram. In our study, we found that the correlation between the computer-estimated percent dense area and radiologists' manual segmentation was 0.94 and 0.91, respectively, for CC and MLO views, with a mean bias of less than 2%. These results indicated that the automated image analysis tool can be an efficient and reproducible method for breast density segmentation. (Publications J1, J2, J3, P1, P2, A1, A2, A3, A6, A8)

(C) Correlation between Percent Dense Area and Percent Volumetric Fibroglandular Tissue

We performed a comparison between fibroglandular tissue volume segmented from 3-dimensional MR images and the percent dense area estimated on corresponding DFMs. The correlation of manually segmented percent dense area of the CC and MLO views by radiologists with the percent volumetric fibroglandular tissue on MR images was found to be 0.91 and 0.91, respectively. The correlation of automated MDEST segmented percent dense area of the CC and MLO views with the percent volumetric fibroglandular tissue on MR images was found to be 0.91 and 0.89, respectively. The mean differences between the percent dense area and the MR percent volumetric density range from 3% to 6%. The high correlation indicated that the percent dense area on mammograms can be used as a surrogate for monitoring breast density changes. (Publications J2, P1, A1, A3)

(D) Correlation between Breast Density estimated on Digitized Screen-Film Mammograms and Full Field Digital Mammograms

We studied the correlation between the percent dense area estimated on DFMs and DMs. Pairs of two-view DFM and DM from 99 patients were used. All mammograms were manually segmented by 5 experienced breast radiologists using a graphical user interface developed in this project. The mean difference in the segmented density between DFMs and DMs was about 3.5%

and the mean ratio was about 1.30, indicating that the mammographic density was about 30% higher, on average, on the DFMs. The difference in the mammographic density may be attributed to the harder beam quality used and the digital image processing applied to the DMs. The lower density may improve the mammographic sensitivity for lesion detection on dense breasts. However, for patients with DFMs and DMs taken over time, comparison of serial mammograms for breast density changes will be problematic. (Publications A4)

(E) Automated Breast Density Analysis System for Digital Mammograms

The MDEST system was originally developed using DFMs. To adapt the system to DMs, we incorporated an Expectation-Maximization (EM) algorithm to extract gray level features from the histogram. A rule-based classifier was trained to estimate a gray level threshold for segmenting the dense area from the breast region adaptively. The performance of the MDEST system for DMs was improved after the incorporation of the EM estimation and re-training. The correlation between the computer-estimated percent dense area and the radiologists' manual segmentation improved from 0.85 and 0.87 to 0.94 and 0.92, respectively, for CC and MLO views. The root-mean-square (RMS) errors improved from 7.3% and 5.7% to 4.2 and 4.4%, respectively, for CC and MLO views. (Publications P2, A5, A6, A8)

(F) Breast Segmentation: Breast Boundary Detection and Pectoral Muscle Trimming

Breast density estimation depends strongly on the accuracy of breast boundary segmentation and, for the MLO view, pectoral muscle segmentation. We developed an improved breast boundary tracking method using Sobel edge detection in combination with multiple edge thresholding to generate contour images and active contour for segmentation. For the detection of pectoral boundary, we developed a method that included gradient-based directional filtering, gradient-based texture analysis to generate an orientation image, mean shift smoothing, and ridge-tracking. For a data set of 118 MLO view mammograms, the new method identified 99.2% of the pectoral muscles. (Publications P3, A7)

(G) Comparison of Mammographic Density by MDEST Radiologists' Estimates and BI-RADS Categories

The fully-automated MDEST computer program was used to estimate breast density on digitized mammograms from 65 women, ages 24-89 (mean = 53). Five breast radiologists prospectively assigned qualitative BI-RADS density categories and visually estimated percent density of 260 mammograms. Qualitative BI-RADS assessments were compared to new quantitative BI-RADS standards. The reference standard density for this study was established by allowing the 5 radiologists to manipulate the MDEST gray level threshold, which segmented mammograms into dense and non-dense areas. It was found that there was close correlation between the reference standard and radiologist-estimated density (correlation = .90-.95) and MDEST density (correlation = 0.89). MDEST had tighter agreement with the reference standard, with an average overestimate of 1% (-15% to +18%). MDEST correlated better with percent density than qualitative BI-RADS categories. There was large overlap and range of percent density in qualitative BI-RADS categories 2 through 4. Qualitative BI-RADS categories correlated poorly with the new quantitative BI-RADS categories. 6% (16/260) of views were erroneously classified by MDEST. This study indicated that MDEST compared favorably to radiologist estimates of percent density, and is more reproducible than radiologist estimates

using qualitative BI-RADS density categories. Qualitative and quantitative BI-RADS density assessments differed markedly. (Publication J3)

(6) Key Research Accomplishments

- Collected over 670 cases digitized film mammograms and digital mammograms for development of the automated density segmentation program (Task 1).
- Developed an automated image analysis tool Mammographic Density ESTimator (MDEST) for estimation of the percentage of breast dense area on mammograms. (Task 2)
- Evaluated the segmentation accuracy by comparison with radiologists' manual segmentation and found that the correlation between the computer-estimated percent dense area and radiologists' manual segmentation was 0.94 and 0.91, respectively, for CC and MLO views, with a mean bias of less than 2%. (Task 2)
- Evaluated correlation between mammographic density and BI-RADS density ratings and found that there was a close correlation between the reference standard and radiologist-estimated density ($R=.90-.95$) and MDEST density ($R = 0.89$). MDEST correlated better with percent density than qualitative BI-RADS categories. (Task 2)
- Evaluated correlation between mammographic density and volumetric density measurement with MR breast images and found that the correlation of automated MDEST segmented percent dense area of the CC and MLO views with the percent volumetric fibroglandular tissue on MR images was 0.91 and 0.89, respectively. The mean differences between the percent dense area and the MR percent volumetric density ranged from 3% to 6%. (Task 3)
- Evaluated the correlation between mamomgraphic density estimated on DFMs and DMs and found that the mean difference in the segmented density between DFMs and DMs was about 3.5% and the mean ratio was about 1.30, indicating that the mammographic density was about 30% higher, on average, on the DFMs. (Tasks 2 and 4)
- Adapted the MDEST density segmentation system to digital mammograms successfully and the system now can be applied to both modalities (Task 4)
- Compared automated density segmentation results on DMs with radiologists' manual segmentation and found that the correlation between the computer-estimated percent dense area and the radiologists' manual segmentation was 0.94 and 0.92, respectively, for CC and MLO views. The root-mean-square (RMS) errors were 4.2 and 4.4%, respectively, for CC and MLO views (Task 4).
- Developed a breast boundary detection method and a pectoral muscle trimming method to improve the accuracy of breast area estimation for calculation of percent dense area. For a data set of 118 MLO view mammograms, the pectoral muscle trimming method identified 99.2% of the pectoral muscles (Task 2 and Task 4).

(7) Reportable Outcomes

As a result of the support by the USAMRMC BCRP grant, we have developed an automated breast density analysis system for both DMs and DFMs. We have also developed graphical user interfaces for interactive thresholding of breast density on mammograms by radiologists efficiently. The results of these investigations have been presented in international conferences or published in peer-reviewed journals. The publications from this project are listed below. Many of these have been reported in the annual reports.

Journal Articles:

- J1. Zhou C, Chan HP, Petrick N, Helvie MA, Goodsitt MM, Sahiner B, Hadjiiski LM. Computerized image analysis: Estimation of breast density on mammograms. Medical Physics 2001; 28: 1056-1069.
- J2. Wei J, Chan HP, Helvie MA, Roubidoux MA, Sahiner B, Hadjiiski L, Zhou C, Paquerault S, Chenevert T, Goodsitt MM. Correlation between Mammographic Density and Volumetric Fibroglandular Tissue Estimated on Breast MR Images. Medical Physics 2004; 31: 933-942.
- J3. Martin KE, Helvie MA, Zhou C, Roubidoux MA, Bailey JE, Paramagul C, Blane CE, Klein K, Sonnad S, Chan HP. Mammographic density measured by an automatic computer-aided quantitative method: Comparison with radiologists' estimates and BI-RADS categories. Radiology 2006; 240: 656-665.

Conference Proceeding:

- P1. Chan HP, Hadjiiski LM, Roubidoux MA, Helvie MA, Paquerault S, Sahiner B, Chenevert T, Goodsitt MM. Breast density estimation: correlation of mammographic density and MR volumetric density. In: Digital Mammography IWDM 2002: 6th International Workshop on Digital Mammography. Ed. Peitgen HO. (Springer, Berlin) 2003: 281-284.
- P2. Zhou C, Chan HP, Wei J, Helvie MA, Roubidoux MA, Paramagul C, Nees A, Hadjiiski LM, Sahiner B. Performance evaluation of an automated breast density estimation system for digital mammograms and digitized film mammograms. In: Digital Mammography IWDM 2004: 7th International Workshop on Digital Mammography. Ed. Pisano E. 425-429.
- P3. Zhou C, Hadjiiski LM, Paramagul C, Sahiner B, Chan HP, Wei J. Computerized pectoral muscle identification on MLO-view mammograms for CAD applications. Proc SPIE 5747; 2005: 852-857.

Conference Abstracts and Presentations:

- A1. Chan HP, Hadjiiski LM, Roubidoux MA, Helvie MA, Paquerault S, Sahiner B, Chenevert T, Goodsitt MM. Breast density estimation: correlation of mammographic density and MR volumetric density. Poster presentation at the 6th International Workshop on Digital Mammography. IWDM-2002. Bremen, Germany. June 22-25, 2002.
- A2. Chan HP, Helvie MA, Wei J, Hadjiiski LM, Zhou C, Goodsitt MM, Sahiner B, Roubidoux MA. Automated analysis of mammographic breast density for breast cancer risk estimation. Presented at the Era of Hope Meeting, U. S. Army Medical Research and Materiel Command, Department of Defense, Breast Cancer Research Program, Orlando, Florida, September 25-28, 2002.
- A3. Wei J, Chan HP, Helvie MA, Hadjiiski LM, Sahiner B, Roubidoux MA, Zhou C, Paquerault S, Chenevert T, Goodsitt MM. Breast density estimation on mammograms and MR images: A tool for assessment of breast cancer risk. Presentation at the 88th Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL, December 1-6, 2002. Radiology 2002; 225(P): 600.
- A4. Chan HP, Wei, J, Zhou C, Helvie MA, Roubidoux MA, Bailey J, Paramagul C, Nees A, Hadjiiski LM, Sahiner B. Comparison of mammographic density estimated on digital mammograms and screen-film mammograms. Presentation at the 89th Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL, November 30-December 5, 2003. RSNA Program 2003; 424.
- A5. Zhou C, Hadjiiski LM, Sahiner B, Chan HP, Helvie MA, Wei, J. Computerized mammographic breast density estimation: Expectation-Maximization estimation and neural network classification of breast density. Presented at the 89th Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL, November 30-December 5, 2003. RSNA Program 2003; 389.
- A6. Zhou C, Chan HP, Wei J, Helvie MA, Roubidoux MA, Paramagul C, Nees A, Hadjiiski LM, Sahiner B. Performance evaluation of an automated breast density estimation system for digital mammograms and digitized film mammograms. Presented at the 7th International Workshop on Digital Mammography. IWDM-2004. Durham, North Carolina. June 18-21, 2004.
- A7. Zhou C, Hadjiiski LM, Paramagul C, Sahiner B, Chan HP, Wei J. Computerized pectoral muscle identification on MLO-view mammograms for CAD applications. Poster presentation at the SPIE International Symposium on Medical Imaging, San Diego, CA, February 12-17, 2005.
- A8. Zhou C, Chan HP, Helvie MA, Wei J, Ge J, Hadjiiski LM, Sahiner B, Computerized mammographic breast density estimation on full field digital mammogram and digitized film mammogram. Presentation at the 91st Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL. November 27-December 2, 2005. RSNA Program 2005; 271.

(8) Conclusions

We have developed an automated mammographic density segmentation system, referred to as Mammographic Density ESTimator (MDEST), for both DMs and DFMs. Our studies showed that the automated MDEST system can provide percent dense area estimates that are highly correlated with radiologists' interactive thresholding results and the percent volumetric fibroglandular tissue estimates from MR breast images. The quantitative estimates are superior to the radiologists' qualitative BI-RADS density assessment. The MDEST system can provide a consistent and reproducible estimation of percent dense area on routine clinical mammograms. This will facilitate studies of various factors associated with breast cancer risk and mammographic sensitivity, and monitoring the effects of interventional or preventive strategies. The image analysis tool will therefore contribute to the understanding of the relationship of density to breast cancer risk, detection, prognosis, and to the prevention and treatment of breast cancers.

(9) Personnel

Heang-Ping Chan, Ph.D.	PI
Berkman Sahiner, Ph.D.	Co-investigator
Thomas Chenevert, Ph.D.	Co-investigator
Mark A. Helve, M.D.	Clinical Co-investigator
Marilyn A. Roubidoux, M.D.	Clinical Co-investigator
Chuan Zhou, Ph.D.	Co-investigator
Jun Wei, Ph.D.	Postdoctoral fellow
Jun Ge, Ph.D.	Postdoctoral fellow
Yiheng Zhang, Ph.D.	Postdoctoral fellow

(10) References

None

(11) Appendix

The following publications are enclosed with this report:

Journal Articles:

1. Zhou C, Chan HP, Petrick N, Helvie MA, Goodsitt MM, Sahiner B, Hadjiiski LM. Computerized image analysis: Estimation of breast density on mammograms. Medical Physics 2001; 28: 1056-1069.

2. Wei J, Chan HP, Helvie MA, Roubidoux MA, Sahiner B, Hadjiiski L, Zhou C, Paquerault S, Chenevert T, Goodsitt MM. Correlation between Mammographic Density and Volumetric Fibroglandular Tissue Estimated on Breast MR Images. Medical Physics 2004; 31: 933-942.
3. Martin KE, Helvie MA, Zhou C, Roubidoux MA, Bailey JE, Paramagul C, Blane CE, Klein K, Sonnad S, Chan HP. Mammographic density measured by an automatic computer-aided quantitative method: Comparison with radiologists' estimates and BI-RADS categories. Radiology 2006; 240: 656-665.

Conference Proceeding:

1. Chan HP, Hadjiiski LM, Roubidoux MA, Helvie MA, Paquerault S, Sahiner B, Chenevert T, Goodsitt MM. Breast density estimation: correlation of mammographic density and MR volumetric density. In: Digital Mammography IWDM 2002: 6th International Workshop on Digital Mammography. Ed. Peitgen HO. (Springer, Berlin) 2003: 281-284.
2. Zhou C, Chan HP, Wei J, Helvie MA, Roubidoux MA, Paramagul C, Nees A, Hadjiiski LM, Sahiner B. Performance evaluation of an automated breast density estimation system for digital mammograms and digitized film mammograms. In: Digital Mammography IWDM 2004: 7th International Workshop on Digital Mammography. Ed. Pisano E, 425-429.
3. Zhou C, Hadjiiski LM, Paramagul C, Sahiner B, Chan HP, Wei J. Computerized pectoral muscle identification on MLO-view mammograms for CAD applications. Proc SPIE 5747; 2005: 852-857.

Conference Abstracts and Presentations:

- A1. Chan HP, Hadjiiski LM, Roubidoux MA, Helvie MA, Paquerault S, Sahiner B, Chenevert T, Goodsitt MM. Breast density estimation: correlation of mammographic density and MR volumetric density. Poster presentation at the 6th International Workshop on Digital Mammography. IWDM-2002. Bremen, Germany. June 22-25, 2002.
- A2. Chan HP, Helvie MA, Wei J, Hadjiiski LM, Zhou C, Goodsitt MM, Sahiner B, Roubidoux MA. Automated analysis of mammographic breast density for breast cancer risk estimation. Presented at the Era of Hope Meeting, U. S. Army Medical Research and Materiel Command, Department of Defense, Breast Cancer Research Program, Orlando, Florida, September 25-28, 2002.
- A3. Wei J, Chan HP, Helvie MA, Hadjiiski LM, Sahiner B, Roubidoux MA, Zhou C, Paquerault S, Chenevert T, Goodsitt MM. Breast density estimation on mammograms and MR images: A tool for assessment of breast cancer risk. Presentation at the 88th Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL, December 1-6, 2002. Radiology 2002; 225(P): 600.
- A4. Chan HP, Wei, J, Zhou C, Helvie MA, Roubidoux MA, Bailey J, Paramagul C, Nees A, Hadjiiski LM, Sahiner B. Comparison of mammographic density estimated on digital mammograms and screen-film mammograms. Presentation at the 89th Scientific Assembly

and Annual Meeting of the Radiological Society of North America, Chicago, IL, November 30-December 5, 2003. RSNA Program 2003; 424.

- A5. Zhou C, Hadjiiski LM, Sahiner B, Chan HP, Helvie MA, Wei J. Computerized mammographic breast density estimation: Expectation-Maximization estimation and neural network classification of breast density. Presented at the 89th Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL, November 30-December 5, 2003. RSNA Program 2003; 389.
- A6. Zhou C, Chan HP, Wei J, Helvie MA, Roubidoux MA, Paramagul C, Nees A, Hadjiiski LM, Sahiner B. Performance evaluation of an automated breast density estimation system for digital mammograms and digitized film mammograms. Presented at the 7th International Workshop on Digital Mammography. IWDM-2004. Durham, North Carolina. June 18-21, 2004.
- A7. Zhou C, Hadjiiski LM, Paramagul C, Sahiner B, Chan HP, Wei J. Computerized pectoral muscle identification on MLO-view mammograms for CAD applications. Poster presentation at the SPIE International Symposium on Medical Imaging, San Diego, CA, February 12-17, 2005.
- A8. Zhou C, Chan HP, Helvie MA, Wei J, Ge J, Hadjiiski LM, Sahiner B. Computerized mammographic breast density estimation on full field digital mammogram and digitized film mammogram. Presentation at the 91st Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL. November 27-December 2, 2005. RSNA Program 2005; 271.

Computerized image analysis: Estimation of breast density on mammograms

Chuan Zhou, Heang-Ping Chan,^{a)} Nicholas Petrick, Mark A. Helvie, Mitchell M. Goodsitt, Berkman Sahiner, and Lubomir M. Hadjiiski

Department of Radiology, The University of Michigan, Ann Arbor, Michigan 48109-0030

(Received 15 September 2000; accepted for publication 4 April 2001)

An automated image analysis tool is being developed for the estimation of mammographic breast density. This tool may be useful for risk estimation or for monitoring breast density change in prevention or intervention programs. In this preliminary study, a data set of 4-view mammograms from 65 patients was used to evaluate our approach. Breast density analysis was performed on the digitized mammograms in three stages. First, the breast region was segmented from the surrounding background by an automated breast boundary-tracking algorithm. Second, an adaptive dynamic range compression technique was applied to the breast image to reduce the range of the gray level distribution in the low frequency background and to enhance the differences in the characteristic features of the gray level histogram for breasts of different densities. Third, rule-based classification was used to classify the breast images into four classes according to the characteristic features of their gray level histogram. For each image, a gray level threshold was automatically determined to segment the dense tissue from the breast region. The area of segmented dense tissue as a percentage of the breast area was then estimated. To evaluate the performance of the algorithm, the computer segmentation results were compared to manual segmentation with interactive thresholding by five radiologists. A “true” percent dense area for each mammogram was obtained by averaging the manually segmented areas of the radiologists. We found that the histograms of 6% (8 CC and 8 MLO views) of the breast regions were misclassified by the computer, resulting in poor segmentation of the dense region. For the images with correct classification, the correlation between the computer-estimated percent dense area and the “truth” was 0.94 and 0.91, respectively, for CC and MLO views, with a mean bias of less than 2%. The mean biases of the five radiologists’ visual estimates for the same images ranged from 0.1% to 11%. The results demonstrate the feasibility of estimating mammographic breast density using computer vision techniques and its potential to improve the accuracy and reproducibility of breast density estimation in comparison with the subjective visual assessment by radiologists. © 2001 American Association of Physicists in Medicine. [DOI: 10.1118/1.1376640]

Key words: mammography, computer-aided diagnosis, breast density, breast cancer risk, image segmentation, thresholding

I. INTRODUCTION

Breast cancer is one of the leading causes for cancer mortality among women.¹ One in every eight women will develop breast cancer at some point in their lives. The most successful method for the early detection of breast cancer is screening mammography. Currently, mammograms are analyzed visually by radiologists. Because of the subjective nature of visual analysis, qualitative responses may vary from radiologist to radiologist. Therefore, a computerized method for analyzing mammographic features would be useful as a supplement to the radiologist’s assessment. Previous research efforts in computer-aided diagnosis (CAD) for breast cancer detection mainly concentrated on detection and characterization of masses and microcalcifications on mammograms by using computer vision techniques. It has been demonstrated that an effective CAD algorithm can improve the diagnostic accuracy of breast cancer characterization on mammograms, which, in turn, may reduce unnecessary biopsies. In this work, we are studying the feasibility of develop-

ing a CAD system for an analysis of breast density on mammograms. Studies have shown that there is a strong positive correlation between breast parenchymal density on mammograms and breast cancer risk.^{2–9} The relative risk is estimated to be about 4 to 6 times higher for women whose mammograms have parenchymal densities over 60% of the breast area, as compared to women with less than 5% of parenchymal densities.

An important difference between breast density as a risk factor and most other risk factors is the fact that breast tissue density can be changed by dietary or hormonal interventions.^{6,10,11} Although there is no direct evidence that changes in mammographic breast densities will lead to changes in breast cancer risk, the strong correlation between breast density and breast cancer risk has prompted researchers to use mammographic density as an indicator for monitoring the effects of intervention as well as for studying breast cancer etiology.^{6,11–13}

Different methods have been used for the evaluation of

mammographic breast density. Earlier studies used a subjective visual assessment of the breast parenchyma primarily based on the four patterns described by Wolfe² (N1 is comprised entirely of fat; P1 has up to 25% nodular densities; P2 has over 25% nodular mammographic densities; DY contains extensive regions of homogeneous mammographic densities). The subjectivity in classifying the mammographic patterns introduced large variability in the risk estimation. Later studies used more quantitative estimates, such as planimetry, to measure the dense area in the breast manually outlined by radiologists on mammograms.^{3,7} These studies indicate that the percentage (%) of mammographic densities relative to the breast area can predict the breast cancer risk more accurately than a qualitative assessment of mammographic patterns. Warner *et al.*¹⁵ conducted a meta-analysis of the studies published between 1976 and 1990 to investigate the effect of different methods of classification on estimates of cancer risk. They found that the mammographic parenchymal pattern does correlate with the breast cancer risk. The magnitude of the risk varies according to the method used to evaluate the mammograms. With the quantitative estimates of mammographic density, the difference in risk between the highest and the lowest risk category is substantial and is greater than the risks associated with most other risk factors for breast cancer. More recent studies used fractal texture and the shape of the gray level histogram¹⁴ to quantify the parenchymal pattern or used interactive thresholding on digitized mammograms to segment the dense area.^{11,15} It was reported that the thresholding method provided a higher risk value than the texture measure or the histogram shape.¹⁶ Other researchers have attempted to calculate a breast density index to model the radiologists' perception.¹⁷

In clinical practice, radiologists routinely estimate the breast density on mammograms by using the BI-RADS lexicon as recommended by the American College of Radiology¹⁸ in order to provide a reference for mammographic sensitivity. Because of the lack of a quantitative method for breast density estimation, researchers often use the BI-RADS rating for monitoring responses to preventive or interventional treatment and the associated changes in breast cancer risk.¹⁹ We have found that there is a large interobserver variability in the BI-RADS ratings among experienced mammographers.^{20,21} An automated and quantitative estimation, as investigated in this study, will provide not only an efficient means to measure mammographic density, but also a reproducible estimate that will reduce the inter- and intraobserver variability of mammographic density measurements. This image analysis tool will therefore allow researchers to study more definitively the relationship of mammographic density to breast cancer risk, detection, prognosis, and mammographic sensitivity, and to better monitor the response of a patient to preventive or interventional treatment of breast cancers.

In this paper, we will describe the image processing techniques used in our automated breast density segmentation algorithm. The performance of the computer segmentation was evaluated by a comparison with the average segmenta-

tion by 5 radiologists using interactive thresholding in the same data set.

II. MATERIALS AND METHODS

A. Database

A data set consisting of 260 mammograms of 65 patients was used for the development of the histogram analysis method in this study. Each case contains the craniocaudal (CC) view and the mediolateral oblique (MLO) view of both breasts of the patient. The first 50 mammograms were consecutive screening cases from the patient files in the Radiology Department at the University of Michigan. After data analysis, it was found that there were very few dense breasts in the initial data set. An additional 15 cases visually judged by radiologists to be dense breasts were then randomly selected and mixed with the initial set. The images were processed individually without knowing their BI-RADS categories. The mammograms were acquired with mammography systems approved by the Mammography Quality Standards Act (MQSA) and were digitized with a LUMISYS 85 laser film scanner with a pixel size of $50\ \mu\text{m} \times 50\ \mu\text{m}$ and 4096 gray levels. The gray levels are linearly proportional to optical densities (O.D.) from 0.1 to greater than 3 O.D. units. The nominal O.D. range of the scanner is 0–4 with large pixel values in the digitized mammograms corresponding to low O.D. The full resolution mammograms were first smoothed with a 16×16 box filter and subsampled by a factor of 16, resulting in $800\ \mu\text{m} \times 800\ \mu\text{m}$ images of approximately 225×300 pixels in size for small films and 300×375 pixels for large films.

B. Breast segmentation and image enhancement

The breast image is first segmented from the surrounding image background by boundary detection. The detected boundary separated the breast from other background features such as the directly exposed area, patient identification information, and lead markers. The density analysis was performed only within the breast region. An automated breast boundary tracking technique developed previously^{22,23} was modified to improve its performance. Briefly, the technique used a gradient-based method to search for the breast boundary. The background of the image was estimated initially by searching for the largest background peak from the gray level histogram of the image. After subtracting this background level from the breast region, a simple edge was found by a line-by-line gradient analysis from the top to the bottom of the image. The criterion used in detecting the edge points was the steepness of the gradient of four adjacent pixels along the horizontal direction. The steeper the gradient, the greater the likelihood that an edge existed at that corresponding image point. The simple edge served as a starting point for a more accurate tracking algorithm that followed. The tracking of the breast boundary started from approximately the middle of the breast image and moved upward and downward along the boundary. The direction to search for a new edge point was guided by the previous edge points. The edge

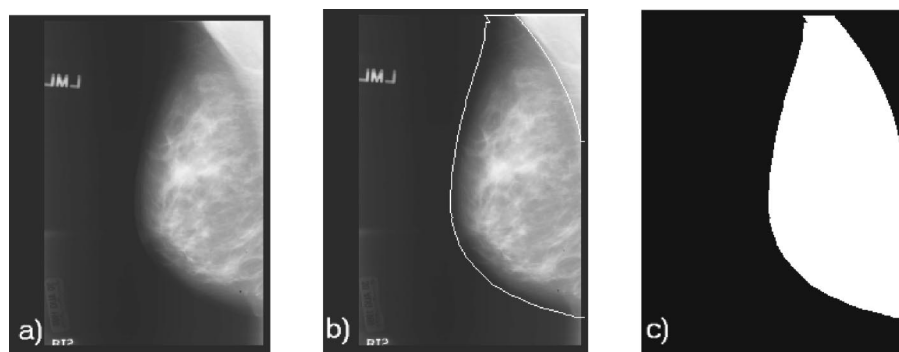


FIG. 1. (a) A mammogram from our image database; (b) the image superimposed with the detected breast boundary and pectoral muscle boundary; (c) the binary map of the segmented breast region.

location was again determined by searching for the maximum gradient along the gray level profile normal to the tracking direction. Since the boundary tracking was guided by the simple edge and the previously detected edge points, it could steer around the breast boundary and was less prone to diversion by noise and artifacts. The accuracy of the boundary tracking technique was evaluated in our previous study²³ by quantifying the root-mean-square differences between the detected and manually identified breast boundaries. In the current study, the performance of the boundary tracking technique for this data set was determined by superimposing the detected boundary on the breast image and visually judged if the detected boundary coincided with the perceived breast boundary. The breast image and its boundary were displayed by appropriately adjusting the contrast and brightness. Incomplete, jagged and mistracked boundaries were considered incorrect tracking.

The unexposed film area around the film edges was detected automatically. After the breast boundary was found, a region growing algorithm was used to fill the enclosed breast region. The result was a binary map that distinguished the breast region from the background areas. An example of the tracked breast boundary and the breast binary map is shown in Figs. 1(a)–1(c).

For the MLO view mammograms, an additional step has to be performed for segmentation of the pectoral muscle. The initial edge in the pectoral region was found as the maximum gradient point by a line-by-line gradient analysis from the chest wall to the breast boundary. The false pectoral muscle edge points were discarded by an edge validation process. First, a straight line was fitted to the initial edge points, and the points that did not lie close to the fitted line were removed. Second, the remaining edge points that were connected were identified by an 8-connectivity criterion. An edge segment was removed if its direction was inconsistent with the pectoral edge direction relative to the breast image. Finally, a second order curve was fitted to the remaining edge points to separate the pectoral muscle from the breast region. The pixels in the pectoral muscle region were excluded from the histogram analysis and breast area calculation. The accuracy of the pectoral muscle detection was also judged visually in this study, similar to the method used for the breast boundary described above. Figure 1 shows the pectoral muscle trimming result for an MLO view mammogram.

To facilitate histogram analysis, a dynamic range compression method was developed to reduce the gray level range of the histograms. With our digitization, the gray levels of the dense tissue are higher than those of the adipose tissue. Because of variations in exposure condition and breast thickness near the periphery, the gray level distribution corresponding to the breast parenchymal pattern is superimposed on a low frequency background that mainly represents the global variations in exposure. This low frequency background distorts the characteristic features of the histogram due to the density pattern. To reduce the distortion, an adaptive dynamic range compression technique was applied to the breast image. For a given breast image, $F(x,y)$, which contains low frequency background and higher frequency breast tissue structures, a smoothed image, $F_B(x,y)$, was obtained by applying a large-scale box filter to $F(x,y)$ to remove the high frequency components while retaining the low frequency components. The image $F_B(x,y)$ was then compressed by a scale factor k :

$$F_C(x,y) = kF_B(x,y). \quad (1)$$

To reconstruct the high frequency components, $F_C(x,y)$, was subtracted from a constant gray level G , and added to the original image, $F(x,y)$:

$$F_D(x,y) = G - F_C(x,y), \quad (2)$$

$$F_E(x,y) = F_D(x,y) + F(x,y). \quad (3)$$

Histogram analysis was applied to the dynamic-range-compressed image $F_E(x,y)$. Figure 2 shows an example of the resulting images and gray level histograms obtained from this procedure, where the size of box filter is 35×35 , the scale factor k is 0.5, and the constant gray level G is the maximum gray level of the compressed image $F_C(x,y)$. The values of these parameters were chosen experimentally as a balance between reducing the dynamic range and preserving the image features in the compressed image.

C. Breast density segmentation and estimation

A rule-based threshold technique was developed to segment the dense areas from the breast background. The histogram of the breast region on the dynamic-range-compressed mammogram was generated and smoothed. The histograms of these images in the database were analyzed to formulate an automatic thresholding routine. The histograms were

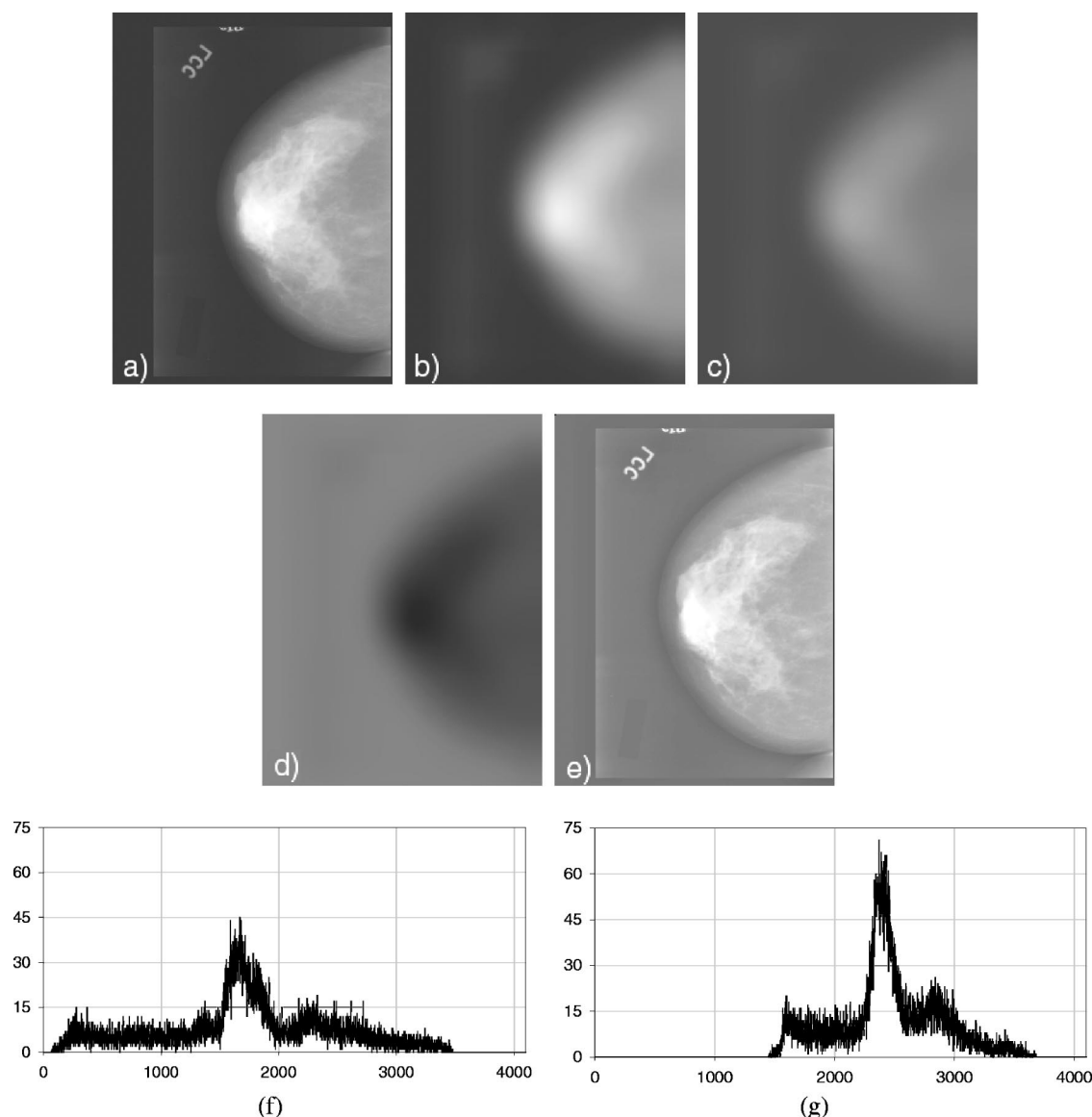


FIG. 2. (a) A typical mammogram from our image database; (b) the low frequency image $F_B(x,y)$ obtained by an 35×35 box filter; (c) the compressed image $F_C(x,y)$; (d) the inverted image $F_D(x,y)$; (e) the enhanced image $F_E(x,y)$; (f) the gray level histogram within the breast region of the original image $F(x,y)$; and (g) the gray level histogram of the breast region of the enhanced image $F_E(x,y)$.

grouped into four classes based on the characteristic shapes of their histograms. It was observed that the grouping corresponded approximately to the four BI-RADS breast density ratings: Class I corresponded to breasts of almost entirely fat, Class II corresponded to scattered fibroglandular densities, Class III corresponded to heterogeneously dense and Class IV corresponded to extremely dense breasts. Examples of typical histograms for these four classes are shown in Fig. 3. The histograms seemed to follow two basic patterns. In one pattern, there was only one dominant peak, which represented most of the breast structures in the breast region. In the other pattern, in addition to a large peak in the histogram, there was one or two smaller peaks on the right or left side of the large peak. In a majority of the cases, the smaller peak was distinguishable from the large one when the random fluctuation on the histogram was smoothed.

1. Peak detection and feature description

The gray level histogram within the breast area was generated and normalized, and passed through an averaging window to smooth out the random fluctuations. We estimated the window size to be in the range of 30 to 50 gray levels by experimentally evaluating the histogram shapes and density segmentation at different window sizes. Too small a window size cannot smooth out the fluctuation and too large a window size will blur the useful features. A window size of 30 was used in this study. The second derivative of every point on the histogram curve was computed. An example of the histogram and its second derivative curve are shown in Fig. 4. The zero crossing locations were detected by scanning for the positive-to-negative and negative-to-positive changes on the latter curve. If the second derivative was negative be-

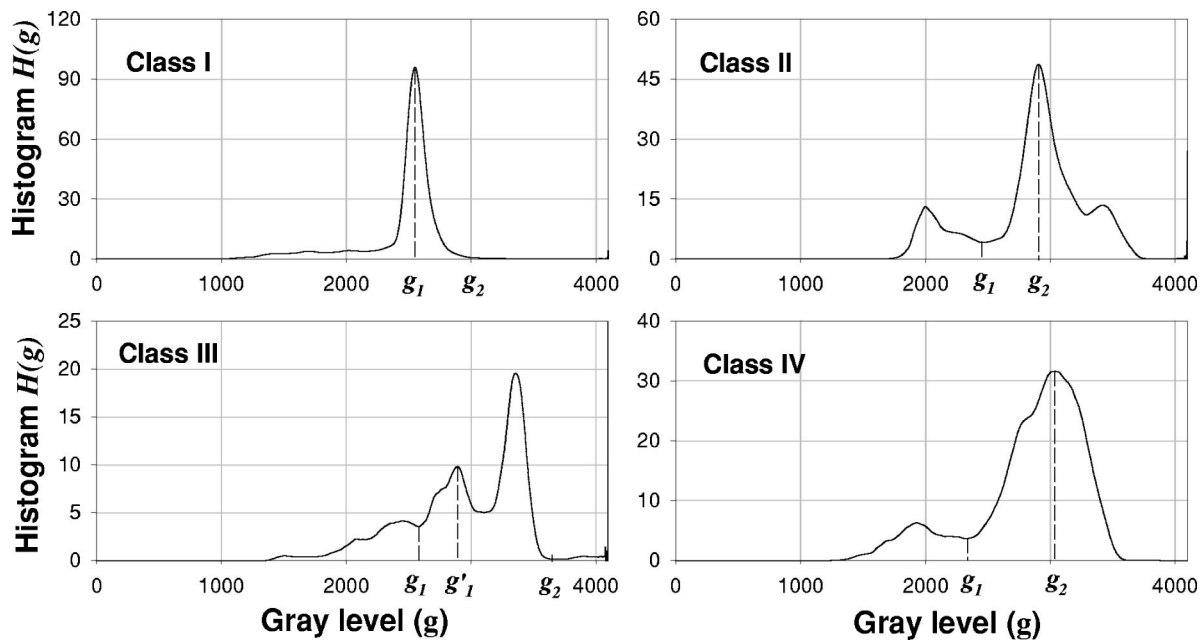


FIG. 3. Four typical classes of histograms and the setting of gray level interval $[g_1, g_2]$ for the threshold calculation.

tween two zero crossing points, it indicated that a peak existed between these two points on the histogram. Normally, as shown in Fig. 4, a peak included the peak point P_0 and two valley points P_1 and P_2 located on the two sides of the peak point. The peak point P_0 was determined by searching for the maximum histogram value between the zero crossing points Z_2 and Z_3 , and the P_1 and P_2 points were obtained by searching for the point with minimum histogram value between zero crossing points Z_1, Z_2 and Z_3, Z_4 , respectively.

The following peak features can be defined by peak point P_0 and valley points P_1 and P_2 :

$$\text{Energy: } E = \frac{1}{A} \sum_{i=P_1}^{P_2} f(i) * f(i), \quad (4)$$

$$\text{left-side energy: } E_L = \frac{1}{A} \sum_{i=P_1}^{P_0} f(i) * f(i), \quad (5)$$

$$\text{right-side energy: } E_R = \frac{1}{A} \sum_{i=P_0}^{P_2} f(i) * f(i), \quad (6)$$

$$\text{likelihood: } L = E/E', \quad (7)$$

where $f(\cdot)$ is the histogram, A is the total energy of the entire histogram and $A = \sum_{i=0}^N f(i) * f(i)$, N is the maximum gray level of the histogram. E' is the energy calculated by approximating the histogram in the interval $[P_1, P_2]$ using two straight lines, P_1P_0 and P_0P_2 . The energy E of the peak is used to compare the sizes of the peaks on the histogram, higher energy means bigger size of the peak. E_L and E_R split the energy E into two parts from the peak point for calculating the ratio of the energy in these two parts. The likelihood L describes how close the real peak is to the triangle represented by the three points P_0 , P_1 and P_2 .

2. Rule-based histogram classification

A rule-based histogram classifier was developed to classify the gray level histogram of the breast area into four classes. As shown in Fig. 3, a typical Class I breast is almost entirely fat, it has a single narrow peak on the histogram. Class II has scattered fibroglandular densities, it has two peaks, other than the tail part on the left, on the histogram, with the smaller peak on the right of the bigger one. Class III is heterogeneously dense, it also has two peaks, but the smaller peak is on the left of the bigger one. Class IV is extremely dense, which has a single dominant peak on the histogram, but it is wider compared with the peak in the Class I histogram, and a second small peak sometimes occurs to the left of the main peak.

The classification is performed in two steps. In the first step, the computer determines whether there is only one single peak in the histogram. The biggest peak (main peak)

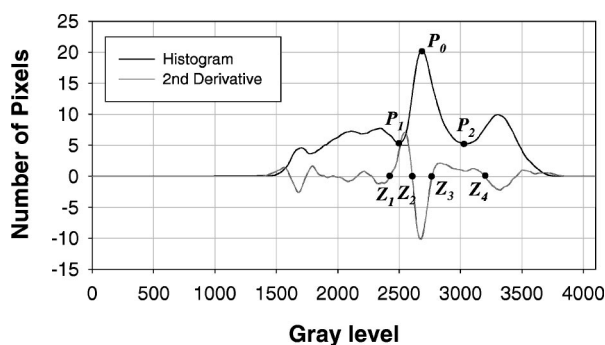


FIG. 4. The gray level histogram (solid curve) and the second derivative (dot curve). P_0 is the peak point, P_1 and P_2 are the valley points of the peak on the two sides of the peak point P_0 . Points Z_1 , Z_2 , Z_3 and Z_4 are zero crossing points on the second derivative curve, which are used for searching the points P_0 , P_1 and P_2 .

P_M and its location are detected by comparing the energy of the peaks on the histogram. The single peak feature is mainly determined by the energy E under the main peak and the features E_L and E_R . If the histogram is found to have a single-peak pattern, in general, a narrow peak corresponds to very fatty breast (Class I), and a wider peak corresponds to very dense breast (Class IV). However, in some cases, the histogram of these two classes is very similar, as discussed below (Fig. 9), and it is difficult to distinguish them by their gray level histogram distributions. Two additional image features were analyzed to classify very fatty and very dense breasts. One feature is the gray level standard deviation (Std) in the entire breast area, defined as

$$\text{Std} = \left(\frac{1}{N} \sum_{x \in \text{MAP}} \sum_{y \in \text{MAP}} (f(x,y) - \bar{f}(x,y))^2 \right)^{1/2}, \quad (8)$$

where MAP is the breast binary map region, N is the pixel numbers within MAP. Another feature is the number of single pixels and single pixel-size holes (NSH) counted in the breast area of a segmented binary image using the biggest histogram peak point P_M as a threshold. For a very fatty mammogram, the breast mainly consists of a fatty background with some fibrous structures and fibroglandular tissue scattered in the breast area. The NSH value was found to be larger (greater than 50 pixels on average), and Std smaller (less than 500 on average), compared with a mammogram of a very dense breast.

In the second step, if the histogram is found to have more than one peak, decision rules are used to decide if the second major peak is on the left side or on the right side of P_M by the features E , E_L , E_R and L , and the relative position of the two peaks. If the second major peak is on the right, then the histogram is classified to be Class II; otherwise, it is classified to be Class III.

3. Gray level thresholding

Gray level thresholding is essentially a pixel classification problem. Its objective is to classify the pixels of a given image into two classes: one includes pixels with gray values that are below or equal to a certain threshold; the other includes those with gray values above the threshold. Thresholding is a popular tool for image segmentation, a variety of techniques have been proposed over the years. In our study, two threshold selection methods are used: one is the Discriminant Analysis (DA) method²⁴ and the other is the Maximum Entropy Principle (MEP) based method.²⁵ The DA method assumes that the image gray levels can be classified into two classes by a threshold. To estimate the threshold, a discriminant criterion based on the within-class variance and between-class variance is introduced. An optimal threshold is selected by the discriminant criterion to maximize the separability of the resultant classes in terms of gray levels. This method is well-suited for the cases where the gray level histogram is bimodal. In an ideal situation, the histogram has a deep and sharp valley between the two peaks representing objects and background, respectively, and the optimum corresponds to the gray level at the bottom of this valley. A

more detailed description of the DA method can be found in Appendix A.

For the MEP method, the optimal threshold value is determined by maximizing the *a posteriori* entropy subject to certain inequality constraints that are derived by means of special measures characterizing the uniformity and the shape of the regions in the image. As is well-known,²⁶ the maximum *a posteriori* probability can serve as a criterion to select *a priori* probability distributions when very little is known about the probability distribution. Compared with the DA method, MEP can provide a better thresholding result if the gray level histogram does not have a bimodal distribution. A more detailed description of the MEP method can be found in Appendix B.

The gray level histograms of the mammograms in our study are very complex, the histogram may be unimodal, bimodal or multi-modal. It is difficult to select an appropriate threshold by one general threshold selection method. Therefore, we combined both the DA and the MEP methods, to select a threshold according to the characteristic features of the histogram that has been classified into one of the four classes. Suppose $f(g)$ is the gray level histogram of the breast area. Let $T = \text{Method}(f(g)|g_1 < g < g_2)$ represent the threshold, T , that is selected by use of Method in the interval $[g_1, g_2]$ of the histogram $f(g)$, where Method can be either the DA or MEP method. The settings of the interval $[g_1, g_2]$ for the four classes are discussed below and shown in Fig. 3.

Class I: The histogram is unimodal so that the threshold is selected as

$$T = \text{MEP}(f(g)|g_1 < g < g_2),$$

where, g_1 is the main peak point; g_2 is the valley point on the right side of main peak.

Class II: The histogram is not unimodal and the histogram is classified as Class II; the threshold is selected by averaging two thresholds that are computed in two different intervals of the histogram by the DA method:

$$T_1 = \text{DA}(f(g)|g > g_1),$$

$$T_2 = \text{DA}(f(g)|g > g_2),$$

$$T = (T_1 + T_2)/2,$$

where g_1 is the valley on the left of the main peak; g_2 is the main peak point.

Class III: The histogram is not unimodal; there are two possibilities in the histogram distribution: there is a valley between the main peak and its left side peak, as shown in Fig. 3, or no obvious valley exists between the main peak and its left side peak. In two different intervals of the histogram, two thresholds are computed as

$$T_1 = \text{DA}(f(g)|g_1 < g < g_2),$$

$$T_2 = \text{DA}(f(g)|g'_1 < g < g_2),$$

where g_1 is the left valley point of the left-side peak (P_{LM}) of the main peak, g'_1 is the peak point of P_{LM} and g_2 is right valley point of the main peak. If there is an obvious valley, $T = (T_1 + T_2)/2$, otherwise $T = T_1$.

Class IV: Since the histogram is considered unimodal, the threshold is computed by the MEP method, $T = \text{MEP}(f(g)|_{g_1 < g < g_2})$, where, g_1 is the left valley point of the main peak; g_2 is the main peak point.

D. Radiologists' segmentation of dense breast tissue

In order to evaluate the accuracy of the computer segmentation method, the computer segmentation results were compared to radiologists' manual segmentation in the data set of 65 patient cases. Details of the observer study for estimation of the breast density and statistical analysis of the results were discussed elsewhere.²⁷ Briefly, a graphical interface was developed for displaying the mammograms and recording the observer's evaluation. The CC-view and MLO-view mammograms for a given breast were displayed side-by-side; a radiologist observer examined the mammograms and gave a BI-RADS rating and a visual estimation of the percent breast density with 10% increments. After the subjective evaluation, each view was displayed sequentially, together with the histogram of the dynamic-range-compressed image. The radiologist would interactively choose a threshold by moving a slider along the abscissas of the histogram plot. The segmented binary image, displayed side-by-side with the mammogram, would change instantaneously when the threshold was changed. The radiologist could inspect if the segmented area corresponded to the dense area on the mammogram. Once the radiologist was satisfied with the segmentation of the dense area, the gray level threshold and the percent dense area derived from this threshold were recorded. The display then moved to the next view of the same breast for evaluation. The mammograms of the other breast for the same patient would then be displayed and evaluated in the same way. The entire process was repeated for each patient until all patients in the data set were evaluated.

Five MQSA-approved radiologists participated in the experiment. To familiarize the radiologists with the procedures and to assist them in their visual estimation of the percent breast density, we had them trained on a separate set of 25 patient cases prior to the evaluation of the actual data set. During the training session, the computer displayed the percent breast dense area to the radiologist, which was obtained by the radiologist's interactive thresholding of the image. The radiologist could then compare the manually segmented percentage with their visually assessed percent density for the image. This feedback helped "calibrate" the radiologists' visual estimates of the percent dense breast area. The percent dense area obtained by interactive thresholding was not displayed during the actual study.

III. RESULTS

An example of a typical mammogram from each of the four classes and its corresponding enhanced image, its histogram, the selected threshold and the segmented image are shown in Figs. 5(a)–5(d), respectively.

The average percent breast density obtained from manual segmentation by the five trained radiologists for each mammogram was used as the "true standard" of the percent

breast density for that mammogram. The breast region was segmented by the breast boundary tracking technique, and the pectoral muscle was trimmed for the MLO-view mammograms. The breast boundary was accurately tracked on 92.3% (240/260) of the mammograms, and the pectoral muscle was correctly trimmed on 74.6% (97/130) of the MLO views. The histograms of 6% (8 CC views and 8 MLO views) of the breast regions did not exhibit the typical characteristic features of the four classes and were misclassified by the computer, resulting in poor segmentation of the dense region.

Figure 6 shows a comparison of the percent breast density visually estimated by radiologists against the true standard for the 94% of the 260 mammograms that were classified correctly by the computer. Table I summarizes the comparison of the radiologists' visual estimates with the true standard. The "difference" between the estimated % breast density and the true standard was calculated for each case, and the mean and the standard deviation of this difference over all cases were estimated for each radiologist and shown in the table. Therefore, the mean difference was the average bias of the estimated % breast density from the true standard over all images in the data set. It can be seen that almost all radiologists had a positive bias, on average, when they visually estimated mammographic density, except for Radiologist 5 who had a small negative average bias on the CC-view reading. For a given radiologist, the over-estimation increased as the breast density increased. Although the correlation coefficients were high, ranging from 0.90 to 0.95, the deviations from the diagonal line were systematic. The average bias from the true standard varied from less than 1% to 11%, depending on the radiologist. The root-mean-square (RMS) errors of the five radiologists relative to the true standard ranged from 7.5% to 16.3%.

Figure 7 shows the comparison of the percent breast density between the computer segmentation and the true standard for the 94% of mammograms whose histograms were considered to be correctly classified. There was a trend of over-estimation in the very fatty breasts. In the medium dense range, the variances from the true standard were high. Some images had a large deviation from the diagonal line, indicating that the threshold was incorrectly determined. Table II summarizes the comparison between the computer performance and the true standard. For the CC views with correct histogram classification, the correlation between the computer-estimated percent dense area and the true percent breast density was 0.94, and between the computer and the radiologists' average visual estimate was 0.87 (not plotted). These correlation coefficients were 0.91 and 0.82, respectively, for the MLO views with correct classification. Although the correlation coefficients of the computer segmentation with the true standard were not better than those of the visual estimates, the average biases of the computer segmentation from the true standard were less than 2%, which were substantially less than those of visual estimates (Table I). This indicates that computerized segmentation is a good alternative to manual segmentation although variances of the automated method will need to be further reduced. The RMS

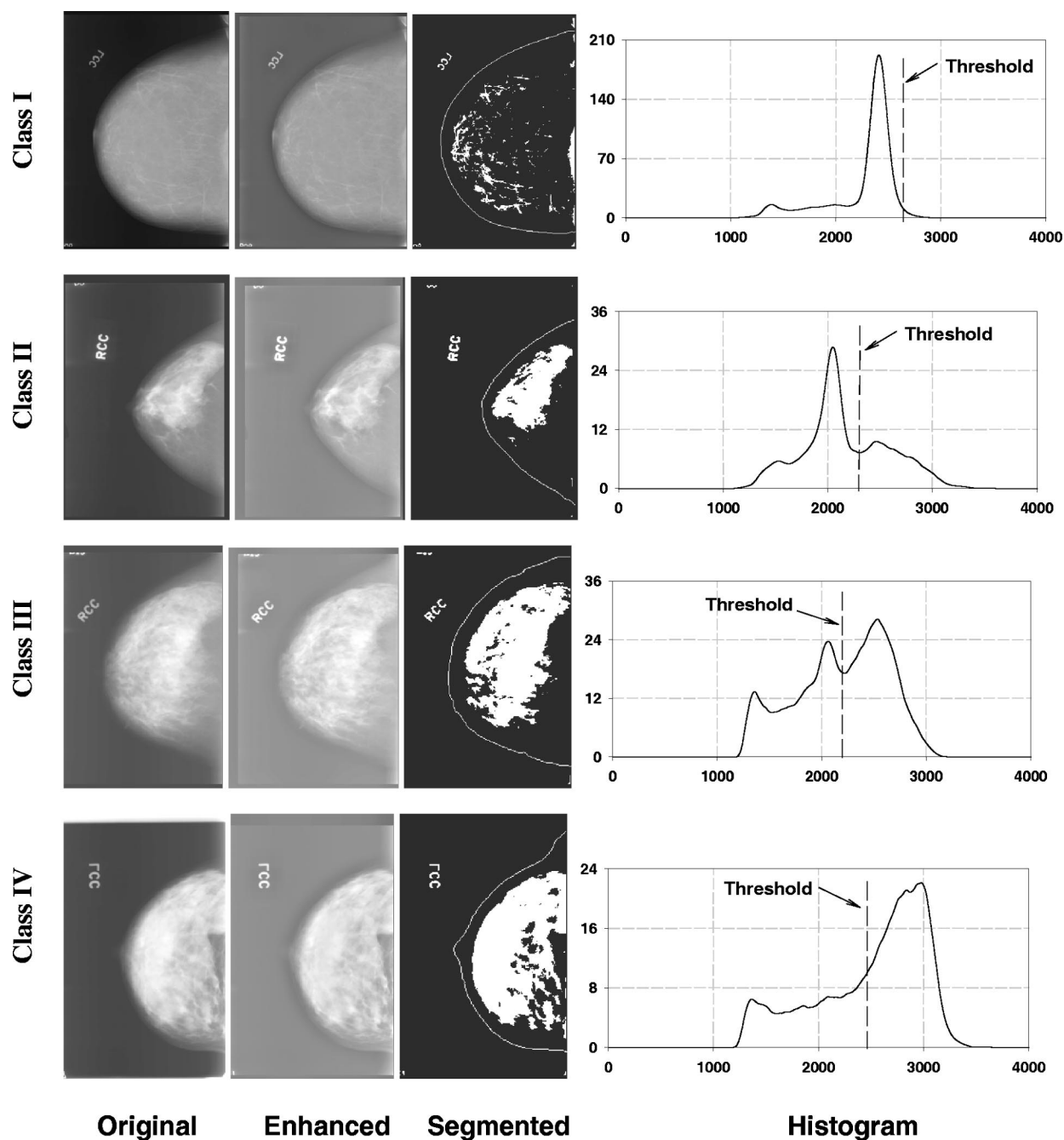


FIG. 5. Four classes of typical mammograms and corresponding enhanced and segmented image, histogram and threshold.

errors of the computer segmentation were also less than those of the radiologists' visual estimates, at 6.1% and 7.2%, respectively, for the CC view and MLO view, when the histograms were correctly classified. The biases and RMS errors for the different subsets of images are also shown in Table II. It can be seen that correct histogram classification was the most important factor in reducing the biases and the RMS errors. The contributions by breast boundary detection and pectoral muscle segmentation were minor, on average, for improving the estimation of the percent dense breast area.

Figure 8 shows the comparison of the individual radiologists' manual segmentation against the true standard. For CC views, the RMS difference in the percent breast density be-

tween an individual radiologist's manual segmentation and the true standard varied from 2.9% to 5.9% among the five radiologists. For MLO views, the RMS difference varied from 2.8% to 6.2%. The average biases of the five radiologists ranged from -2.8% to 2.2% for the CC views and from -3.1% to 3.0% for the MLO views. The maximum biases of the five radiologists varied from 4.4% to 22.6% for the CC views and from 5.2% to 23% for the MLO views.

The five radiologists provided BI-RADS density ratings for each breast. Although the BI-RADS ratings exhibited large inter-observer variations,²⁰ it is interesting to compare the computer's histogram classification with the BI-RADS ratings. Since there were 260 images, each with 5 radiolo-

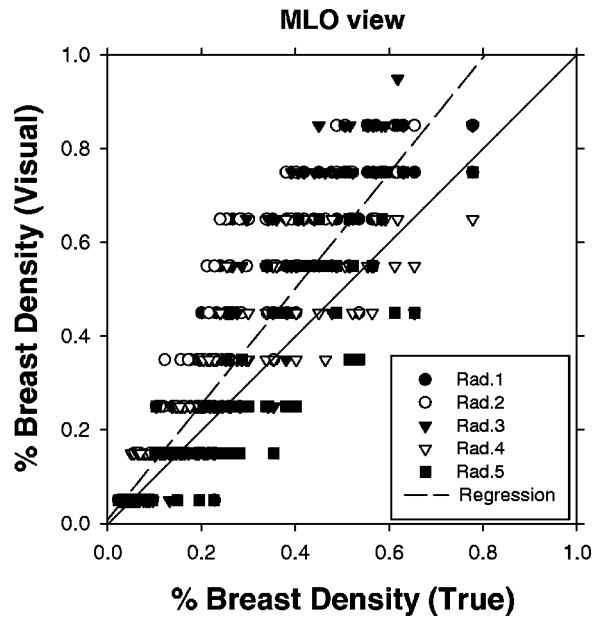


FIG. 6. A comparison of the percent breast density between five radiologists' visual estimates and the true standard. The dashed line represents the linear regression of all data points on the plot. The MLO view is shown. The trend for the CC view is similar.

gists' ratings, there were a total of 1300 rating comparisons. The comparison of the computer and the radiologists' BI-RADS ratings is shown in Table III. It was found that 87.4% of Class I classification have BI-RADS ratings 1 or 2, 92.0% of Class II classifications have density ratings 2 or 3, 83.4% of Class III classifications have density ratings 3 or 4

and 57.1% of Class IV classifications have density rating 4. More detailed analysis of the variability of radiologists' BI-RADS ratings was discussed by Martin *et al.*²¹

IV. DISCUSSION

Radiologists routinely estimate mammographic breast density using the four BI-RADS categories. In studies that require breast density estimation, radiologists' visual estimates of mammographic density were often used as the density measure. Our observer study indicates that interobserver variation between the BI-RADS ratings of five experienced radiologists ranged from -1 to $+1$. The subjectively estimated percent dense area can deviate from the true standard by as much as 40%, as shown in Fig. 6. These results indicate the need to develop an objective method for the estimation of mammographic breast density in order to improve the accuracy and reproducibility of the estimation. A computerized image analysis method for mammographic breast density estimation will be a useful tool for study of breast cancer risk factors and for monitoring the change of breast cancer risk with preventive or interventional treatments.

In this study, we used the average of the percent breast area obtained with interactive thresholding by five experienced radiologists as the true standard. The gray level thresholding method used in this study could achieve a reasonable segmentation of the dense areas on the mammogram because the image was preprocessed with dynamic range compression. The image-based analysis of breast density will not provide the actual percentage of fibroglandular tissue in the breast volume. However, the previous studies that estab-

TABLE I. A comparison of the radiologists' visual estimate of mammographic breast density with the true standard. The "difference" was defined as the difference between the estimated % breast density and the true standard for each case, and the mean and the standard deviation of this difference are tabulated.

Image subsets	No. of images	Radiologist	Correlation	RMS error	Mean difference	Std. dev. of difference
CC view:						
All	130	Rad. 1	0.942	13.3%	6.9%	11.5%
		Rad. 2	0.931	14.5%	9.8%	10.7%
		Rad. 3	0.923	13.3%	6.3%	11.8%
		Rad. 4	0.934	7.5%	2.9%	7.0%
		Rad. 5	0.901	9.6%	−1.4%	9.6%
Histogram correctly classified	122	Rad. 1	0.946	13.7%	7.2%	11.3%
		Rad. 2	0.936	14.7%	10.3%	10.8%
		Rad. 3	0.929	14.2%	6.7%	11.6%
		Rad. 4	0.929	7.7%	3.1%	7.1%
		Rad. 5	0.900	9.7%	−1.3%	9.4%
MLO view:						
All	130	Rad. 1	0.933	14.5%	8.3%	12.0%
		Rad. 2	0.914	16.1%	11.2%	11.5%
		Rad. 3	0.915	14.4%	7.7%	12.2%
		Rad. 4	0.919	8.8%	4.3%	7.7%
		Rad. 5	0.910	9.2%	0.1%	9.2%
Histogram correctly classified	122	Rad. 1	0.932	15.0%	8.3%	12.0%
		Rad. 2	0.914	16.3%	10.9%	11.4%
		Rad. 3	0.919	14.7%	7.8%	12.2%
		Rad. 4	0.916	9.0%	4.3%	7.7%
		Rad. 5	0.909	9.4%	0.3%	9.2%

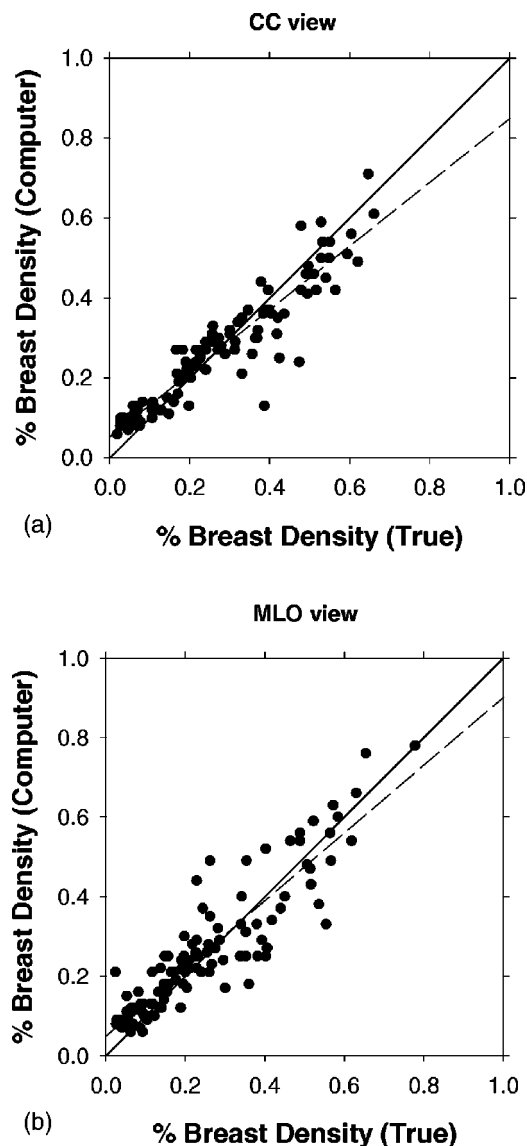


FIG. 7. A comparison of the percent breast density between the computer segmentation and the true standard. The dashed line represents the linear regression of the data on the plot. (a) CC view, (b) MLO view.

lished the correlation between breast density and breast cancer risk were all based on mammographic density. This indicated that mammographic density is a sufficiently sensitive marker for breast cancer risk, although it may be less accurate than volumetric density. An actual measurement of the percentage of fibroglandular tissue volume in the breast, for example, by x-ray penetration with correction for scatter and beam hardening, is difficult because it requires accurate x-ray sensitometry or phantom calibration for each image. These requirements will limit its use to a few laboratories that have specialized equipment and expert physicists. Magnetic resonance breast imaging can also provide volume measurement of dense tissue but it is expensive and not easily accessible. It can be expected that the estimation of mammographic breast density by a computerized image analysis method will be a more practical and viable approach, espe-

cially when direct digital mammography becomes more widely used in the future.

Our preliminary study indicates that breast density estimation can be performed automatically and accurately (Fig. 7). Although the accuracy of our current algorithm still needs to be improved, it can be seen that the computer segmentation can provide an estimate of the percent breast density with a very small bias (Table II). More importantly, computer segmentation will be more reproducible and consistent than visual estimates. This will improve the sensitivity of studies that depend on evaluation of the change in mammographic density over time or before and after a certain treatment.

In this study, we reduced the spatial resolution to a pixel size of $800\ \mu\text{m} \times 800\ \mu\text{m}$ for image processing. The small matrix size of the reduced images improves the computational efficiency. The reduction in resolution has two major effects: reducing the image noise and blurring the details. Since the significant dense tissue in the breast that contributes to the parenchyma is relatively large compared to $800\ \mu\text{m}$, it is not expected that processing at this pixel size will have a strong effect on the accuracy of the estimated percent breast density. Differences in the segmented area may occur mainly along the boundary of the dense tissue region, but the effect may be averaged out statistically along boundaries of reasonable lengths. The residual errors in the estimation of the dense area should not be substantial in comparison with the inter- and intra-radiologists' variations in their manual segmentation.

Successful segmentation of dense tissue depends strongly on whether a mammogram can be classified correctly into a proper class. A successful classification will likely result in the selection of a near optimal threshold. Conversely, if a mammogram is classified into a wrong class, the threshold will be selected incorrectly. For the mammograms of very fatty breasts, the gray level histogram has the characteristics of Class I, which contains one large single peak. These histograms can be distinguished relatively easily from most of the other classes of histograms if those histograms exhibit the typical features. For mammograms of BI-RADS category 2 or 3, there are scattered fibroglandular or heterogeneous densities in the breast. A small peak may be located on the left or on the right, or on both sides of the main peak on the histogram. The histogram could be classified into Class I if the small peak is not large enough and is not detected as a second peak. Otherwise, it would be classified into Class II or Class III, depending on the location of that small peak relative to the main peak of the histogram. For the two-peak pattern histogram, the DA threshold selection method is robust if there is an obvious valley between the two peaks. If the valley is flat or not obvious, averaging the two thresholds obtained by the DA method in two different intervals, as designed for this study, can reduce the chance of calculating an incorrect threshold that differs greatly from the optimum, but it also reduces the chance of finding the optimal threshold. Overall, the rules designed for classification of the two-peak patterns seem to perform consistently well for this data set. One of the difficult situations is to distinguish between

TABLE II. A comparison of computer segmentation with the true standard. The “difference” was defined as the difference between the estimated % breast density and the true standard for each case, and the mean and the standard deviation of this difference are tabulated.

Image subsets	No. of images	Correlation	RMS error	Mean difference	Std. dev. of difference
CC view:					
All	130	0.746	12.3%	1.3%	12.3%
Boundary correctly tracked	120	0.780	11.4%	1.4%	11.4%
Histogram correctly classified	122	0.943	6.1%	0.2%	6.2%
Boundary and histogram correctly done	113	0.953	5.6%	0.8%	5.6%
MLO view:					
All	130	0.780	11.6%	1.9%	11.5%
Boundary correctly tracked	120	0.766	11.9%	2.1%	11.7%
Histogram correctly classified	122	0.914	7.2%	1.5%	7.1%
Pectoral muscle correctly trimmed	97	0.733	11.6%	1.6%	11.6%
Boundary and histogram correctly done	112	0.912	7.2%	1.7%	7.1%
Boundary, histogram and pectoral muscle correctly done	83	0.891	7.1%	1.9%	6.8%

Class I and Class IV, when the histogram of a very dense breast mimics that of a very fatty breast, as shown in Fig. 9. This image was correctly classified with the additional features, Std and NSH. However, there were other cases that failed in spite of the additional criteria. The large difference in the optimal threshold locations between these two classes will lead to a large error in the estimated percent breast density if the histogram is misclassified. Further study is needed to more accurately distinguish these two classes.

The dynamic range reduction technique reduces the variability of the gray level histograms and enhances their characteristics. This pre-processing facilitates the classification of the image into the correct class. There are many image smoothing techniques published in the literature. Low-pass filtering with a box filter is the simplest choice. The effective-

ness of background correction with a box filtered image depends on the box size. We found that a 35×35 -pixel filter is a good balance between computation time and the capability to remove the high frequency components. The subtraction of the low-pass filtered image from the original image is a form of unsharp masking. The breast boundary is generally enhanced as shown in Fig. 2(e). The pixels at the enhanced breast boundary contribute a small peak to the left tail of the gray level histogram of the breast area. Moreover, if dense tissue is present close to the breast boundary, it may not be segmented correctly due to intensity reduction. Other low frequency estimation techniques such as wavelet decomposition will be investigated in future studies.

In this feasibility study, we used a small data set of mammograms to develop a rule-based classifier for the histogram analysis. Although a large fraction of the histograms manifest characteristic features that can be grouped into four classes, corresponding approximately to the four BI-RADS breast density ratings, there are many exceptions. One such example is shown in Fig. 9. This causes misclassification and incorrect thresholding by the histogram classifier. It will be

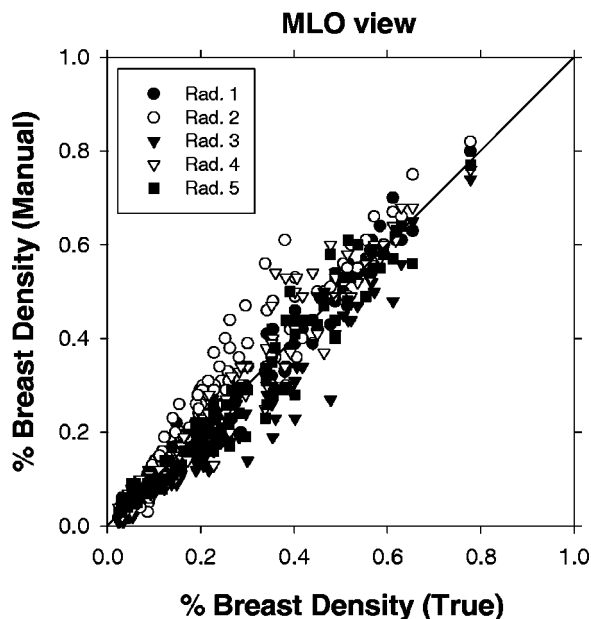


FIG. 8. A comparison of the percent breast density obtained from the five radiologists' manual segmentation with their average for the same mammograms. The MLO view is shown. The trend for the CC view is similar.

TABLE III. A comparison of computer classification and radiologists' BI-RADS breast density ratings.

Computer classification	BI-RADS 1	BI-RADS 2	BI-RADS 3	BI-RADS 4	Total
Class I	210 (16.2%)	262 (20.2%)	52 (4%)	16 (1.2%)	540 (41.5%)
Class II	0 (0%)	92 (7.1%)	184 (14.2%)	24 (1.8%)	300 (23.1%)
Class III	1 (0.1%)	52 (4%)	167 (12.8%)	100 (7.7%)	320 (24.6%)
Class IV	5 (0.4%)	12 (0.9%)	43 (3.3%)	80 (6.2%)	140 (10.8%)
Total	216 (16.6%)	418 (32.2%)	446 (34.3%)	220 (16.9%)	1300 (100%)

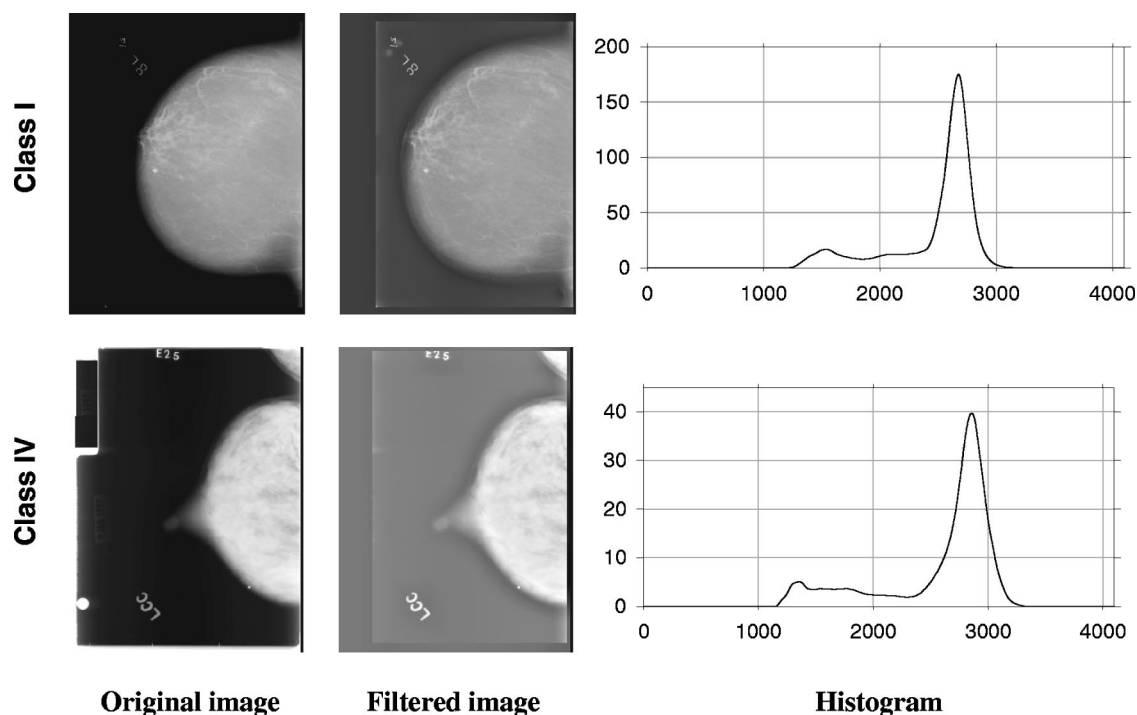


FIG. 9. The gray level histograms of two mammograms classified by radiologists as BI-RADS rating 1 (upper mammogram) and BI-RADS rating 4 (lower mammogram). The shapes of the histograms are very similar and cannot be distinguished by our current histogram analysis method. These two examples were correctly classified with the additional Std and NSH criteria.

necessary to investigate if other classification strategies can be more effective than a rule-based method. Furthermore, we have not performed a systematic study to optimize the many parameters used in the segmentation algorithm. Further work will be required to investigate the dependence of the segmentation accuracy on the various parameters. The parameter selection and the performance of the computer classifier will have to be improved by training with a larger data set and its generalizability evaluated with unknown cases. The generalization of the algorithm to images acquired with other digitizers or direct digital mammography systems will also need to be investigated.

V. CONCLUSION

We are developing an image analysis method for automated segmentation of the dense area from mammograms and estimation of the percent mammographic density. Our preliminary study indicates the feasibility of our approach. The computer-estimated mammographic breast density correlate closely with the average manual segmentation by five experienced radiologists and the average bias is much less than that of the radiologists' visual estimation. We have found that correct classification of the histogram shapes is the most crucial step in our approach. The histograms of many mammograms have distinctive characteristics that can be recognized by a rule-based classifier. However, some histograms deviate from these rules and this can lead to misclassification. A further investigation will be needed to design more robust rules or classifiers to improve the classification accuracy. Despite these limitations, we have

demonstrated in this preliminary study that the estimation of mammographic density can be performed efficiently and accurately by the automated image analysis tool. The fully automated algorithm can provide an objective and reproducible quantitative estimation of mammographic breast density that is expected to be superior to subjective visual assessment and comparable to manual segmentation by radiologists.

ACKNOWLEDGMENTS

This work is supported by USPHS Grant No. CA48129, U.S. Army Medical Research and Material Command grants (DAMD 17-99-1-9294) and (DAMD 17-01-1-0326) and by a Career Development Award (B.S) from the USAMRMC (DAMD 17-96-1-6012). The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned in this paper should be inferred.

APPENDIX A: GRAY-LEVEL THRESHOLDING—DISCRIMINANT ANALYSIS (DA) METHOD

Suppose the probability of the gray level n_i in an image with L gray levels can be estimated as

$$p_i = n_i / N, \quad N = \sum_{i=1}^L n_i. \quad (\text{A1})$$

If the pixels in the image are classified into two classes C_0 and C_1 by the threshold k , then the probabilities of class occurrence and the class mean levels are given by

$$\omega_0 = \sum_{i=1}^k p_i = P(k), \quad \omega_1 = \sum_{i=k+1}^L p_i = 1 - P(k), \quad (\text{A2})$$

$$\mu_0 = \sum_{i=1}^k i p_i / \omega_0 = \mu(k) / \omega_0, \quad (\text{A3})$$

$$\mu_1 = \sum_{i=k+1}^L i p_i / \omega_1 = \frac{\mu_T - \mu(k)}{1 - P(k)},$$

where

$$P(k) = \sum_{i=1}^k p_i, \quad \mu(k) = \sum_{i=1}^k i p_i \quad \text{and} \quad \mu_T = \sum_{i=1}^L i p_i, \quad (\text{A4})$$

are the zeroth- and the first-order cumulative moments of the histogram up to the k th level, and the total mean level of original image, respectively.

The between-class variance is defined as

$$\sigma_B^2(k) = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2$$

$$= \omega_0 \omega_1 (\mu_1 - \mu_0)^2 = \frac{[\mu_T P(k) - \mu(k)]^2}{P(k)[1 - P(k)]}, \quad (\text{A5})$$

and the optimal threshold k^* is given by

$$\sigma_B^2(k^*) = \max_{1 \leq k \leq L} \sigma_B^2(k). \quad (\text{A6})$$

APPENDIX B: GRAY-LEVEL THRESHOLDING—MAXIMUM ENTROPY PRINCIPLE (MEP) METHOD

Suppose the probability of the gray level n_i in an image with L gray levels can be estimated as

$$p_i = n_i / N, \quad N = \sum_{i=1}^L n_i. \quad (\text{B1})$$

After thresholding the image by threshold k , the *a posteriori* probability of the pixels with gray level value less than k , is given by

$$F(k) = \sum_{i=0}^k p_i. \quad (\text{B2})$$

And the *a posteriori* probability of all those pixels with values greater than or equal to k is $1 - F(k)$. Thus the Shannon entropy of the thresholded image is

$$H(F(k)) = -F(k) \log F(k) - (1 - F(k)) \log(1 - F(k)). \quad (\text{B3})$$

The optimal threshold k maximizes $H(F(k))$.

^{a)}Correspondence: Heang-Ping Chan, Ph.D., University of Michigan, Department of Radiology, 1500. Medical Center Drive, UHB1F510B, Ann Arbor, Michigan 48109-0030. Electronic mail: chanhp@umich.edu

¹S. H. Landis, T. Murray, S. Bolden, and P. A. Wingo, "Cancer statistics, 1998," *Ca-Cancer J. Clin.* **48**, 6–29 (1998).

²J. N. Wolfe, "Breast patterns as an index of risk for developing breast cancer," *Am. J. Roentgenol.* **126**, 1130–1139 (1976).

- ³J. N. Wolfe, A. F. Saftlas, and M. Salane, "Evaluation of mammographic densities: A case-control study," *Am. J. Roentgenol.* **148**, 1087–1092 (1987).
- ⁴N. F. Boyd et al., "Mammographic patterns and breast cancer risk: methodologic standards and contradictory results," *J. Natl. Cancer Inst.* **72**, 1253–1259 (1984).
- ⁵A. F. Saftlas and M. Szklo, "Mammographic parenchymal patterns and breast cancer risk," *Epidemiol. Rev.* **9**, 146–174 (1987).
- ⁶J. Brisson, R. Verreault, A. S. Morrison, D. Tennina, and F. Meyer, "Diet, mammographic features of breast tissue, and breast cancer risk," *Am. J. Epidemiol.* **130**, 14–24 (1989).
- ⁷A. F. Saftlas, R. N. Hoover, L. A. Brinton, M. Szklo, D. R. Olson, M. Salane, and J. N. Wolfe, "Mammographic densities and risk of breast cancer," *Cancer (N.Y.)* **67**, 2833–2838 (1991).
- ⁸A. M. Oza and N. F. Boyd, "Mammographic parenchymal patterns: A marker of breast cancer risk," *Epidemiol. Rev.* **15**, 196–208 (1993).
- ⁹C. Byrne, "Studying mammographic density: Implications for understanding breast cancer," *J. Natl. Cancer Inst.* **89**, 531–532 (1997).
- ¹⁰W. Leung, F. Goldberg, B. Zee, and E. Sterns, "Mammographic density in women on postmenopausal hormone replacement therapy," *Surgery (St. Louis)* **122**, 669–674 (1997).
- ¹¹N. F. Boyd, C. Greenberg, G. Lockwood, L. Little, L. Martin, J. Byng, Y. Martin, and D. Tritchler, "Effects at two years of a low-fat, highcarbohydrate diet on radiologic features of the breast: Results from a randomized trial," *J. Natl. Cancer Inst.* **89**, 488–497 (1997).
- ¹²D. V. Spicer, G. Ursin, Y. R. Parisky, J. G. Pearce, D. Shoupe, A. Pike, and M. C. Pike, "Changes in mammographic densities induced by a hormonal contraceptive designed to reduce breast cancer risk," *J. Natl. Cancer Inst.* **86**, 431–436 (1994).
- ¹³C. Byrne, C. Schairer, J. N. Wolfe, N. Parekh, M. Salane, L. A. Brinton, R. Hoover, and R. Haile, "Mammographic features and breast cancer risk: Effects with time, age, and menopause status," *J. Natl. Cancer Inst.* **87**, 1622–1629 (1995).
- ¹⁴J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe, "Automated analysis of mammographic densities," *Phys. Med. Biol.* **41**, 909–923 (1996).
- ¹⁵S. J. Graham, M. J. Bronskill, J. W. Byng, M. J. Yaffe, and N. F. Boyd, "Quantitative correlation of breast tissue parameters using magnetic resonance and x-ray mammography," *Br. J. Cancer* **73**, 162–168 (1996).
- ¹⁶M. J. Yaffe, N. F. Boyd, J. W. Byng, R. A. Jong, R. Fishell, G. A. Lockwood, L. E. Little, and D. L. Tritchler, "Breast cancer risk and measured mammographic density," *Eur. J. Cancer Prev.* **7**, S47–S55 (1998).
- ¹⁷J. M. Boone, K. K. Lindfors, C. S. Veatty, and J. A. Seibert, "A breast density index for digital mammograms based on radiologists' ranking," *J. Digital Imaging* **11**, 101–115 (1998).
- ¹⁸*American College of Radiology. Breast Imaging—Reporting and Data System (BI-RADS)*, 3rd ed. (American College of Radiology, Reston, VA, 1998).
- ¹⁹E. White, P. Velentgas, M. T. Mandelson, C. D. Lehman, J. G. Elmore, P. Porter, Y. Yasui, and S. H. Taplin, "Variation in mammographic breast density by time in menstrual cycle among women aged 40–49 years," *J. Natl. Cancer Inst.* **90**, 906–910 (1998).
- ²⁰C. Zhou, H. P. Chan, N. Petrick, B. Sahiner, H. M. A. M. A. Roubidoux, L. M. Hadjiiski, and M. M. Goodsitt, "Computerized image analysis: Estimation of breast density on mammograms," *Proc. SPIE* **3979**, 1615–1624 (2000).
- ²¹K. E. Martin, M. A. Helvie, C. Zhou, M. A. Roubidoux, J. E. Bailly, C. Paramagul, and H. P. Chan, "Automatic computer-aided quantitative assessment of mammographic density: A validation study," *100th Annual Meeting of American Roentgen Ray Society*, Washington, DC, 7–12 May 2000.
- ²²A. R. Morton, H. P. Chan, and M. M. Goodsitt, "Automated model-guided breast segmentation algorithm," *Med. Phys.* **23**, 1107–1108 (1996).
- ²³M. M. Goodsitt, H. P. Chan, B. Liu, A. R. Morton, S. V. Guru, S. Keshavmurthy, and N. Petrick, "Classification of compressed breast shape for the design of equalization filters in mammography," *Med. Phys.* **25**, 937–948 (1998).
- ²⁴N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).
- ²⁵A. K. C. Wong, "A gray-level threshold selection method based on

maximum entropy principle," IEEE Trans. Syst. Man Cybern. **19**, 866–871 (1989).

²⁶J. N. Kapur, "Twenty-five years of maximum-entropy principle," J. Math. Phys. **17**, 103–156 (1983).

²⁷K. E. Martin, M. A. Helvie, C. Zhou, M. A. Roubidoux, J. Bailey, P. C., C. Blane, K. Klein, S. Sonnad, and H. P. Chan, "Mammographic density measured by an automatic computer-aided quantitative method: a feasibility study," Am. J. Roentgenol. (submitted).

Correlation between mammographic density and volumetric fibroglandular tissue estimated on breast MR images

Jun Wei,^{a)} Heang-Ping Chan, Mark A. Helvie, Marilyn A. Roubidoux, Berkman Sahiner, Lubomir M. Hadjiiski, Chuan Zhou, Sophie Paquerault, Thomas Chenevert, and Mitchell M. Goodsitt

Department of Radiology, University of Michigan, Ann Arbor, Ann Arbor, Michigan 49109

(Received 18 June 2003; revised 26 November 2003; accepted for publication 21 January 2004; published 26 March 2004)

Previous studies have found that mammographic breast density is highly correlated with breast cancer risk. Therefore, mammographic breast density may be considered as an important risk factor in studies of breast cancer treatments. In this paper, we evaluated the accuracy of using mammograms for estimating breast density by analyzing the correlation between the percent mammographic dense area and the percent glandular tissue volume as estimated from MR images. A dataset of 67 cases having MR images (coronal 3-D SPGR T1-weighted pre-contrast) and corresponding 4-view mammograms was used in this study. Mammographic breast density was estimated by an experienced radiologist and an automated image analysis tool, Mammography Density ESTimator (MDEST) developed previously in our laboratory. For the estimation of the percent volume of fibroglandular tissue in breast MR images, a semiautomatic method was developed to segment the fibroglandular tissue from each slice. The tissue volume was calculated by integration over all slices containing the breast. Interobserver variation was measured for 3 different readers. It was found that the correlation between every two of the three readers for segmentation of MR volumetric fibroglandular tissue was 0.99. The correlations between the percent volumetric fibroglandular tissue on MR images and the percent dense area of the CC and MLO views segmented by an experienced radiologist were both 0.91. The correlation between the percent volumetric fibroglandular tissue on MR images and the percent dense area of the CC and MLO views segmented by MDEST was 0.91 and 0.89, respectively. The root-mean-square (rms) residual ranged from 5.4% to 6.3%. The mean bias ranged from 3% to 6%. The high correlation indicates that changes in mammographic density may be a useful indicator of changes in fibroglandular tissue volume in the breast. © 2004 American Association of Physicists in Medicine. [DOI: 10.1118/1.1668512]

Key words: mammography, breast density, MR images, correlation

I. INTRODUCTION

Studies have shown that there is a strong positive correlation between breast parenchymal density imaged on mammograms and breast cancer risk.^{1–3} The relative risk is estimated to be about 4 to 6 for women whose mammograms have parenchymal densities over 60% of the breast area, as compared to women with less than 5% densities. Other cohort studies^{4–13} also found that breast cancer risk in the category with the most extensive dense tissue was 1.8 to 6 times as high as that in the category with the least extensive dense tissue. Mammographic density as the risk indicator is greater than almost all other risk factors of breast cancer.^{2,14} Although there is no direct evidence that changes in mammographic breast densities will result in changes in breast cancer risk, the strong correlation between breast density and breast cancer risk has prompted researchers to use mammographic density for monitoring the effects of intervention as well as for studying breast cancer etiology.^{14–17}

A number of researchers have investigated image analysis techniques to estimate breast density.^{15,18–28} The common approaches are to analyze the textural pattern or the percentage of mammographic densities relative to the breast area. It has been found that the texture measures were corre-

lated with parenchymal density patterns but they appeared to be less sensitive measures of relative risk than the percent dense area.^{1,25,29} In current practice, breast density is estimated mainly by radiologists' visual judgment of the fibroglandular tissue imaged on mammograms following the Breast Imaging—Reporting and Data System (BI-RADS) lexicon.^{30,31} Because of the qualitative and subjective nature of visual judgment, there are large intraobserver and interobserver variations in the estimated breast density. The large variability may reduce the observed correlation between breast cancer risk and breast density. It may also reduce the sensitivity of studies using mammographic density for monitoring the effect of risk modifying treatments. We have developed an automated image analysis system, Mammographic Density ESTimator (MDEST), to assist radiologists in estimating breast density on mammograms. A computerized analysis is expected to increase the reproducibility and consistency in the estimation of mammographic density, thereby improving the accuracy of the related studies. In our previous study, we have found that the percent mammographic density segmented by MDEST agreed closely with that estimated by radiologists' interactive thresholding.³²

The high correlation between breast cancer risk and breast

density indicates that breast cancer risk may be closely related to the volume of glandular tissue in the breast. Among the modalities available for breast imaging at present, magnetic resonance (MR) imaging is likely to be the most accurate method for volumetric dense tissue estimation because fibroglandular tissue and adipose tissue can be well distinguished in MR images when a proper image acquisition technique is used.³³ However, MR imaging is expensive, making it difficult to use MR imaging as a routine monitoring tool.^{33,34} On the other hand, a mammogram is a two-dimensional (2-D) projection image of a three-dimensional (3-D) object. The area of dense tissue measured on a mammogram is not an accurate measure of the volume of fibroglandular tissue in the breast because no thickness information is used. However, mammography is a widely available low cost procedure that may be used for monitoring breast density change during preventive and interventional treatment or other studies. Women who participate in screening will also have mammograms readily available for retrospective review. Therefore, mammography will most likely be the method of choice for breast density estimation.

In this study, we investigated the correlation between the volumetric fibroglandular tissue in the breast and the projected breast dense area on mammograms by analyzing the percent volumetric fibroglandular tissue in MR breast images and the percent dense area in corresponding mammograms. Our purpose in this study is not to evaluate the usefulness of either MR fibroglandular tissue volume or mammographic density as an indicator for breast cancer risk, which have been studied by other investigators. Rather, we used the MR breast images to estimate the volumetric fibroglandular tissue in the breast and explored the reason that a change in mammographic density (2-D) can be used as an indicator of breast density change (3-D). These comparisons will provide a better understanding of their relationship, and may lead to improved methods for utilizing mammographic density as a surrogate marker for breast cancer risk.

II. MATERIALS AND METHOD

A. Dataset

In a previous study, gadolinium contrast enhanced MR dynamic imaging was employed to characterize malignant and benign breast lesions. A dataset was collected with IRB approval which included MR images and corresponding mammograms acquired between detection and before biopsy for a given patient. In the MR study, several series of images were acquired for each patient. Patients were scanned prone using a commercial dual phased-array breast coil. The imaging protocol included a series was the coronal 3-D T1-weighted pre-contrast series (coronal sections 2–5 mm thick, 32 slices; 3-D Spoiled Gradient-Recalled Echo (SPGR); TE = 3.3 ms; TR = 10 ms, Flip = 40°, matrix = 256 × 128, FOV = 28–32 cm right/left, 14–16 cm superior/inferior, scan time = 2 min 38 sec). This 3-D SPGR sequence produces full volume coverage of both breasts with contiguous image sections. The dense parenchyma and fat tissue are well separated with this heavily T1-weighted acquisition. We used a

set of 67 patients to study the correlation between the 2-D projected percentage of dense area on a mammogram and the percentage of dense tissue volume estimated from the 3-D MR images.

The mammograms consisting of the craniocaudal (CC) view and the mediolateral oblique (MLO) view of both breasts of the patient were digitized with a LUMISYS 85 laser film scanner at a pixel size of 50 μm × 50 μm . The digitizer has a gray level resolution of 12 bits and a nominal optical density (O.D.) range of 0 to 4. For density segmentation, it is not necessary to use very high-resolution images. To reduce processing time, the full resolution mammograms were first smoothed with a 16 × 16 box filter and subsampled by a factor of 16, resulting in 800 μm × 800 μm images for this study.

B. Estimation of fibroglandular tissue volume on MR images

Since it is not our intention to routinely segment MR images for breast density estimation, we did not attempt to develop an automated method for this application. Our algorithm for segmentation of volumetric fibroglandular tissue on MR images used a semi-automatic method. The computer performed an initial segmentation. A graphical user interface (GUI) was developed to allow a user to review the segmentation of every slice and make modifications if necessary. The method consists of four steps. First, the breast boundary was detected automatically on each slice. A deformable model and manual modification were used to correct for incorrectly detected boundaries that usually occurred in slices near the chest wall where there were no well-defined breast boundaries. Because of inhomogeneity of the breast coil sensitivity, the signal intensity in the breast region was not uniform across the field of view. A background correction technique that estimated the low frequency background from the gray levels along the breast boundary was developed to reduce this systematic nonuniformity. Manual interactive thresholding of the gray level histogram in the breast region was then used to separate the fibroglandular from the fatty region. Morphological erosion was used to exclude the skin voxels along the breast boundary. Finally, the volume of fibroglandular tissue was calculated by integration over all slices containing the breast. A flow chart of our algorithm is shown in Fig. 1.

C. Breast boundary detection

A two-step algorithm was developed for the detection of breast boundary on each slice. First, we used a seeded pixel thresholding algorithm (SPTA) for the initial assessment of a breast boundary. Second, a 2-D active contour algorithm further refined the boundary. For slices close to the chest wall where no clear boundary can be seen, manual modification was used to outline an estimated boundary.

The SPTA determined the optimal threshold by iteratively partitioning the MR image into two parts and using the gradient value along the boundary of the partition as a guide in optimizing the threshold. First, the center of gravity was se-

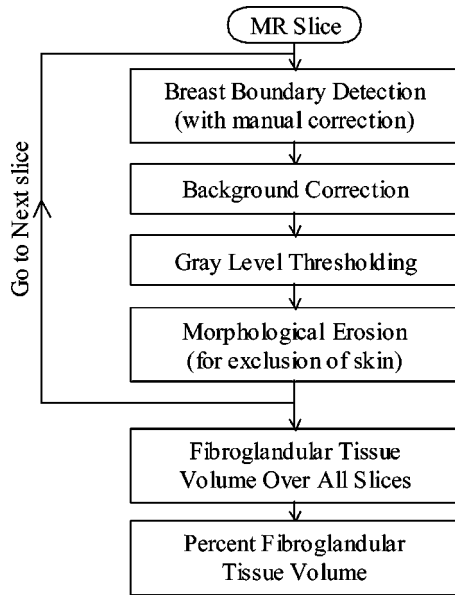


FIG. 1. The flow-chart for the segmentation of the fibroglandular tissue on MR images.

lected as the starting pixel on each slice. The gray level of the starting pixel was used as a threshold to create a binary partition of the image in which all pixels greater than the threshold were set to one and all other pixels were set to zero. Second, the gradient value of each pixel on the boundary of the binary partition was calculated by applying the Sobel filter to the original image. The gradient assessment for this particular binary partition was defined as the average gradient magnitude of these boundary pixels. The threshold value was reduced to zero in a stepwise manner. The partition for each threshold value was created and the gradient assessment for each partition was calculated as described above. The partition with the maximum gradient assessment was considered to be the initial segmentation result for the breast, and the boundary of this partition was considered to be the initial breast boundary.

After the initial segmentation, a deformable contour method was used to further refine the boundary. The movement of the boundary pixel was controlled by an energy function which consisted of internal energy and external energy. The internal energy components used in this study were the continuity and curvature of the contour, as well as the homogeneity of the segmented partition. The external energy components were the negative of the smoothed image gradient magnitude, and a balloon force that exerted pressure at a normal direction to the contour. The energy function was defined as the following:

$$E = \sum_{c=1}^N [E_{\text{inter}}(c) + E_{\text{extert}}(c)], \quad (1)$$

where E_{inter} and E_{extert} are the internal energy and the external energy, respectively, as defined in Eq. (2) and Eq. (3):

$$E_{\text{inter}} = w_{\text{curv}}E_{\text{curv}}(c) + w_{\text{cont}}E_{\text{cont}}(c) + w_{\text{hom}}E_{\text{hom}}, \quad (2)$$

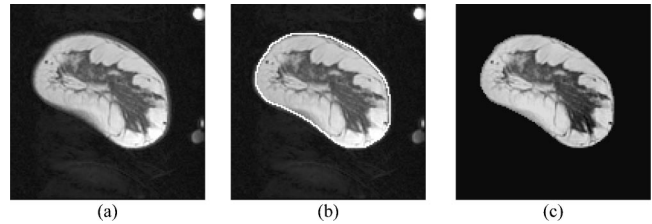


FIG. 2. An example of the first three processing blocks in Fig. 1. (a) Original MR slice; (b) automatically-detected breast boundary superimposed on the image; and (c) the background-corrected image.

$$E_{\text{extert}} = w_{\text{grad}}E_{\text{grad}}(c) + w_{\text{bal}}E_{\text{bal}}(c), \quad (3)$$

where curv , cont , grad , bal , hom denoted curvature, continuity, gradient, balloon force and homogeneity, respectively, and each energy term was associated with a weight, w . The detailed definition for each term can be found in the literature.³⁵ An example of a MR slice of a breast is shown in Fig. 2(a), and the segmented boundary is shown in Fig. 2(b). Note that the two breasts of a patient were scanned together but each breast was analyzed separately.

D. Background correction

To reduce the nonuniformity of the MR signal intensity in the breast region, a background correction technique³⁶ using the pixel values around the segmented breast region was employed. For a given pixel (i, j) inside the breast region, the gray value of the background image was estimated as shown in Eq. (4):

$$B(i, j) = \left[\frac{L}{d_l} + \frac{R}{d_r} + \frac{U}{d_u} + \frac{D}{d_d} \right] / \left[\frac{1}{d_l} + \frac{1}{d_r} + \frac{1}{d_u} + \frac{1}{d_d} \right], \quad (4)$$

where L , R , U and D are the average gray values inside a breast background estimation region (BBER) centered at the left, right, upper and lower pixels on the breast boundary, respectively. A BBER was defined as the intersection of a 21×21 -pixel box and the breast region. The center pixels for the left and right boxes were the intersection points between the breast boundary and a horizontal line passing through the given pixel (i, j) . Similarly, the upper and lower center pixels for the upper and lower boxes were the intersection points between the breast boundary and a vertical line passing through the given pixel (i, j) . Only the pixels that were within the intersected area between the 21×21 -pixel box and the breast region were included in the definition of the BBER and the calculation of the average gray value. The contributions of the average gray levels to the background pixel (i, j) were inversely weighted by their distances d_l, d_r, d_u, d_d from the given pixel (i, j) . An example of the background corrected image is shown in Fig. 2(c).

E. Segmentation of fibroglandular tissue

We developed a GUI that allowed the user to perform a combination of manual and automatic operations to segment the breast boundary and the fibroglandular tissue on the MR

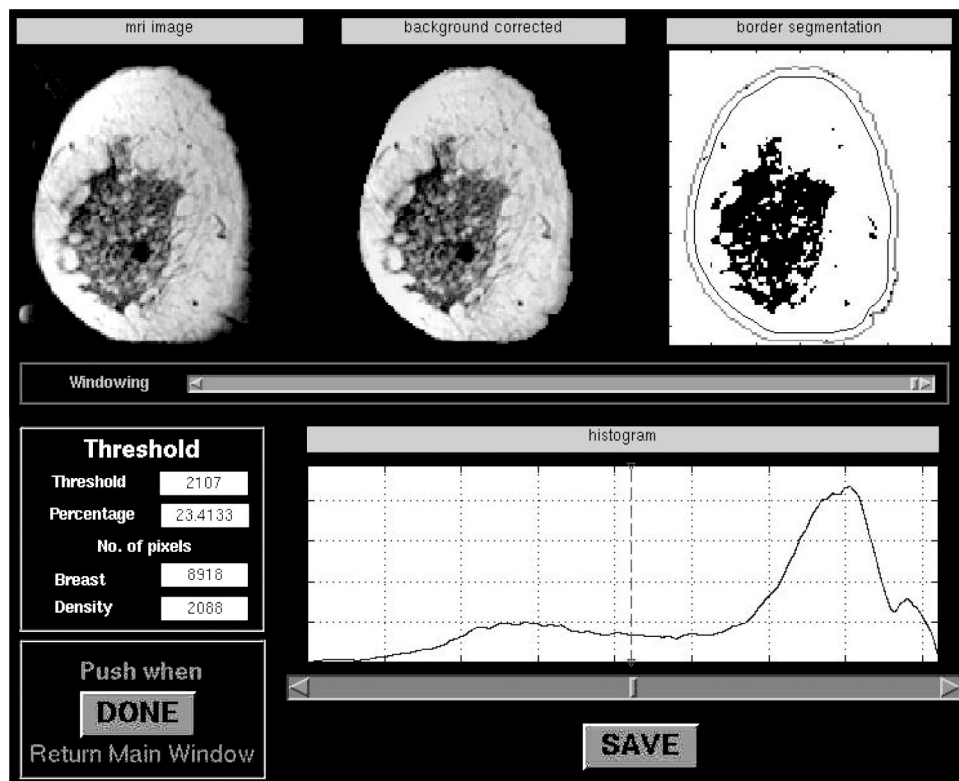


FIG. 3. The graphic user interface for the segmentation of the fibroglandular tissues on the MR slice. The upper row shows the original MR slice (left), the background-corrected image (middle) and the segmented binary image (right). The segmented image responds to the reader's adjustment of the gray level threshold (lower row) in real time so that the reader can choose the appropriate threshold by inspecting the segmented image visually. The dark area in the segmented image indicates the fibroglandular tissue and the white area indicates the adipose tissue. The inner line along the breast boundary is the boundary obtained by morphological erosion to exclude the skin voxels for calculating the fibroglandular tissue volume.

images. The first window (not shown) displayed the MR series and the corresponding mammogram of each breast to give the user an overview of the breast. The segmentation of the fibroglandular tissue on each MR slice was processed in the second window, shown in Fig. 3. The original MR slice, the corresponding background corrected image and the segmented binary image were shown in the upper part of the window. At the lower part of the window, the histogram of the voxel values in the breast region was shown. The user performed interactive thresholding on the histogram and the segmented binary image corresponding to the chosen threshold was displayed in real time in the upper part. If the breast boundary, which was automatically segmented by the computer initially, had to be corrected, the user could go to the third window and manually move the apices of the polygon outlining the boundary. The voxels contributed by the nipple were excluded. On the slices containing breast skin that had voxel values similar to those of fibroglandular tissue, a morphological erosion operation was applied to the breast boundary to exclude the skin voxels from the calculation of the fibroglandular tissue volume in the slice. The size of the structuring element could be selected interactively on the fourth window and the eroded boundary was displayed instantly for a chosen erosion operation. The user might again change the structuring element if the erosion result of the previous choice was deemed unsatisfactory. Since the eroded boundary only marked the region within which the fibroglandular voxels would be summed and would not be used for the calculation of the breast volume, as described below, it did not need to be precise as long as it excluded the skin voxels while not excluding the fibroglandular voxels.

F. MR fibroglandular tissue volume

After the fibroglandular tissue was segmented for each slice, the total number of voxels containing the fibroglandular tissue was obtained as a summation of these voxels over all slices of the breast. The total volume of the breast was obtained as the summation of the voxels enclosed by the breast boundary before morphological erosion. The ratio of these two volumes provided the percent volumetric fibroglandular tissue in the breast.

G. Mammographic density segmentation

We have previously developed an automated method for segmentation of the dense fibroglandular area on mammograms. The method, referred to as the Mammographic Density ESTimator (MDEST) was described in detail elsewhere.³² In brief, the breast boundary on the digitized mammogram is tracked. A dynamic-range compression technique reduces the gray level range of the breast area. By analyzing the shape of the gray level histogram, a rule-based classifier classifies the breast density into one of four classes. Typically, a Class I breast is almost entirely fat; it has a single narrow peak on the histogram. A Class II breast contains scattered fibroglandular densities. Its histogram has two main peaks, with the smaller peak on the right of the bigger one. A Class III breast is heterogeneously dense. Its histogram also has two peaks, but the smaller peak is on the left of the bigger one. A Class IV breast is extremely dense. Its histogram has mainly a single dominant peak, but the peak is wider compared with the peak in the Class I histogram. A second smaller peak sometimes occurs on the left of the

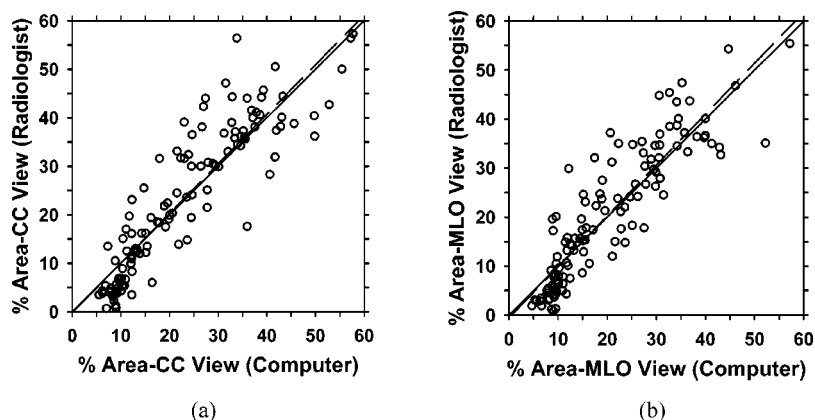


FIG. 4. A comparison of the percent mammographic density obtained from interactive thresholding by an MQSA-qualified radiologist and that estimated by our automated MDEST computer program. (a) CC view, correlation coefficient=0.90, rms residual=6.7, mean difference=0.3; (b) MLO view, correlation coefficient=0.89, rms residual=6.1, mean difference=0.4. Dashed line: linear regression of the data; solid line: diagonal.

main peak. Based on the histogram shape, a threshold is automatically calculated to separate the dense and fatty pixels. The mammographic density was estimated as the percentage of fibroglandular tissue area relative to the total breast area. For MLO view mammograms, the pectoral muscle is detected and excluded from the density area or breast area calculations. In our previous work, the performance of MDEST was verified by comparison with manual segmentation by 5 breast imaging radiologists using a dataset of 260 mammograms from 65 patients that were different from the cases used in the current study. We found that the correlation between the computer-estimated percent dense area and the average segmentation by the 5 radiologists was 0.94 and 0.91, respectively, for CC and MLO views, with a mean bias of less than 2%.

MDEST was applied to the mammograms of the 67 patients used in this study. The percent dense area on mammograms was estimated for the CC-view and the MLO-view mammogram of each breast separately. In addition, an MQSA-qualified radiologist also segmented the dense area by interactive thresholding for each mammogram. The correlation between the mammographic density obtained by manual and automatic segmentation is shown in Figs. 4(a) and 4(b) for the CC view and MLO view, respectively. The correlation coefficients for the CC view and MLO view were 0.90 and 0.89, respectively. The mammographic densities estimated by automatic and manual segmentation were compared with the percent volumetric fibroglandular tissue on MR images as described below.

H. Observer experiments

We performed an experiment to evaluate the variability of the estimated % volumetric fibroglandular tissue due to the uncertainty in the determination of the starting slice of the breast at the chest wall. The starting slice affected the estimation of the breast volume that was calculated by integrating from the starting slice to the anterior of the breast. Twenty-three MR cases from the dataset were randomly selected for this observer experiment. There were a total of 41 breasts because some cases had only one breast. For this subset of cases, each radiologist was asked to select the starting slice from the MR images for each breast. The estimated

% volumetric fibroglandular tissue calculated with all available slices was then compared to that calculated with the selected starting slice.

We also performed observer experiments to evaluate the inter-observer variations in the segmentation of fibroglandular tissue using the semi-automatic method. Two MQSA-qualified radiologists performed the segmentation of the fibroglandular tissue on the MR images of the 41 breasts using the semi-automatic method implemented with the GUI. A Ph.D. researcher who was trained by these radiologists also performed the segmentation independently with the GUI.

After verifying the consistency of segmentation by these observers, the trained Ph.D. completed the segmentation of all MR cases. The correlation between percent volumetric fibroglandular tissue on MR images and percent dense area on mammograms was then examined for the entire dataset.

III. RESULTS

A. Effect of selection of the starting slice

Figure 5(a) shows the correlation of the % volumetric fibroglandular tissue calculated using all available slices for the breast with that calculated using the selected starting slice by radiologist A for the 41 breasts. The correlation coefficient was 0.999. To compare the difference between their results, the mean difference and the root-mean-square (rms) residual, which is the residual from the linear least-squares-fitted line, were also calculated. The mean difference was 0.7 and the rms residual was 0.6. The result is similar for radiologist B (not shown), with a correlation coefficient of 0.999, a mean difference of 0.4 and a rms residual of 0.4. The correlation between the % volumetric fibroglandular tissue calculated using the selected starting slice by radiologist A with that calculated using the selected starting slice by radiologist B was also very high with a correlation coefficient of 0.988, a mean difference of 0.7 and a rms residual of 1.8, as shown in Fig. 5(b). These comparisons indicated that the variability in the selection of the starting slice of the breasts did not have a strong influence on the % volumetric fibroglandular tissue. We therefore used all available slices in the MR dataset for each breast in the following analyses.

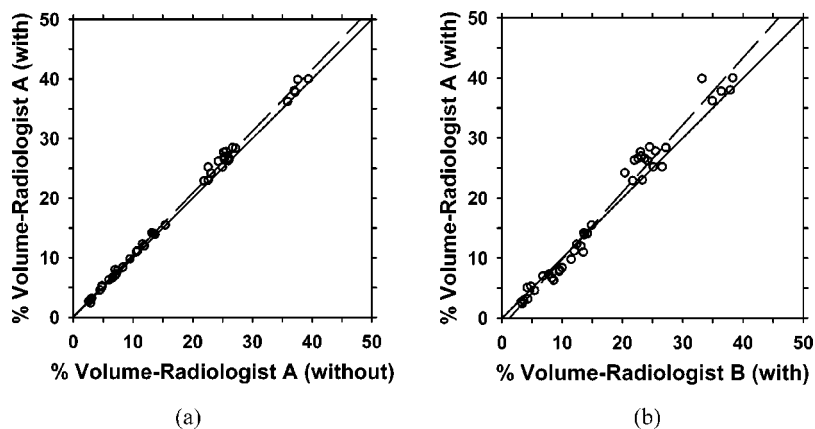


FIG. 5. (a) A comparison of the percent fibroglandular tissue volume calculated using the selected starting slice with that calculated using all available slices for radiologist A, correlation coefficient=0.999. (b) A comparison of the percent fibroglandular tissue volume calculated using the selected starting slice by radiologist B with that by radiologist A, correlation coefficient =0.988. Dashed line: linear regression of the data; solid line: diagonal.

B. Inter-observer variation between radiologists

Figure 6(a) shows the comparison of the percent volumetric fibroglandular tissues on MR images segmented by two radiologists for the 41 breasts. The correlation between the segmentation results of the two radiologists is 0.99. The mean difference was found to be 0.3 and the rms residual was 1.6.

C. Inter-observer variation between radiologists and trained Ph.D.

Figure 6(b) shows the comparison of the percent volumetric fibroglandular tissues segmented by the trained Ph.D. against that segmented by radiologist A. A similar result was obtained by comparing the percent volumetric tissue segmented by the trained Ph.D. and that segmented by radiologist A except that the data points were even closer to the diagonal (not shown). The correlation between the result of the trained Ph.D. and the results of both radiologists was 0.99. The corresponding mean differences were -0.8 and -0.4 , respectively, and the rms residuals were 1.4 and 1.5, respectively.

D. Correlation between percent volumetric fibroglandular tissue on MR images and percent mammographic density

The percent volumetric fibroglandular tissue on MR images was compared with the percent dense area on CC- and

MLO-view mammograms. After verifying that the difference in segmentation between the trained Ph.D. and the radiologists was similar to the interobserver variations between the two experienced radiologists, the trained Ph.D. completed the segmentation of the entire dataset.

Figure 7 shows the comparison of the percent volumetric fibroglandular tissue on MRI and the percent mammographic density segmented by a radiologist. The percent areas on CC- and MLO-view mammograms are higher than the percent volume on MR images with a mean difference of 5.7% and 3.0%, respectively.

Figure 8 shows the comparison of the percent volumetric fibroglandular tissue on MRI and the percent mammographic density segmented by MDEST. The percent areas on CC- and MLO-view mammograms segmented by the computer are higher than the percent volume on MR images with a mean difference of 5.3% and 2.6%, respectively.

The correlation coefficients, the mean differences and the rms residuals between the percent volumetric fibroglandular tissue on MR images and percent dense area on mammograms are compared in Table I. The correlation between the percent volume on MR images and percent area on mammograms of the fibroglandular breast tissue is high, ranging from 0.89 to 0.91. Although it is not expected that the values of percent volume agree with the values of percent area, their mean differences range only from 3% to 6% and the rms residual range from 5.4 to 6.3.

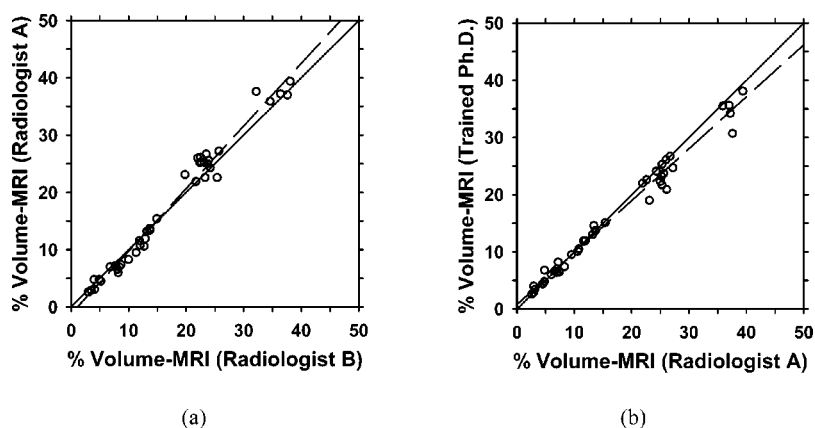


FIG. 6. A comparison of the segmentation of fibroglandular tissue from MR images between two observers: (a) two experienced MQSA-qualified radiologists, correlation coefficient=0.99. (b) The trained Ph.D. and Radiologist A, correlation coefficient=0.99. The correlation between the trained Ph.D. and Radiologist B is also 0.99 but the data points were very close to the diagonal and is not shown. The % volumetric fibroglandular tissue was calculated using all available slices. Dashed line: linear regression of the data; solid line: diagonal.

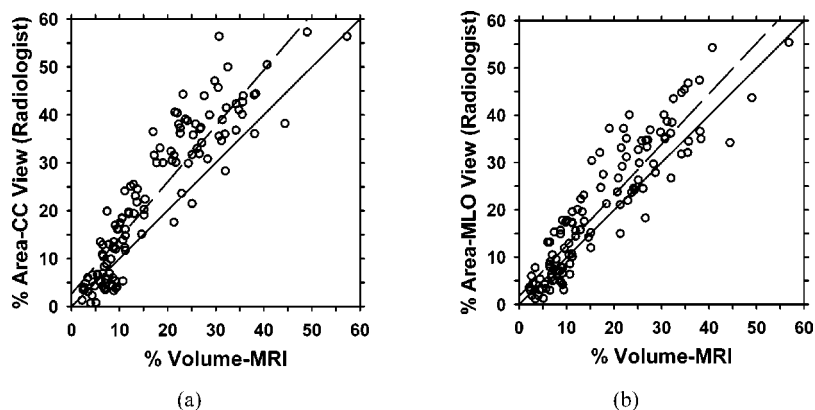


FIG. 7. A comparison of the percent fibroglandular tissue volume on MR images and the percent dense area on mammograms segmented by an experienced radiologist. (a) CC view, correlation coefficient=0.91; (b) MLO view, correlation coefficient=0.91. Dashed line: linear regression of the data; solid line: diagonal.

IV. DISCUSSION

Our purpose in this paper was to investigate the relationship between the percent dense area on mammogram and the percent fibroglandular tissue volume on MR image. We found a direct correlation between mammographic density and MR volumetric density (Fig. 7 and Fig. 8). The correlation coefficients between the percent area on a mammogram and the percent volume on MR images are high at 0.89 and 0.91. These results are more promising than those found in previous studies that attempted to correlate percent dense area on mammograms with MR information. Graham *et al.*³³ investigated the relationship between percent density (projected dense area) on mammogram and two objective MR parameters of breast tissue, relative water content and mean T2 relaxation. Their results with 45 cases showed a positive correlation between percent density and relative water content (Pearson correlation coefficient=0.79) and a negative correlation between percent density and mean T2 value (Pearson correlation coefficient=-0.61). Another study by Lee *et al.*³⁴ analyzed fatty and fibroglandular tissue in different age groups to compare x-ray mammography with T1-weighted MR images. Their study with 40 cases indicated that the correlation between the two techniques is 0.63 when the fat content was more than 45%. However, the correlation coefficient decreased to 0.34 when their analysis included only dense breasts.

It may be noted that although MR imaging is currently the most accurate method for estimating the volumetric fibro-

glandular tissue in the breast, it is still not the ideal tool. Fibrous tissue and glandular tissue are not well separated with current MR imaging techniques. Since the amount of glandular tissue in the breast is the important factor relating to breast cancer risk, further studies are warranted for differentiating the glandular and the fibrous components of the imaged volume. The correlation between the percent glandular tissue volume and percent projected dense area on a mammogram will be a more reliable indicator of the usefulness of mammographic density analysis.

The density on mammograms is a 2-D projected area of the fibroglandular tissues. The percent dense area is not expected to be equal in value to the percent volume. The mean differences between the percent volume and the percent area on CC- and MLO-views, as determined by the radiologist's interactive segmentation, are 5.7 and 3.0, respectively (Table I), with the percent dense area values being higher. We also investigated the rms residual between the percent volume and the percent area when the relationship between them was assumed to be linear. The rms residual between the percent volume and the percent area on CC- and MLO-views are 6.3 and 5.6, respectively (Table I), relative to the straight line obtained from linear least squares fits to the data. One possible factor that may contribute to a higher value of percent dense area on mammograms than the percent volume value on MR images is that the tissue volume imaged by the two modalities is somewhat different. The MR images include more tissue near the chest wall, which is mainly retroglan-

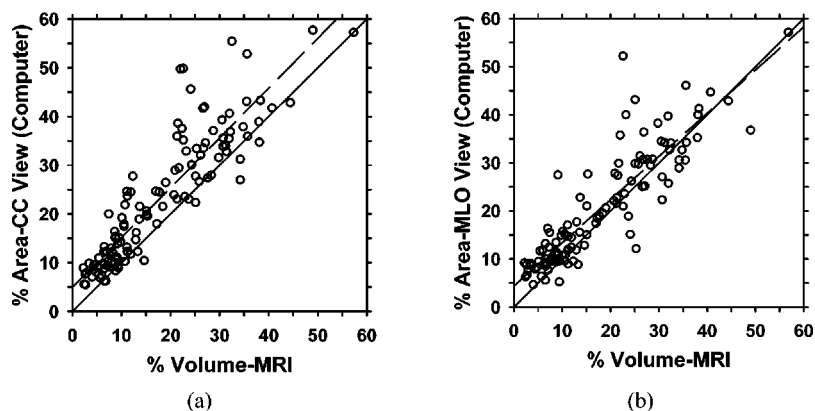


FIG. 8. A comparison of the percent volume on MR images and the percent area on mammogram segmented by our automated MDEST computer program. (a) CC view, correlation coefficient=0.91; (b) MLO view, correlation coefficient=0.89. Dashed line: linear regression of the data; solid line: diagonal.

TABLE I. Statistic analysis of the relationship between percent fibroglandular tissue volume on breast MR images and percent dense area on mammograms segmented by radiologist and MDEST.

	Radiologist		Computer (MDEST)	
	CC vs MRI	MLO vs MRI	CC vs MRI	MLO vs MRI
Correl. coeff.	0.91	0.91	0.91	0.89
rms residual	6.3	5.6	5.8	5.4
Mean diff.	5.7	3.0	5.3	2.6

dular adipose tissue, than a mammogram does, thus reducing the percentage of fibroglandular tissue volume. The reduction in the percent volume values, however, is relatively small, as found in our study evaluating the effects of selecting starting slices for volume calculation (Fig. 5). The main difference may therefore be attributed to the geometric relationship between the volume and the projected 2-D area, explained later.

Geometrically, we do not expect the relationship between volume and its projected 2-D area to be linear. In a hypothetical situation such that the dense tissue volume is a sphere (volume = $\frac{4}{3}\pi r^3$) enclosed inside a concentric spherical shell of fatty tissue volume, the percent projected 2-D area (area = πr^2) of the inner sphere relative to the outer sphere is equal to the percent volume to the power of $2/3$. The relationship between the percent area and the percent volume is therefore not linear, and the percent area is larger in value than the percent volume for any ratio of radii between the two spheres. In general, the compressed breast and the dense tissue are not spherical. To investigate the empirical relationship between the percent area and the percent volume in the nonlinear situation, we applied least squares fits in several polynomial models to the data points in Fig. 7. The results are shown in Table II and Fig. 9. A comparison of Table I and Table II indicates that the $Y = kx^{2/3}$ model (x = percent fibroglandular tissue volume, Y = percent mammographic dense area) resulted in slightly larger rms residuals than the linear model. The model $Y = kx^m$ with m equal to 0.83 and 0.86, respectively, for CC- and MLO-views slightly reduced the rms residuals. The best fit was obtained from the model $Y = k_1x^m + k_2$. However, the

TABLE II. An analysis of the relationship between percent fibroglandular tissue volume (x) on breast MR images and percent dense area (Y) on mammograms segmented by radiologist using three mathematical models. m , k , k_1 and k_2 are constants determined by least squares curve fitting.

Mathematical model		$Y = kx^{2/3}$	$Y = kx^m$	$Y = k_1x^m + k_2$
CC vs MRI	Least squares Fit	$Y = 0.82x^{2/3}$	$Y = 1.03x^{0.83}$	$Y = 1.02x^{0.48} - 0.19$
	rms residual	6.5	6.0	5.6
	Coefficient of determination	0.82	0.85	0.87
MLO vs MRI	Least squares Fit	$Y = 0.73x^{2/3}$	$Y = 0.96x^{0.86}$	$Y = 0.90x^{0.60} - 0.09$
	rms residual	6.0	5.5	5.3
	Coefficient of determination	0.80	0.84	0.85

situation that the percent projected area was negative when the percent volume was zero would not occur physically. Note that if the model was fitted to the percent area data segmented by MDEST (Fig. 8), the k_2 values would become positive, indicating that the nonzero k_2 values are likely caused by segmentation biases.

Overall, these models demonstrate that there is no simple mathematical relationship between the percent volume and the percent projected area but the values for the exponents appeared to be in a reasonable range. The relationship between the percent volumes of two 3-D objects, one within another, and their percent projected 2-D area depends on their shapes. For example, the closer the two volumes are to concentric cylinders of the same height, the closer the exponent is to unity. The spread of the data points can therefore be attributed to the various irregular shapes of the fibroglandular tissue in the breasts, the changes in the shapes of the fatty and fibroglandular tissue due to compression, as well as the uncertainties in the segmentation of both the mammograms and the MR images. Although the spread of the data points in the correlation plots is large, one can expect that when the mammographic density of a given patient is monitored over time, the variations in the projected dense area due to the geometric factors, described above, will actually be much less than that observed from the scatter plots among a large number of patients. In other words, the uncertainty in the estimated percent density from the serial mammograms of a given patient should be much less than those shown in

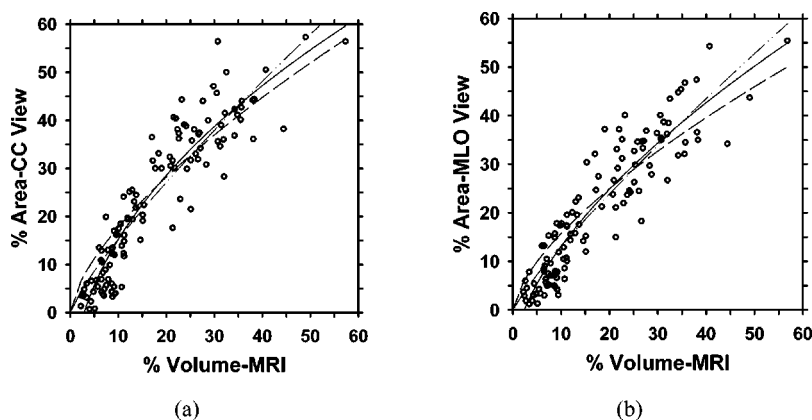


FIG. 9. Nonlinear fitting of the relationship between the percent volume and the percent area segmented by a radiologist with the least squares method. (a) CC view, (b) MLO view. Dashed line: $y = kx^{2/3}$; dashed-dotted line: $y = kx^m$; solid line: $y = k_1x^m + k_2$. The fitted parameters of the models, m , k , k_1 and k_2 , are shown in Table II.

Fig. 7. The strong correlation observed between the percent dense area on mammograms and the percent volumetric fibroglandular tissue on MR images therefore indicates that a change in mammographic density can be a useful indicator of a change in percent fibroglandular tissue volume in the breast.

Recently, some researchers attempted to estimate the thickness of the fibroglandular tissue in local regions of the mammograms from the projected density.³⁷ This approach is expected to provide a more accurate estimation of the fibroglandular tissue volume if the true thicknesses of the fibroglandular tissue and fatty tissue can be determined at various locations of the projected breast region. The volume of the fibroglandular tissue can then be summed over the pixels in the breast region and the percent volume calculated. However, to obtain accurate measurements, this approach requires the knowledge of the sensitometric curve for the screen-film mammogram at the imaging facility (or use of a digital detector with linear response) and other physical parameters such as the scatter fraction, the beam quality and beam hardening, in addition to the compressed breast thickness and the breast shape profile at the periphery. Some of the requirements may be circumvented by using a look-up table predetermined with a phantom calibration. Other factors may have to be approximated or ignored, or require further corrections by imaging each mammogram with a calibration phantom placed adjacent to the breast. This method is still being developed and the accuracy of estimating the thickness of the local fibroglandular tissue from a mammogram is yet to be determined. To our knowledge, no study to date has demonstrated that fibroglandular tissue volume estimated from mammograms has a higher correlation with the percent volumetric fibroglandular tissue volume estimated from MR images or other volumetric methods than we found in our current study. Furthermore, even if the local fibroglandular tissue thickness on mammograms can be measured in a laboratory or in an academic center using elaborate calibration schemes, it is doubtful that these methods can be translated into routine clinical measurement in mammography clinics. Its use may then be limited to controlled clinical trials. An estimation of the percent dense area projected on mammograms is likely a more practical approach for breast density assessment. The high correlation between the percent dense area and the percent fibroglandular tissue volume on MR images as demonstrated in the current study further supports the validity of this approach.

V. CONCLUSION

In this study, we investigated the correlation between the percent mammographic dense area and the percent volumetric fibroglandular tissue as measured on MR images. A semi-automatic method was developed for segmentation of the MR images and a fully automated computerized method, MDEST, was used to segment the mammograms. The performance of MDEST on the set of mammograms used in this study was verified with an experienced radiologist's manual segmentation. The inter-observer variability in segmentation

of MR images was found to be small with correlation coefficients of 0.99. The correlation between the percent volume on MR images and percent area segmented by a radiologist for either CC- view or MLO-view is 0.91. The correlation between percent volume and percent area estimated by MDEST is 0.91 and 0.89, respectively, for CC and MLO views. Mammographic density is thus highly correlated with the percent volumetric fibroglandular tissue in the breast. The high correlation indicates that changes in mammographic density may be a useful indicator of changes in fibroglandular tissue volume in the breast. Our computerized image analysis tool, MDEST, can provide a consistent and reproducible estimation of percent dense area on routine clinical mammograms. The automated image analysis tool may improve the sensitivity of quantifying mammographic density changes, thereby contributing to the understanding of the relationship of mammographic density to breast cancer risk, detection, and prognosis, and the prevention and treatment of breast cancer.

ACKNOWLEDGMENTS

This work is supported by U.S. Army Medical Research and Materiel Command Grants No. DAMD 17-01-1-0326, No. DAMD 17-02-1-0214, and No. DAMD 17-99-1-9294. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred.

^{a)}Author to whom correspondence should be addressed. Jun Wei, Ph.D., Department of Radiology, University of Michigan, CGC B2103, 1500 E. Medical Center Drive, Ann Arbor, Michigan 48109. Phone: 734-647-8553; fax: 734-615-5513; electronic mail: jvwei@umich.edu

¹A. F. Saftlas, R. N. Hoover, L. A. Brinton, M. Szklo, D. R. Olson, M. Salane, and J. N. Wolfe, "Mammographic densities and risk of breast cancer," *Cancer (N.Y.)* **67**, 2833–2838 (1991).

²N. F. Boyd, G. A. Lockwood, J. W. Byng, D. L. Trichler, and M. J. Yaffe, "Mammographic densities and breast cancer risk," *Cancer Epidemiology Biomarkers & Prevention* **7**, 1133–1144 (1998).

³C. M. Vachon, C. C. Kuni, K. Anderson, V. E. Anderson, and T. A. Sellers, "Association of mammographically defined percent breast density with epidemiologic risk factors for breast cancer (United States)," *Cancer Causes & Control* **11**, 653–662 (2000).

⁴P. M. Krook, "Mammographic parenchymal patterns as risk indicators for incident cancer in a screening program: an extended analysis," *Am. J. Roentgenol.* **131**, 1031–1035 (1978).

⁵R. L. Egan and R. C. Mosteller, "Breast cancer mammography patterns," *Cancer (N.Y.)* **40**, 2087–2090 (1977).

⁶B. Threatt, J. M. Norbeck, N. S. Ullman, R. Kummer, and P. Roselle, "Association between mammographic parenchymal pattern classification and incidence of breast cancer," *Cancer (N.Y.)* **45**, 2550–2556 (1980).

⁷M. Moskowitz, P. Gartside, and C. McLaughlin, "Mammographic patterns as markers for high-risk benign breast disease and incident cancers," *Radiology* **134**, 293–295 (1980).

⁸I. Witt, H. S. Hansen, and S. Brunner, "The risk of developing breast cancer in relation to mammography findings," *Eur. Radiol.* **4**, 65–67 (1984).

⁹S. Ciatto and M. Zappa, "A prospective study of the value of mammographic pattern as indicators of breast cancer risk in a screening experience," *Eur. Radiol.* **17**, 122–125 (1993).

¹⁰E. Thurfjell, C. C. Hsieh, L. Lipworth, A. Ekblom, H. O. Adami, and D. Trichopoulos, "Breast size and mammographic pattern in relation to breast cancer risk," *Eur. J. Cancer Prevention* **5**, 37–41 (1996).

¹¹I. Kato, C. Beirart, A. Bleich, S. Su, M. Kim, and P. G. Toniolo, "A nested case-control study of mammographic patterns, breast volume and

- breast cancer (New York City, NY, United States),” *Cancer Causes & Control* **6**, 431–438 (1995).
- ¹² E. Sala, R. Warren, J. McCann, S. Duffy, N. Day, and R. Luben, “Mammographic parenchymal patterns and mode of detection: implications for the breast screening programme,” *J. Medical Screening* **5**, 207–212 (1998).
 - ¹³ T. M. Salminen, I. E. Saarenmaa, M. M. Heikkilä, and M. Hakama, “Is a dense mammographic parenchymal pattern a contraindication to hormonal replacement therapy?,” *Acta Oncol.* **39**, 969–972 (2000).
 - ¹⁴ C. Byrne, C. Schairer, J. N. Wolfe, N. Parekh, M. Salane, L. A. Brinton, R. Hoover, and R. Haile, “Mammographic features and breast cancer risk: Effects with time, age, and menopause status,” *J. Natl. Cancer Inst.* **87**, 1622–1629 (1995).
 - ¹⁵ N. F. Boyd, C. Greenberg, G. Lockwood, L. Little, L. Martin, J. Byng, Y. Martin, and D. Tritchler, “Effects at two years of a low-fat, high-carbohydrate diet on radiologic features of the breast: Results from a randomized trial,” *J. Natl. Cancer Inst.* **89**, 466–467 (1997).
 - ¹⁶ D. V. Spicer, G. Ursin, Y. R. Parisky, J. G. Pearce, D. Shoupe, A. Pike, and M. C. Pike, “Changes in mammographic densities induced by a hormonal contraceptive designed to reduce breast cancer risk,” *J. Natl. Cancer Inst.* **86**, 431–436 (1994).
 - ¹⁷ J. Brisson, R. Verreault, A. S. Morrison, D. Tennina, and F. Meyer, “Diet, mammographic features of breast tissue, and breast cancer risk,” *Am. J. Epidemiol.* **130**, 14–24 (1989).
 - ¹⁸ J. N. Wolfe, “Mammography: Ducts as a sole indicator of breast carcinoma,” *Radiology* **89**, 206–210 (1967).
 - ¹⁹ J. N. Wolfe, “The prominent duct pattern as an indicator of cancer risk,” *Oncology* **23**, 149–158 (1969).
 - ²⁰ J. N. Wolfe, “Breast patterns as an index of risk for developing breast cancer,” *Am. J. Roentgenol.* **126**, 1130–1139 (1976).
 - ²¹ J. N. Wolfe, “Risk for breast cancer development determined by mammographic parenchymal pattern,” *Cancer (N.Y.)* **37**, 2486–2492 (1976).
 - ²² I. E. Magnin, F. Cluzeau, C. L. Odet, and A. Bremond, “Mammographic texture analysis: An evaluation of risk for developing breast cancer,” *Opt. Eng.* **25**, 780–784 (1986).
 - ²³ J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe, “Automated analysis of mammographic densities,” *Phys. Med. Biol.* **41**, 909–923 (1996).
 - ²⁴ J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe, “The quantitative-analysis of mammographic densities,” *Phys. Med. Biol.* **39**, 1629–1638 (1994).
 - ²⁵ M. J. Yaffe, N. F. Boyd, J. W. Byng, R. A. Jong, R. Fishell, G. A. Lockwood, L. E. Little, and D. L. Tritchler, “Breast cancer risk and measured mammographic density,” *Eur. J. Cancer Prevention* **7**, S47–S55 (1998).
 - ²⁶ Z. Huo, M. L. Giger, D. E. Wolverton, and W. Zhong, “Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: Feature selection,” *Med. Phys.* **27**, 4–12 (2000).
 - ²⁷ J. J. Heine and R. P. Velthuisen, “A statistical methodology for mammographic density detection,” *Med. Phys.* **27**, 2644–2651 (2000).
 - ²⁸ J. M. Boone, K. K. Lindfors, C. S. Veatty, and J. A. Seibert, “A breast density index for digital mammograms based on radiologists’ ranking,” *J. Digit Imaging* **11**, 101–115 (1998).
 - ²⁹ J. N. Wolfe, A. F. Saftlas, and M. Salane, “Evaluation of mammographic densities: A case-control study,” *Am. J. Roentgenol., Radium Ther. Nucl. Med.* **148**, 1087–1092 (1987).
 - ³⁰ *American College of Radiology. Breast Imaging—Reporting and Data System (BI-RADS)*, 3rd ed. (American College of Radiology, Reston, VA, 1998).
 - ³¹ E. White, P. Velentgas, M. T. Mandelson, C. D. Lehman, J. G. Elmore, P. Porter, Y. Yasui, and S. H. Taplin, “Variation in mammographic breast density by time in menstrual cycle among women aged 40–49 years,” *J. Natl. Cancer Inst.* **90**, 906–910 (1998).
 - ³² C. Zhou, H. P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, B. Sahiner, and L. M. Hadjiiski, “Computerized image analysis: Estimation of breast density on mammograms,” *Med. Phys.* **28**, 1056–1069 (2001).
 - ³³ S. J. Graham, M. J. Bronskill, J. W. Byng, M. J. Yaffe, and N. F. Boyd, “Quantitative correlation of breast tissue parameters using magnetic resonance and x-ray mammography,” *Br. J. Cancer* **73**, 162–168 (1996).
 - ³⁴ N. A. Lee, H. Rusinek, J. Weinreb, R. Chandra, H. Toth, C. Singer, and G. Newstead, “Fatty and fibroglandular tissue volumes in the breasts of women 20–83 years old: comparison of x-ray mammography and computer-assisted MR imaging,” *Am. J. Roentgenol.* **168**, 501–506 (1997).
 - ³⁵ B. Sahiner, N. Petrick, H. P. Chan, L. M. Hadjiiski, C. Paramagul, M. A. Helvie, and M. N. Gurcan, “Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization,” *IEEE Trans. Med. Imaging* **20**, 1275–1284 (2001).
 - ³⁶ B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, “Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images,” *IEEE Trans. Med. Imaging* **15**, 598–610 (1996).
 - ³⁷ O. Pawluczyk, B. J. Augustine, M. J. Yaffe, D. Rico, J. Yang, G. E. Mawdsley, and N. F. Boyd, “A volumetric method for estimation of breast density on digitized screen-film mammograms,” *Med. Phys.* **30**, 352–364 (2003).

Mammographic Density Measured with Quantitative Computer-aided Method:

Comparison with Radiologists' Estimates and BI-RADS Categories¹

Katherine E. Martin, MD
Mark A. Helvie, MD
Chuan Zhou, PhD
Marilyn A. Roubidoux, MD
Janet E. Bailey, MD
Chintana Paramagul, MD
Caroline E. Blane, MD
Katherine A. Klein, MD
Seema S. Sonnad, PhD
Heang-Ping Chan, PhD

Purpose:

To retrospectively compare computer-aided mammographic density estimation (MDEST) with radiologist estimates of percentage density and Breast Imaging Reporting and Data System (BI-RADS) density classification.

Materials and Methods:

Institutional Review Board approval was obtained for this HIPAA-compliant study; patient informed consent requirements were waived. A fully automated MDEST computer program was used to measure breast density on digitized mammograms in 65 women (mean age, 53 years; range, 24–89 years). Pixel gray levels in detected breast borders were analyzed, and dense areas were segmented. Percentage density was calculated by dividing the number of dense pixels by the total number of pixels within the borders. Seven breast radiologists (five trained with MDEST, two not trained) prospectively assigned qualitative BI-RADS density categories and visually estimated percentage density on 260 mammograms. Qualitative BI-RADS assessments were compared with new quantitative BI-RADS standards. The reference standard density for this study was established by allowing the five trained radiologists to manipulate the MDEST gray-level thresholds, which segmented mammograms into dense and nondense areas. Statistical tests performed include Pearson correlation coefficients, Bland-Altman agreement method, κ statistics, and unpaired *t* tests.

Results:

There was a close correlation between the reference standard and radiologist-estimated density ($R = 0.90$ – 0.95) and MDEST density ($R = 0.89$). Untrained radiologists overestimated percentage density by an average of 37%, versus 6% for trained radiologists ($P < .001$). MDEST showed better agreement with the reference standard (average overestimate, 1%; range, -15% to $+18\%$). MDEST correlated better with percentage density than with qualitative BI-RADS categories. There were large overlaps and ranges of percentage density in qualitative BI-RADS categories 2–4. Qualitative BI-RADS categories correlated poorly with new quantitative BI-RADS categories, and 16 (6%) of 260 views were erroneously classified by MDEST.

Conclusion:

MDEST compared favorably with radiologist estimates of percentage density and is more reproducible than radiologist estimates when qualitative BI-RADS density categories are used. Qualitative and quantitative BI-RADS density assessments differed markedly.

© RSNA, 2006

¹ From Associated Radiologists Limited, Mesa, Ariz (K.E.M.); Department of Radiology, University of Michigan Health Center, Ann Arbor, Mich (M.A.H., C.Z., M.A.R., J.E.B., C.P., C.E.B., K.A.K., H.P.C.); and Department of Surgery, University of Pennsylvania School of Medicine, Philadelphia, Pa (S.S.S.). Received November 16, 2004; revision requested January 18, 2005; revision received April 3; accepted May 2; final version accepted November 7. Supported in part by U.S. Army Medical Research and Material Command grant DAMD 17-01-1-0326. Address correspondence to K.E.M., East Valley Diagnostic Imaging, 1125 E Southern Ave, Suite 200, Mesa, AZ 85204 (e-mail: yango@cox.net).

Previous breast density assessments performed by using the Breast Imaging Reporting and Data System (BI-RADS) have been completely qualitative. The new (fourth edition) BI-RADS involves combined qualitative and quantitative assessments (1). The quantitative assessments are divided into quartiles, with category 1 indicating breast tissue that is less than 25% glandular; category 2, breast tissue that is approximately 25%–50% glandular; category 3, breast tissue that is approximately 51%–75% glandular; and category 4, breast tissue that is more than 75% glandular (1). The qualitative descriptive terms remain the same (1). The correlation between the new quantitative assessments and the conventional qualitative density assessments has not been well studied. The main purpose of using the BI-RADS density categories is to indicate the relative sensitivity of the mammographic examination in the detection of breast carcinoma, which may be lower in cases of dense breasts (1). Quantitative assessment of breast density may enable a more precise determination of differences in breast density.

Mammographic density is important for two main reasons: First, the sensitivity of mammography in the detection of breast carcinoma is lower in dense breasts because dense fibroglandular tissue may obscure calcifications and masses (1–3). Second, there is a direct association between increased mammographic density and increased risk of developing breast cancer (4–10). In addition, investigators who use quantitative assessment of mammographic density report higher odds ratios for the development of breast carcinoma in women with dense breasts compared with the odds ratios reported by investigators who use subjective assessment of density (7,8,11). Boyd et al (6) confirmed the importance of using precise methods to determine mammographic density: They observed a 2% increase in the relative risk of breast cancer for every 1% increase in mammographic density percentage.

There is also evidence that hormonal therapies, including estrogen and

tamoxifen treatments, can change mammographic density (9,12–14) and alter the risk of breast carcinoma (15–18). Whether this relationship is causal remains to be proved. A simple and accurate method of measuring breast density would be a useful tool for investigating breast cancer risk–mammographic density relationships.

Several methods to objectively quantitate mammographic density exist. The original method, described by Wolfe et al (11) in 1987, involved the use of manual planimetry to compute the density percentage: The dense white areas on mammograms were manually traced. However, as the authors themselves noted, this method was “tedious and time-consuming.” More recent techniques have been facilitated by the advent of digital methods of acquiring and viewing mammographic data. Although these methods involve the use of computers, some of them are only partially automated (19,20). One such method was based on an ordinal ranking system rather than on a density percentage system (20). More recent computerized programs have been fully automated (10,21,22).

We developed a method in which a fully automated mammographic density estimation (MDEST) program is used to rapidly determine the perimeter of the breast and quantitate the mammographic density percentage (23). Thus, the purpose of our study was to retrospectively compare mammographic densities determined by using this MDEST program with both radiologists’ estimates of density percentage and BI-RADS breast density categories.

Materials and Methods

Mammogram Selection and Digitization

The authors had control of the data and the information submitted for publication. The data set comprised the four-view craniocaudal (CC) and mediolateral oblique (MLO) mammograms obtained in 65 patients who were randomly selected from a National Cancer Institute–designated comprehensive cancer center (University of Michigan Health

Center) after institutional review board approval was obtained. This is the same mammogram data set used in a prior study (23). The requirement for individual patient informed consent was waived. Our study was HIPAA compliant. We originally selected 50 consecutive normal four-view screening mammograms for analysis. Four months later, to include more qualitative BI-RADS density category 3 and 4 mammograms—since these were underrepresented in the original sample—we selected an additional 15 consecutive normal four-view mammograms that had been qualitatively determined to be dense. The ages of the 65 women ranged from 24 to 89 years (mean, 53 years).

The mammograms had been acquired by using Mammography Quality Standards Act (MQSA)-approved GE DMR mammography units (GE Medical Systems, Milwaukee, Wis) with Kodak MR2000 (Kodak, Rochester, NY) screen and film systems. All images were digitized by using a LUMISYS 85 laser film scanner (Lumisys, Mountain View, Calif) with a pixel size of 0.05×0.05 mm and 4096 gray levels. The gray levels were linearly proportional to the optical densities, from 0.1 to approximately 4.0 optical density units. The nominal optical density range of the scanner is 0–4, with large pixel values corresponding to low optical

Published online before print

10.1148/radiol.2402041947

Radiology 2006; 240:656–665

Abbreviations:

BI-RADS = Breast Imaging Reporting and Data System
CC = craniocaudal
MDEST = mammographic density estimation
MLO = mediolateral oblique
MQSA = Mammography Quality Standards Act

Author contributions:

Guarantors of integrity of entire study, K.E.M., M.A.H.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, K.E.M., M.A.H., M.A.R., C.P., C.E.B., H.P.C.; clinical studies, M.A.H., M.A.R., C.P., K.A.K.; statistical analysis, M.A.H., C.Z., C.P., S.S.S.; and manuscript editing, K.E.M., M.A.H., C.Z., M.A.R., C.P., C.E.B., K.A.K., S.S.S., H.P.C.

Authors stated no financial relationship to disclose.

density. Since the breast density pattern does not have to be analyzed in high spatial resolution (ie, pixel size of 0.05 mm or less), the full-spatial-resolution mammograms were first smoothed with a 16×16 box filter and subsampled by a factor of 16 to result in 0.8-mm pixel size images that were approximately 256×256 pixels in size for the analysis. This process reduced the processing time and image noise. The technical details are described elsewhere (23). However, a different software version of the density program was used for this study.

Mammogram Density Analysis with MDEST

The computer first tracked the breast boundary by using a gradient-based edge-tracking algorithm, which has been described previously (23). The tracking of the boundaries of a given breast started from approximately the middle of the breast image and continued both upward and downward along the boundary. The direction in which to search for a new edge point was guided by the previous edge points. The edge location was determined by using a gradient criterion along a band of pixels perpendicular to the tracking direction. The detected boundary separated the breast from other background features, including the directly exposed area, patient identification information, and lead markers, which were excluded in the subsequent analyses. Figure 1 shows examples of the breast boundaries determined on typical CC- and MLO-view mammograms. A separate edge-tracking algorithm was used to detect the edge of the pectoral muscle on the MLO-view mammograms (Fig 1). The detected edge usually is not very smooth owing to noise on the image. A second-order polynomial was fitted to the detected edge points to segment the pectoral region. The pectoral muscle on the MLO views was excluded from the subsequent gray-level histogram analyses and breast area calculations.

A dynamic range-compression method was used to reduce the gray-level range of the histograms without affecting the relative areas of the dense tissue region and the entire breast region. The histogram of the breast region on

the dynamic-range-compressed mammogram was generated, normalized, and smoothed. The histograms were analyzed by the computer to formulate an automatic thresholding routine.

After histogram classification, a gray-level threshold was automatically calculated to separate the fat and dense glandular tissue regions. The gray-level threshold depends on the shape (or class) of the histogram. If the histogram has a single peak, the maximum entropy principle-based method (24) is used to calculate the threshold. If the histogram has more than one peak, the discriminant analysis method (25) is used. The threshold is used to separate the pixels in the breast region into two classes: The class of pixel values above the threshold corresponds to dense tissue, and the class of pixel values below the threshold corresponds to fat tissue. This classification is represented on a binary image (ie, segmented image), on which dense pixels are represented by

white and fat pixels are represented by black (Fig 2). The percent breast density is then calculated as the number of pixels in the dense area divided by the total number of pixels in the entire breast region.

Mammogram Density Analysis Performed by Readers

Seven MQSA-certified radiologists (K.E.M., M.A.H., M.A.R., J.E.B., C.P., C.E.B., K.A.K.) independently evaluated the data set. Five of these radiologists (K.E.M., M.A.H., M.A.R., J.E.B., C.P.) were involved in a prior study (23). Before evaluation of the actual data set, a set of 25 training cases was used to familiarize five of the radiologists with the MDEST program and assist them in the visual estimation of the percentage of dense area on the mammograms. Thus, five radiologists were considered to be trained, and two were not. The experience of the radiologists in interpreting mammograms ranged

Figure 1

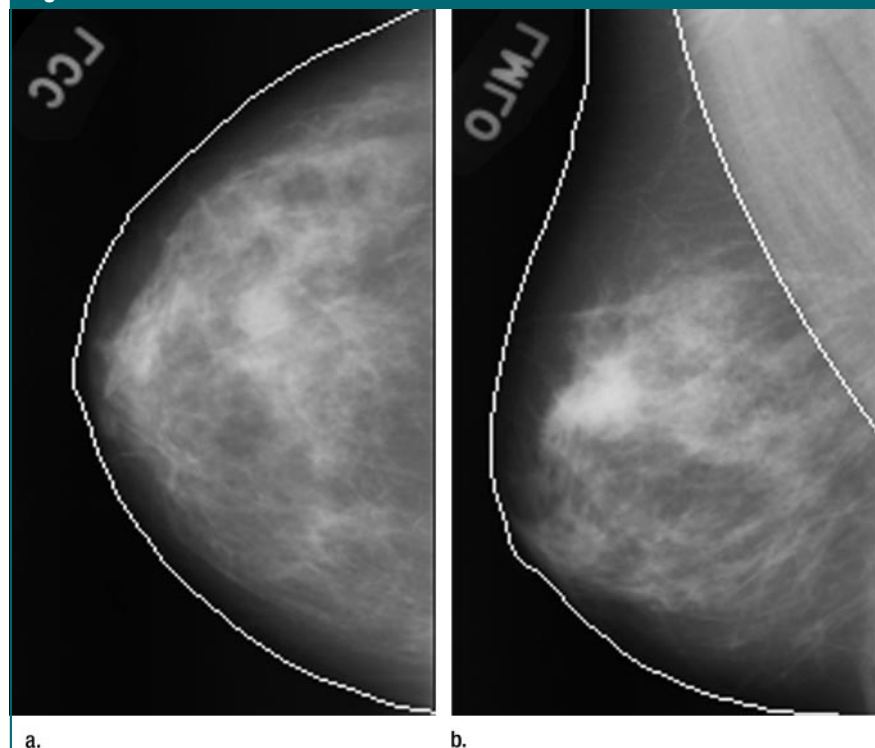


Figure 1: Mammogram from our image database superimposed with the detected breast boundary. (a) CC view. (b) MLO view with anterior breast boundary and pectoral muscle boundary.

from 1 to 25 years (median, 9 years; mean, 10.4 years). The seven radiologists received their residency education at six different institutions. Two were educated at the same institution.

A graphical interface for displaying and recording the radiologists' evaluations was developed. For a given breast, the CC- and MLO-view mammograms

were first displayed side-by-side on a high-spatial-resolution 22-inch Compaq AlphaStation monitor (Compaq, Palo Alto, Calif). This monitor has a display matrix size of 1280×1024 pixels. It is not Digital Imaging and Communications in Medicine calibrated, but it allows one to adjust contrast and brightness settings, and we adjusted these

at the beginning of the study according to the subjective impressions of an experienced MQSA-certified radiologist (M.A.H.). For each mammogram, the radiologists were able to adjust the window and level settings on the display screen.

Qualitative BI-RADS density classifications.—The radiologist first assigned each two-view mammogram to one of the four conventional BI-RADS qualitative density categories (eg, category 1, indicating fat tissue). This BI-RADS density assessment system does not include any quantitative classification used in the new (fourth edition) American College of Radiology BI-RADS (1), which was not published at the time of the study. Herein, the scores used in the new BI-RADS classification system are referred to as "qualitative BI-RADS categories."

Quantitative estimate of density percentage.—Next, the radiologist visually estimated the density percentage on each mammogram by selecting one of the 10% density ranges displayed on the screen. Ten density percentage increments (eg, 1%–10%, 10%–20%) were used because we believed that it would be too difficult for the radiologists to visually estimate density to the nearest 1%.

Determination of reference-standard density.—After the subjective radiologist evaluation, each view (CC or MLO) was displayed sequentially. The displayed material included the original mammogram, the enhanced mammogram, the histogram of the breast region in that view, and the corresponding binary image created by thresholding the histogram. The enhanced image was generated by the MDEST program during the density segmentation. This image was basically a version of the original mammogram with the contrast of structures enhanced. The radiologist was then able to manipulate the gray-level threshold by interactively moving a slider along the horizontal axis of the histogram. The binary image changed simultaneously with the chosen threshold so that the radiologist could determine whether the segmented white area corresponded to the dense white area on the mammogram. The radiologist

Figure 2

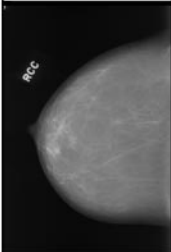
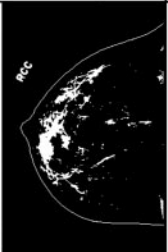
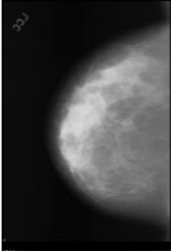
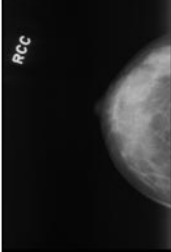
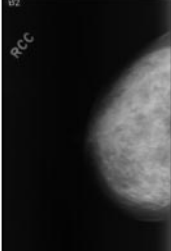
Mammogram	MDEST Segmented Mammogram	Density Reference Standard	Density MDEST	Qualitative BIRADS (range)	Quantitative BIRADS
		2.8%	10.1%	1.2(1-2)	1
		17.2%	19.0%	2.8 (2-3)	1
		33.0%	35.0%	3 (3-3)	2
		52.5%	59.1%	3.6 (3-4)	3

Figure 2: Four representative CC mammograms with corresponding segmented images, density estimates, and BI-RADS categories determined by trained radiologists. Note the difference between the quantitative and qualitative BI-RADS categories. Although the two middle (second and third images from top) mammograms have similar amounts of segmented breast density, the density percentage is greater in the smaller breast owing to less fat tissue.

was instructed to change the amount of segmented dense area to resemble the area that he or she would trace by hand if he or she were performing manual planimetry (11). When the radiologist determined that the segmented area was accurate, he or she clicked a button to record the gray-level threshold and density percentage for each image. Since no reference standard exists for breast density measurements, we used this value—averaged for five radiologists previously trained with the training cases—as the reference-standard density percentage for each view.

The radiologist was blinded to his or her own estimated density percentage value obtained and thus could not attempt to match his or her density percentage estimate for the different views or for different breasts of the same patient. The mammogram of the contralateral breast of the same patient was then displayed and evaluated in the same way. The entire process was repeated for each patient until the imaging data of all patients in the data set were evaluated. We also recorded how long it took the radiologists to complete their evaluations of the mammograms.

During the training session for the five radiologists, both the percentage of dense area derived by the MDEST program and that determined by using interactive thresholding were presented to the radiologists so that they could compare these two percentages with their visually estimated density percentage for each image. The percent dense areas derived by using MDEST and interactive thresholding were not displayed during the actual study. To assess the effect of training, two additional breast imaging radiologists, who had not undergone training to visually estimate density percentage with the 25 training cases, evaluated the same set of study images.

Statistical Analyses

Pearson correlation coefficients were calculated to examine the associations of the qualitative BI-RADS-, MDEST-, and trained radiologist-estimated mammographic densities with the true (ie, reference-standard) mammographic density. To assess the agreement between

Table 1

Descriptive Statistics for Breast Density on CC Views, Estimated by Seven Radiologists

Datum*	Mean Value†	Minimal Value	Maximal Value
Reader 1			
BI-RADS category	2.38 ± 0.85	1	4
MDEST density (%)	0.25 ± 0.15	0.07	0.76
Radiologist density estimate (%)	0.30 ± 0.25	0.05	0.85
Reference-standard density (%)	0.22 ± 0.16	0.02	0.59
Reader 2			
BI-RADS category	2.44 ± 0.90	1	4
MDEST density (%)	0.25 ± 0.15	0.07	0.76
Radiologist density estimate (%)	0.23 ± 0.21	0.05	0.85
Reference-standard density (%)	0.23 ± 0.18	0.02	0.67
Reader 3			
BI-RADS category	2.65 ± 0.88	1	4
MDEST density (%)	0.25 ± 0.15	0.07	0.76
Radiologist density estimate (%)	0.27 ± 0.19	0.05	0.65
Reference-standard density (%)	0.25 ± 0.19	0.02	0.72
Reader 4			
BI-RADS category	2.52 ± 1.08	1	4
MDEST density (%)	0.25 ± 0.15	0.07	0.76
Radiologist density estimate (%)	0.34 ± 0.25	0.05	0.85
Reference-standard density (%)	0.27 ± 0.20	0.01	0.72
Reader 5			
BI-RADS category	2.52 ± 1.02	1	4
MDEST density (%)	0.25 ± 0.15	0.07	0.76
Radiologist density estimate (%)	0.31 ± 0.26	0.05	0.85
Reference-standard density (%)	0.25 ± 0.18	0.02	0.73
Reader 6			
BI-RADS category	2.5 ± 0.90	1	4
MDEST density (%)	0.25 ± 0.15	0.07	0.76
Radiologist density estimate (%)	0.42 ± 0.26	0.05	0.95
Reader 7			
BI-RADS category	2.32 ± 1.06	1	4
MDEST density (%)	0.25 ± 0.15	0.07	0.76
Radiologist density estimate (%)	0.44 ± 0.27	0.05	0.95

* Readers 1–5 were trained in using the MDEST method, and readers 6 and 7 were not. Radiologist estimates of breast density were based on visual estimates of density percentage.

† Mean values ± standard deviations.

the reference-standard density and both the MDEST- and the trained radiologist-estimated densities and to obtain 95% limits of agreement, the method of Bland and Altman was used (26). Interreader agreement among the radiologists was measured by using κ statistics (27). The strengths of agreement were expressed in κ values: A value of 0.20 or less indicated poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; and 0.81–1.00, very good agreement. The significance of differences in overestimations of density between the trained and untrained radiol-

ogists was estimated by using the unpaired *t* test. For the statistical calculations, the radiologists' density percentage estimates were expressed as the mean of the 10% range (eg, for 1%–10%, 5% was used). Software, including SAS (SAS Institute, Cary, NC) and Microsoft Excel (Redmond, Wash), was used to perform all statistical analyses.

Results

We excluded 16 (6%) of the 260 mammographic views owing to technical prob-

lems that were secondary to improper breast boundary detection or the MDEST program's gross misclassification of the gray-level histograms. The MDEST program performed well in most cases. Descriptive statistics for the four methods of evaluating mammographic density are presented, according to radiologist and view, in Tables 1 and 2.

Density Analyses

Pearson correlation coefficients for correlations between the reference-standard

density and the qualitative BI-RADS-, MDEST-, and trained radiologist-estimated densities showed strong positive linear relationships (Table 3). These positive correlations indicate that as the reference-standard density percentage increased, the values obtained with the other methods also tended to increase. Of these three estimated densities, the trained radiologists' estimates had the highest correlation with the reference-standard density ($R = 0.90$), the MDEST measurements had the second

highest correlation ($R = 0.89$), and the qualitative BI-RADS categories had the third highest correlation ($R = 0.85$). The correlation of each method with the reference-standard method was better on the CC views than on the MLO views. Correlation coefficients for agreement between the CC and MLO views of the same breast were 0.85 with MDEST; 0.96 with the reference-standard method; and 0.94, 0.95, 0.94, 0.93, and 0.94 with the estimates of the five trained radiologists.

Agreement of Reference-Standard Density with MDEST- and Trained Radiologist-estimated Densities

The method of Bland and Altman (26) was used to assess agreement between the reference-standard density and both the MDEST- and the trained radiologist-estimated densities. The mean overall bias for the comparison between the trained radiologists' density estimates and the reference-standard measurement was an overestimation of 6%, compared with an overestimation of 1% by the MDEST program (Table 4). The 95% limits of agreement between the trained radiologist-estimated and reference-standard densities were wider (-16% to $+27\%$), indicating greater error in measuring density by using radiologist estimates than by using MDEST. The overall limits of agreement between the reference-standard and MDEST densities ranged from -15% to $+18\%$. The MDEST program tended to overestimate mammographic density (up to $+18\%$) more than it underestimated it (up to -15%).

Untrained Radiologists

The densities estimated by the two untrained breast imagers had excellent correlation with the reference-standard measurement ($R = .95$). The untrained radiologists overestimated the mammographic density percentage to a greater extent than did the trained radiologists. The untrained radiologists overestimated density by 37% with respect to the reference-standard density (Fig 3). This was in contrast to the 6% overestimation of the trained radiologists ($P < .001$). There was no significant differ-

Table 2

Descriptive Statistics for Breast Density on MLO Views, Estimated by Seven Radiologists

Datum*	Mean Value†	Minimal Value	Maximal Value
Reader 1			
BI-RADS category	2.37 ± 0.87	1	4
MDEST density (%)	0.25 ± 0.17	0.06	0.82
Radiologist density estimate (%)	0.31 ± 0.26	0.05	0.95
Reference-standard density (%)	0.20 ± 0.16	0.01	0.74
Reader 2			
BI-RADS category	2.42 ± 0.91	1	4
MDEST density (%)	0.25 ± 0.17	0.06	0.82
Radiologist density estimate (%)	0.23 ± 0.21	0.05	0.85
Reference-standard density (%)	0.22 ± 0.17	0.03	0.77
Reader 3			
BI-RADS category	2.65 ± 0.87	1	4
MDEST density (%)	0.25 ± 0.17	0.06	0.82
Radiologist density estimate (%)	0.27 ± 0.19	0.05	0.75
Reference-standard density (%)	0.24 ± 0.18	0.02	0.76
Reader 4			
BI-RADS category	2.51 ± 1.07	1	4
MDEST density (%)	0.25 ± 0.17	0.06	0.82
Radiologist density estimate (%)	0.34 ± 0.25	0.05	0.85
Reference-standard density (%)	0.26 ± 0.2	0.01	0.82
Reader 5			
BI-RADS category	2.52 ± 1.03	1	4
MDEST density (%)	0.25 ± 0.17	0.06	0.82
Radiologist density estimate (%)	0.31 ± 0.26	0.05	0.85
Reference-standard density (%)	0.23 ± 0.17	0.02	0.8
Reader 6			
BI-RADS category	2.5 ± 0.90	1	4
MDEST density (%)	0.25 ± 0.17	0.06	0.82
Radiologist density estimate (%)	0.42 ± 0.26	0.05	0.95
Reader 7			
BI-RADS category	2.32 ± 1.06	1	4
MDEST density (%)	0.25 ± 0.17	0.06	0.82
Radiologist density estimate (%)	0.44 ± 0.27	0.05	0.95

* Readers 1–5 were trained in using the MDEST method, and readers 6 and 7 were not. Radiologist estimates of breast density were based on visual estimates of density percentage.

† Mean values ± standard deviations.

ence in the assignment of qualitative BI-RADS density categories between the trained and untrained radiologists ($P > .43$).

Comparison with BI-RADS

A wide range of percent densities were assigned to mammograms classified in three of the four qualitative BI-RADS categories (BI-RADS 2–4) (Table 5). For example, qualitative BI-RADS category 4 included mammograms with densities of between 20% and 82%. There was also a large range of qualitative BI-RADS categories assigned to the same range of percent densities. A mammogram with 0%–24% density may have been assigned to qualitative BI-RADS category 1, 2, or 3 (Table 6) (Fig 2).

In Table 6, the new quantitative BI-RADS density quartiles (0%–24%, 25%–49%, 50%–74%, and 75%–100%) are compared with the conventional qualitative BI-RADS categories. The reference-standard percent densities are grouped to mimic the new quantitative BI-RADS breast density quartiles. There was poor agreement between the qualitative BI-RADS density classifications and the new quantitative BI-RADS density classifications, with use of the reference-standard density as truth. If the new combined qualitative and quantitative BI-RADS category 4—indicating extremely dense breast tissue—corresponded to greater than 75% breast density, far fewer cases would be classified as BI-RADS 4 compared with the number of cases that would be classified as qualitative BI-RADS 4. With use of the qualitative BI-RADS system, 110 (17%) of the 650 cases (130 breasts times five radiologists equals 650 cases) in our study were assigned to category 4. With use of the new quantitative system, none of the 650 cases was assigned to category 4. With use of the old qualitative BI-RADS system, 108 (17%) of the 650 cases were classified as fatty (category 1); with use of the new BI-RADS quantitative system, 370 (57%) cases were classified as fatty (category 1), representing a downstaging of 40%.

Interobserver Agreement

The interobserver agreement values observed for each density measurement

Table 3

Correlation of Reference-Standard Breast Density with Qualitative BI-RADS Categories, MDEST Density, and Trained Radiologists' Density Estimates

Correlation Method	Density Estimate Correlated with Reference Standard		
	Qualitative BI-RADS	MDEST	Trained Radiologist Estimate
By view			
CC	0.8682	0.9013	0.9124
MLO	0.8343	0.8880	0.8962
Overall	0.8509	0.8914	0.9035

Note.—Data are Pearson correlation coefficients.

Table 4

Mean Agreement between Density Measurement Methods

Correlation Method	Mean Bias*	95% Limits of Agreement	
		Lower	Upper
Reference-Standard versus MDEST Density			
By view			
CC	-0.0036 ± 0.0797	-0.1630	0.1558
MLO	-0.0244 ± 0.0824	-0.1892	0.1404
Overall	-0.01384 ± 0.0816	-0.1770	0.1494
Reference-Standard Density versus Trained Radiologists' Estimates			
By view			
CC	-0.0469 ± 0.1029	-0.2527	0.1589
MLO	-0.0634 ± 0.1117	-0.2868	0.1600
Overall	-0.0551 ± 0.1076	-0.2703	0.1601

* Mean bias values ± standard deviations.

method indicate that there was strong agreement among the trained radiologists. Intraclass correlation coefficients were 0.88 for the radiologists' estimates of density percentage and 0.94 for the radiologists' determinations of the reference-standard density percentage, indicating very good agreement among the radiologists' density measurements obtained with these two methods. Pairwise comparison of the radiologists' assignments of qualitative BI-RADS categories revealed good but lower agreement, with κ values (27) ranging from 0.61 to 0.76.

Time

The mean time to complete the qualitative BI-RADS density category assignments and density percentage estimations with the MDEST program was 18 seconds per view (range, 13–22 seconds), with a mean standard

deviation of 8 seconds (range, 7–9 seconds).

Discussion

Our study findings demonstrate that quantified breast density is a more accurate and reproducible measure of breast density than radiologist estimates derived by using the conventional qualitative BI-RADS density categories. In our study, we observed large ranges of percent densities among cases classified in three of the four qualitative BI-RADS density categories, with up to a 62% range in the category 4 cases. We also observed a large overlap in the mammographic density percentages assigned to cases classified in the four qualitative BI-RADS categories by five breast imaging specialists, with up to a 40% range overlap between categories 3 and 4. A mammogram with 30% total

breast density could be assigned to qualitative BI-RADS category 2, 3, or 4; however, under the new system, it would be assigned to category 2.

Similar results were found in a comparison between the qualitative estimates based on the Wolfe parenchymal patterns and the quantitative determinations of density made by using manual planimetry (28). This is not surprising, given the subjective nature of both the qualitative BI-RADS density categories and the Wolfe parenchymal patterns. More recently, Wang et al (21) suggested that the visual density percentage estimates derived by three mammographers may have led to the same mammogram being assigned to different qualitative BI-RADS categories. With use of quantitative percent breast density determinations, one is more likely to detect subtle changes in breast density that may be masked when they

are classified in the same BI-RADS category as the overall breast density.

Our results show that experienced radiologists' subjective density assessments based on qualitative BI-RADS categories may be quite different from density assessments based on quantitative BI-RADS categories. For example, none of the 650 mammographic cases judged to have qualitative category 4 density had greater than 75% breast density according to quantitative BI-RADS measures. In addition, the cases with 0%–24% quantitative BI-RADS category 1 (fatty) breast density would have encompassed a majority (370

[57%] of 650) of the mammograms, many of which were conventionally assigned to qualitative BI-RADS category 2 or 3.

A goal of the BI-RADS system is to facilitate uniformity of physician reports, and our results suggest that additional training would be necessary to enable physicians to accurately translate a visual assessment of density percentage into a quantitative assessment, as recommended by the new BI-RADS standards (1). In our study, the two untrained radiologists overestimated density by 37%. However, radiologists could be rapidly trained to estimate breast density by using a computerized density measurement program so that they could follow the new quantitative BI-RADS density classifications. Furthermore, use of the new quantitative BI-RADS assessment may lead to the "down coding" of breast density and thus the creation of nonuniformity between the old and new standards and the consequent hindering of longitudinal research. A case previously assigned to BI-RADS category 3 may now be assigned to BI-RADS category 1. Since very few breasts have greater than 75% density, the new quantified BI-RADS may functionally approach a three-density-level system, with the majority of mammographic cases assigned to categories 1 and 2.

The MDEST breast density determinations were more accurate than the radiologists' visual estimates of breast density, as indicated by the radiologists' overestimating of breast density to a greater degree and their larger variation in density estimates (relative to the reference-standard density) compared with the MDEST measurements. We also observed good correlations between MDEST-derived density percentage and radiologist-determined reference-standard density percentage ($R = .89$). Although this correlation was slightly lower than that between the trained radiologist-estimated and reference-standard densities ($R = 0.90$), the MDEST-derived densities had tighter agreement with the reference-standard measurements than did the radiologists' estimates. The MDEST program overesti-

Table 5

Range of Mammographic Density Percentages Assigned to Qualitative BI-RADS Categories by Trained Radiologists

Qualitative BI-RADS Category	Density Percentage Range (%)
1	1–11
2	2–45
3	8–60
4	20–82

Figure 3

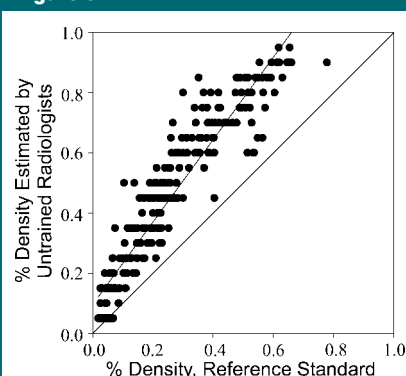


Figure 3: Graph illustrates mammographic density percentage visually estimated by two untrained radiologists versus reference-standard density percentage. The untrained radiologists significantly overestimated the percent breast density relative to the reference-standard density determined by the trained radiologists ($P < .001$).

Table 6

Qualitative BI-RADS versus Quantitative Density Determinations

Quantitative Density Percentage [†]	Qualitative BI-RADS Score*			
	1	2	3	4
0–24	17 (108/650)	29 (189/650)	11 (73/650)	0
25–49	0	3 (20/650)	21 (136/650)	6 (39/650)
50–74	0	0	2 (14/650)	11 (71/650)
75–100	0	0	0	0

* Data are percentages of cases with the given quantitative density percentage that were assigned the given BI-RADS score. The numbers of cases used to calculate the percentages are in parentheses.

[†] The quantitative (ie, reference-standard) breast density ranges correspond to the new quantitative BI-RADS density categories. Each breast (depicted on CC and MLO views) was assigned a single qualitative BI-RADS score by five trained radiologists (130 breasts times five radiologists equals 650 cases).

mated density by a mean of only 1% (range, -15% to +18%), as compared with a mean overestimation of density of 6% by the trained radiologists (with a wider range: -16% to +27%) relative to the reference-standard density percentage.

Correlation coefficients for agreement on density measurement between the CC and MLO views favored the trained radiologists, although both the MDEST program and the radiologists had good correlation. The MDEST program performed better than the untrained radiologists in the estimation of percent breast density. The two untrained breast imagers tended to overestimate density percentage by approximately 37%, which was greater than the percentage of overestimation by the trained breast imagers. These findings are in contrast to those of Lee-Han et al (28): The single radiologist in their study slightly underestimated the density percentage relative to the measured area of density. This result may have been secondary to some form of density percentage estimation training received by the radiologist, although this was not specified.

Many of the computer programs previously used to evaluate breast density have been only partially automated. The density measurement methods used by Byng et al (19) and Boone et al (20) involved manual cropping of the pectoral muscle to determine the breast area on mammograms. In addition, the Byng et al method involved manual determinations of both breast edge and breast density gray-level thresholds. With the Byng et al method, it took less than a minute to evaluate each image (19). Our fully automated program automatically detects the breast edge, crops the pectoral muscle, and estimates the gray-level threshold for density segmentation. In addition, if manual interactive thresholding (the reference-standard method used in the current study) is preferred, the MDEST user interface is fast and simple to use, requiring an average of 18 seconds per view to evaluate both the BI-RADS category-based density and the density percentage. Investigators in two other

studies (10,21) have described fully automated programs for determining breast density.

We found that percent breast density determinations were more accurate on CC views than on MLO views. This was true for both the MDEST densities and the radiologist visual density estimates. The MDEST program overestimated density percentage by a mean of 2.4% on the MLO views and by a mean of 0.4% on the CC views. The radiologists overestimated density by a mean of 6.3% on the MLO views and by a mean of 4.7% on the CC views. These data suggest that in the future, CC views alone may be adequate for assessing percent breast density in temporal measurements.

There were several limitations to our study. The MDEST program had technical errors, which led to a 6% case rejection rate. Technical errors included inaccurate breast border detection and gross misclassification of the gray-scale histograms. Errors in both the anterior breast border detection algorithm and the pectoral muscle detection algorithm occurred and resulted in inaccurate breast tissue area determinations. Misclassification of the gray-scale histograms resulted in improper gray-level threshold determination, inaccurate segmentation of the dense areas, and inaccurate density percentage calculations. These histogram misclassification errors occurred more often on the mammograms with extremely dense and fatty pixels. Thus, MDEST cannot yet be used as a stand-alone density measurement method. Although further development of computer visualization techniques and additional training with a large data set are needed to improve the accuracy and robustness of MDEST, the results of this study demonstrate the feasibility of our approach and the promise of using an automated or semi-automatic system like MDEST to aid future research efforts in the investigation of mammographic breast density.

Another limitation of our study was that the BI-RADS qualitative assessments were subjective and could be institutionally defined. Six (86%) of the seven radiologists received residency

training at different institutions, so strong institutional bias was less likely in this study. Also, there is no reference standard for determining breast density, so there will always be some subjective difference in determining mammographic density, even when manual segmentation is used. The averaging of five radiologists' segmentations may have partially reduced this bias.

Our MDEST program calculates the area of mammographic breast density, which correlates with the area of fibroglandular tissue that is present. However, volume is a more accurate measure of the amount of breast tissue than is area. Accurate determination of the dense tissue volume requires densitometry and scatter and beam-hardening corrections for each mammogram. A rough estimate of dense tissue volume could be determined by multiplying the breast thickness, which is recorded for each mammogram at our institution, by the area of dense tissue. Wang et al (21) described a computer-aided detection method that is more accurate for estimating dense mammographic tissue composition because it involves the use of a tissue-thickness-correction algorithm. This concept of breast tissue volume may be of importance in the study of breast cancer risk, because it is probably the volume of dense glandular breast tissue—rather than the density of breast tissue—that determines risk (9). Wei et al (22) recently observed a high correlation between our automated mammographic MDEST assessment method and volumetric fibroglandular tissue estimation at breast magnetic resonance imaging, suggesting that estimates of change in mammographic density are close surrogates for change in volumetric density. Further investigation is needed to determine whether rough estimates of dense tissue volume can improve the correlation between breast density and breast cancer risk.

In conclusion, the MDEST-derived densities compared favorably to radiologist estimates of percent breast density and were more reproducible than radiologist estimates of the conventional qualitative BI-RADS density categories.

Qualitative and quantitative BI-RADS density assessments differed markedly.

Acknowledgment: We acknowledge the statistical assistance of Acham Gebremariam, MS, of CHOICES at the University of Michigan.

References

1. American College of Radiology. American College of Radiology Breast Imaging Reporting and Data System (BI-RADS). 4th ed. Reston, Va: American College of Radiology, 2003.
2. White E, Velentgas P, Mandelson MT, et al. Variation in mammographic breast density by time in menstrual cycle among women aged 40–49 years. *J Natl Cancer Inst* 1998; 90:906–910.
3. Rosenberg RD, Hunt WC, Williamson MR, et al. Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183 134 screening mammograms in Albuquerque, New Mexico. *Radiology* 1998;209:511–518.
4. Boyd NF, Byng RA, Jong EK, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *J Natl Cancer Inst* 1995;87:670–675.
5. Boyd NF, Lockwood GA, Martin LJ, et al. Mammographic densities and risk of breast cancer among subjects with a family history of this disease. *J Natl Cancer Inst* 1999;91: 1404–1408.
6. Boyd NF, Greenberg C, Lockwood G, et al. Effects at 2 years of a low-fat, high-carbohydrate diet on radiologic features of the breast: results from a randomized trial. *J Natl Cancer Inst* 1997;89:488–496.
7. Byrne C, Schairer CS, Wolfe J, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *J Natl Cancer Inst* 1995;87:1622–1628.
8. Warner E, Lockwood G, Math M, Tritchler D, Boyd NF. The risk of breast cancer associated with mammographic parenchymal patterns: a meta-analysis of the published literature to examine the effect of method of classification. *Cancer Detect Prev* 1992;16: 67–72.
9. Heine JJ, Malhotra P. Mammographic tissue, breast cancer risk, serial image analysis, and digital mammography. *Acad Radiol* 2002;9:298–335.
10. Boyd NF, Dite GS, Stone J, et al. Heritability of mammographic density, a risk factor for breast cancer. *N Engl J Med* 2002;347:886–894.
11. Wolfe JN, Saftlas AF, Salane M. Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a case-control study. *AJR Am J Roentgenol* 1987;148:1087–1092.
12. Stomper PC, Van Voorhis BJ, Ravnikaar VA, Meyer JE. Mammographic changes associated with postmenopausal hormone replacement therapy: a longitudinal study. *Radiology* 1990;174:487–490.
13. Laya MB, Gallagher JC, Schreiman JS, Larson EB, Watson P, Weinstein L. Effect of postmenopausal hormonal replacement therapy on mammographic density and parenchymal pattern. *Radiology* 1995;196: 433–437.
14. Son HJ, Oh KK. Significance of follow-up mammography in estimating the effect of tamoxifen in breast cancer patients who have undergone surgery. *AJR Am J Roentgenol* 1999;173:905–909.
15. Colditz GA, Hankinson SE, Hunter DJ, et al. The use of estrogens and progestins and the risk of breast cancer in postmenopausal women. *N Engl J Med* 1995;332:1589–1593.
16. Ross RK, Paganini-Hill A, Wan PC, Pike MC. Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin. *J Natl Cancer Inst* 2000;92: 328–332.
17. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* 1998;90:1371–1388.
18. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321–333.
19. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. *Phys Med Biol* 1994;39: 1629–1638.
20. Boone JM, Lindfors KK, Beatty CS, Seibert JA. A breast density index for digital mammograms based on radiologists' ranking. *J Digit Imaging* 1998;11:101–115.
21. Wang XH, Good WF, Chapman BE, et al. Automated assessment of the composition of breast tissue revealed on tissue-thickness-corrected mammography. *AJR Am J Roentgenol* 2003;180:257–262.
22. Wei J, Chan HP, Helvie MA, et al. Correlation between mammographic density and volumetric fibroglandular tissue estimated on breast MR images. *Med Phys* 2004;31(4): 933–942.
23. Zhou C, Chan HP, Petrick N, et al. Computerized image analysis: estimation of breast density on mammograms. *Med Phys* 2001; 28:1056–1069.
24. Wong AK. A gray-level threshold selection method based on maximum entropy principle. *IEEE Trans System Man Cybernetics* 1989;19:866–871.
25. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans System Man Cybernetics* 1979;9:62–66.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1: 307–310.
27. Altman DG. Practical statistics for medical research. New York, NY: Chapman & Hall, 1991.
28. Lee-Han H, Cooke G, Boyd NF. Quantitative evaluation of mammographic densities: a comparison of methods of assessment. *Eur J Cancer Prev* 1995;4:285–292.

Breast Density Estimation: Correlation of Mammographic Density and MR Volumetric Density

Heang-Ping Chan, Lubomir M. Hadjiiski, Marilyn A. Roubidoux, Mark A. Helvie,
Sophie Paquerault, Berkman Sahiner, Jun Wei, Chuan Zhou,
Thomas Chenevert, Mitchell M. Goodsitt

Department of Radiology, University of Michigan, Ann Arbor, MI 48109
Chanhp@umich.edu

Abstract. Studies have demonstrated a strong correlation between mammographic breast density and breast cancer risk. Mammographic breast density may therefore be used as a surrogate marker for monitoring the response to treatment in studies of breast cancer prevention or intervention methods. In this study, we evaluated the accuracy of using mammograms for estimating breast density by analyzing the correlation between the percent mammographic dense area and the percent glandular tissue volume as estimated from MR images. A data set of 37 patients who had corresponding MR images and mammograms was collected. The glandular tissue regions in the MR slices were segmented by a semi-automatic method and the percent glandular tissue volume calculated. Mammographic breast density was estimated by an automated image analysis program. It was found that the correlation between the percent dense area of the CC and MLO views and the percent volumetric fibroglandular tissue on MR images was 0.93 and 0.91, respectively, with a mean bias of 4.4%. The high correlation indicates the usefulness of mammographic density as a surrogate for breast density estimation.

1. Introduction

Previous studies have shown that there is a strong positive correlation between breast parenchymal density on mammograms and breast cancer risk [1-3]. The relative risk is estimated to be about 4 to 6 times higher for women whose mammograms have parenchymal densities over 60% of the breast area, as compared to women with less than 5% of parenchymal densities. The change in mammographic breast density is therefore often used as an indicator for monitoring the effects of preventive or interventional treatment of breast cancer.

Breast cancer risk is expected to depend on the volume of glandular tissue in the breast. Mammographic density is a projection of the volume of glandular tissue onto the two-dimensional image plane. To better understand the correlation between mammographic density and breast cancer risk, it is important to investigate the relationship between the projected areal density on mammograms and the volume of glandular tissue in the breast. In this study, we investigate this relationship by analyzing the percent volumetric glandular tissue in magnetic resonance (MR) images and the percent dense area in corresponding mammograms for the same breasts.

2. Materials and Methods

Our data set consisted of corresponding MR breast images and mammograms from 37 patients acquired between detection and biopsy. The MR image series used in this study, which included coronal 3D T_1 -weighted pre-contrast images (coronal sections 2-5 mm thick, 32 slices; 3D Spoiled Gradient-Recalled Echo (SPGR); TE=3.3ms; TR=10ms, Flip=40°, matrix=256x128, FOV=28-32cm right/left, 14-16cm superior/inferior, scan time=2 min 38 sec) was part of a dynamic breast MRI study. This 3D SPGR sequence produced full volume coverage of both breasts with contiguous image slices. An example of images from one breast is shown in Fig. 1. Although this is not the optimal pulse sequence for separating water and fat, the fibroglandular parenchyma (~water) and fatty tissue are well separated with this heavily T_1 -weighted acquisition and therefore the series was chosen for this study.

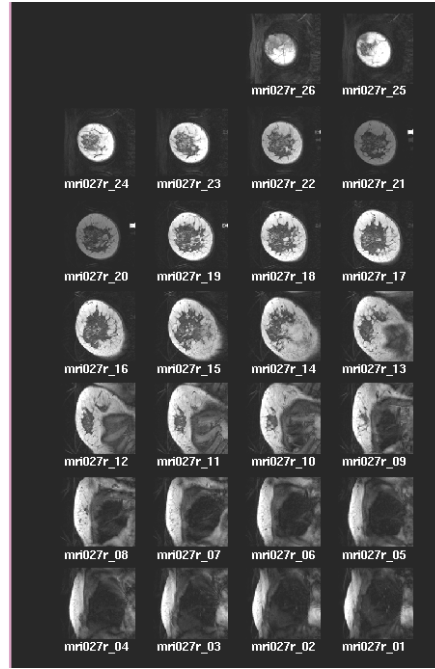


Fig. 1. MR images of the right breast of a patient. The two-view mammogram of the same breast is shown in Fig. 2.

We have developed a graphical user interface that displays the MR series and the corresponding mammogram of each breast. The interface allows the user to perform a combination of manual and automatic operations to segment the MR images. Each MR slice is first thresholded to separate the breast from the surrounding region. For slices close to the chest wall where no clear boundary can be seen, the boundary is manually drawn and evaluated by radiologists. Background correction [4] using the voxel values around the segmented breast region is employed to correct for the non-

uniformity across the breast area due to the breast coil. The histogram of the voxel values in the breast region is then formed and interactive thresholding is used to segment the fibroglandular tissue from the fatty tissue. A morphological erosion operation along the breast boundary then excludes the skin voxels from the calculation of the fibroglandular tissue area in each slice. Finally, an integration of the fibroglandular voxels in all slices relative to the breast volume provides the percent volumetric fibroglandular tissue in the breast.

We have previously developed an automated image analysis tool (Mammographic Density ESTimator) to assist radiologists in estimating mammographic breast density [5]. MDEST performs dynamic range compression, breast boundary tracking, pectoral muscle segmentation for the MLO view, automatic thresholding based on gray level histogram analysis, and calculates the percent dense area on a mammogram. We found that the correlation between the computer-estimated percent dense area and radiologists' manual segmentation was 0.94 and 0.91, respectively, for CC and MLO views, with a mean bias of less than 2%. An example of a mammogram segmented by MDEST is shown in Fig. 2.

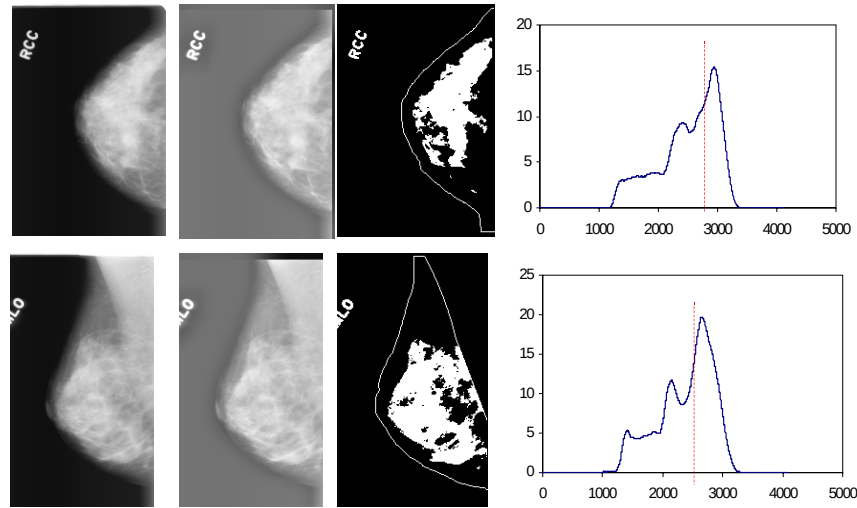


Fig. 2. Automated breast density segmentation from mammograms. Upper row: CC view. Lower row: MLO view.

3. Results

Scatter plots of the percent volumetric fibroglandular tissue versus the percent dense area on mammograms are shown in Fig. 3(a) and 3(b) for the CC- and MLO-view mammograms, respectively. The correlation of percent dense area of the CC and MLO views with the percent volumetric fibroglandular tissue on MR images was found to be 0.93 and 0.91, respectively, with a mean bias of 4.4%.

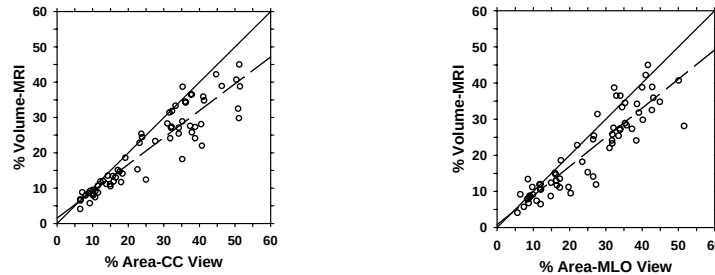


Fig. 3. Correlation of % volumetric fibroglandular tissue on MR images with % dense area on mammograms for 37 patients. The left and right breasts are plotted as separate data points on each graph. The dash lines are linear least squares fits to the data points.

4. Conclusion

Mammographic density is highly correlated with the volumetric fibroglandular tissue in the breast, indicating its usefulness as a surrogate for breast density estimation. The computerized image analysis tool, MDEST, is useful for estimation of mammographic density. The automated analysis is expected to contribute to the understanding of the relationship of mammographic density to breast cancer risk, detection, and prognosis, and to the prevention and treatment of breast cancer.

Acknowledgements

This work is supported by U. S. Army Medical Research and Materiel Command grants DAMD 17-01-1-0326 and DAMD 17-99-1-9294. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred.

References

1. J. Brisson, R. Verreault, A. S. Morrison, D. Tennina and F. Meyer, "Diet, mammographic features of breast tissue, and breast cancer risk," *Am. J. Epidemiology* 130, 14-24 (1989).
2. A. F. Saftlas, R. N. Hoover, L. A. Brinton, M. Szklo, D. R. Olson, M. Salane and J. N. Wolfe, "Mammographic densities and risk of breast cancer," *Cancer* 67, 2833-2838 (1991).
3. N. F. Boyd, G. A. Lockwood, J. W. Byng, D. L. Tritchler and M. J. Yaffe, "Mammographic densities and breast cancer risk," *Cancer Epidemiology Biomarkers & Prevention* 7, 1133-1144 (1998).
4. B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler and M. M. Goodsitt, "Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images," *IEEE Trans Med Imag* 15, 598-610 (1996).
5. C. Zhou, H. P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, B. Sahiner and L. M. Hadjiiski, "Computerized image analysis: Estimation of breast density on mammograms," *Med Phys* 28, 1056-1069 (2001).

Performance evaluation of an automated breast density estimation system for digital mammograms and digitized film mammograms

Chuan Zhou, Heang-Ping Chan, Jun Wei, Mark A. Helvie, Marilyn A. Roubidoux, Chintana Paramagul, Alexis V. Nees, Lubomir M. Hadjiiski, Berkman Sahiner

Department of Radiology, University of Michigan, Ann Arbor, MI 48109 chuan@umich.edu

Abstract. Studies have demonstrated a strong correlation between mammographic breast density and breast cancer risk. We have previously developed a computerized system, mammographic density estimator (MDEST), to estimate breast density automatically on digitized film mammograms (DFM). In this study, we evaluated the performance of the MDEST system on full field digital mammograms (FFDM) and DFMs. The input to the system is a preprocessed dynamic range compressed image. The breast region is first segmented by breast boundary detection. The pectoral muscle is trimmed if it is an MLO view. A rule-based classifier is then used to classify the breast image into one of four classes according to the characteristics of its gray level histogram. The dense area from the breast region is subsequently segmented by automatic gray level thresholding. The breast density is estimated as the percentage of the segmented dense area relative to the breast area. In this study, two-view FFDM and the corresponding DFM from 99 patients with 202 images in each set were used. The dense area on each mammogram was segmented by 4 radiologists using interactive thresholding and their average was used as the “gold standard”. The MDEST system was directly applied to the FFDM and DFM data without any re-training except that the preprocessing filter was modified for FFDMs. We found that the correlation between the estimated percent dense area and the gold standard was 0.850 and 0.873 on FFDM, and 0.885 and 0.824 on DFM, for CC and MLO views, respectively. The results demonstrated the feasibility of estimating breast density automatically on FFDM and DFM using the same MDEST system.

1. Introduction

Studies have demonstrated a strong correlation between breast density on mammograms and breast cancer risk (Saftlas and Szklo 1987; Brisson et al. 1989; Saftlas et al. 1991; Oza and Boyd 1993; Boyd et al. 1998; Yaffe et al. 1998). The relative risk is estimated to be about 4 to 6 times higher for women whose mammograms have parenchymal densities over 60% of the breast area, as compared to women with less than 5% of parenchymal densities. The strong correlation between breast density and breast cancer risk has prompted researchers to use mammographic density as an indicator for monitoring the effects of preventive or interventional treatment of breast cancer.

Because of the subjective nature of visual analysis, qualitative estimation may vary from radiologist to radiologist. A computerized method for measuring mammographic density would be useful as a supplement to the radiologist’s assessment. We have previously

developed a computerized system, mammographic density estimator (MDEST) (Zhou et al. 2001), to estimate breast density automatically on digitized film mammograms (DFM). The MDEST system performs dynamic range compression, breast boundary tracking, pectoral muscle trimming for MLO view, gray level thresholding based on histogram analysis, and calculation of the percent dense area on the mammogram. In the previous study, 260 digitized 4-view mammograms of 65 patients were used. The gold standard of percent dense area of the breast region for each mammogram was obtained by averaging five radiologists' manually segmented percent dense area. We found that the correlation between the computer-estimated percent dense area and radiologists' manual segmentation was 0.94 and 0.91, with RMS errors at 6.1% and 7.2%, respectively, for CC and MLO views.

In this study, we investigate the feasibility of computerized mammographic density estimation on FFDMs and DFMs using the same image segmentation system. The MDEST system was directly applied to the FFDM and the corresponding DFM without any re-training except that the preprocessing filter was modified for FFDMs. The performance was evaluated by analyzing the correlation between the computer-estimated mammographic density and the gold standard obtained by radiologists' interactive thresholding.

2. Materials and Methods

The data sets consisting of FFDM and the corresponding DFM of 99 patients with 202 images in each set were used. Each case contains the craniocaudal (CC) view and the mediolateral oblique (MLO) view. The FFDM was acquired with a GE Senographe 2000D system and the raw GE FFDM was processed by a Laplacian pyramid multi-resolution preprocessing method (Wei et al. 2004). The preprocessed image was downsized to a pixel size of $800\ \mu\text{m} \times 800\ \mu\text{m}$ image and 4096 gray levels. The DFM was acquired with mammography systems approved by the Mammography Quality Standards Act (MQSA) and was digitized with a LUMISYS 85 laser film scanner with a pixel size of $50\ \mu\text{m} \times 50\ \mu\text{m}$ and 4096 gray levels. The digitized mammogram was also downsized to a $800\ \mu\text{m} \times 800\ \mu\text{m}$ image using a 16×16 box filter.

Our previously developed computerized system MDEST was applied to the FFDMs and DFMs to estimate the mammographic density without any re-training. The density estimation was performed in three stages: breast region segmentation, image enhancement, and gray level thresholding based on histogram analysis. First, the breast region was segmented from the surrounding background by an automated breast boundary tracking algorithm for DFM. For FFDM, thresholding was used to separate the breast region from the background. Since our current pectoral muscle trimming program is not 100% accurate, the pectoral muscle was manually trimmed on the MLO view images for both DFM and FFDM in this study in order to separate the errors due to breast density segmentation from those due to pectoral muscle trimming. Second, an adaptive dynamic range compression technique was applied to enhance the DFM image. For FFDM image, a Laplacian pyramid multi-resolution preprocessing method (Wei et al. 2004) was used for image enhancement. At the third stage, for both FFDM and DFM, rule-based classification was used to classify the breast image into one of four classes according to the characteristic features of its gray level histogram (Zhou et al. 2001). For each image in the classified classes, a gray level threshold was determined adaptively to segment the dense area from the breast region. The breast density was estimated as the percentage of the segmented dense area relative to the breast area. As an example, typical mammograms in the four classes with the corresponding enhanced images, histograms, selected thresholds and the segmented image are shown in Figure 1. To evaluate the performance of MDEST, the computer segmentation results were compared to those by

manual segmentation with interactive thresholding by four MQSA radiologists. The “gold standard” of percent dense area for each mammogram was obtained by averaging the manually segmented percent dense areas of the four radiologists.

3. Results

Figures 2(a)-(d) show the comparison of the percent dense area between the estimation by the MDEST system and the gold standard on FFDM and DFM for CC- and MLO-view mammograms, respectively. Table 1 summarizes the comparison between the MDEST performance and the gold standard for FFDM and DFM, respectively. The correlation between the computer-estimated percent dense area and the gold standard is 0.850 and 0.873

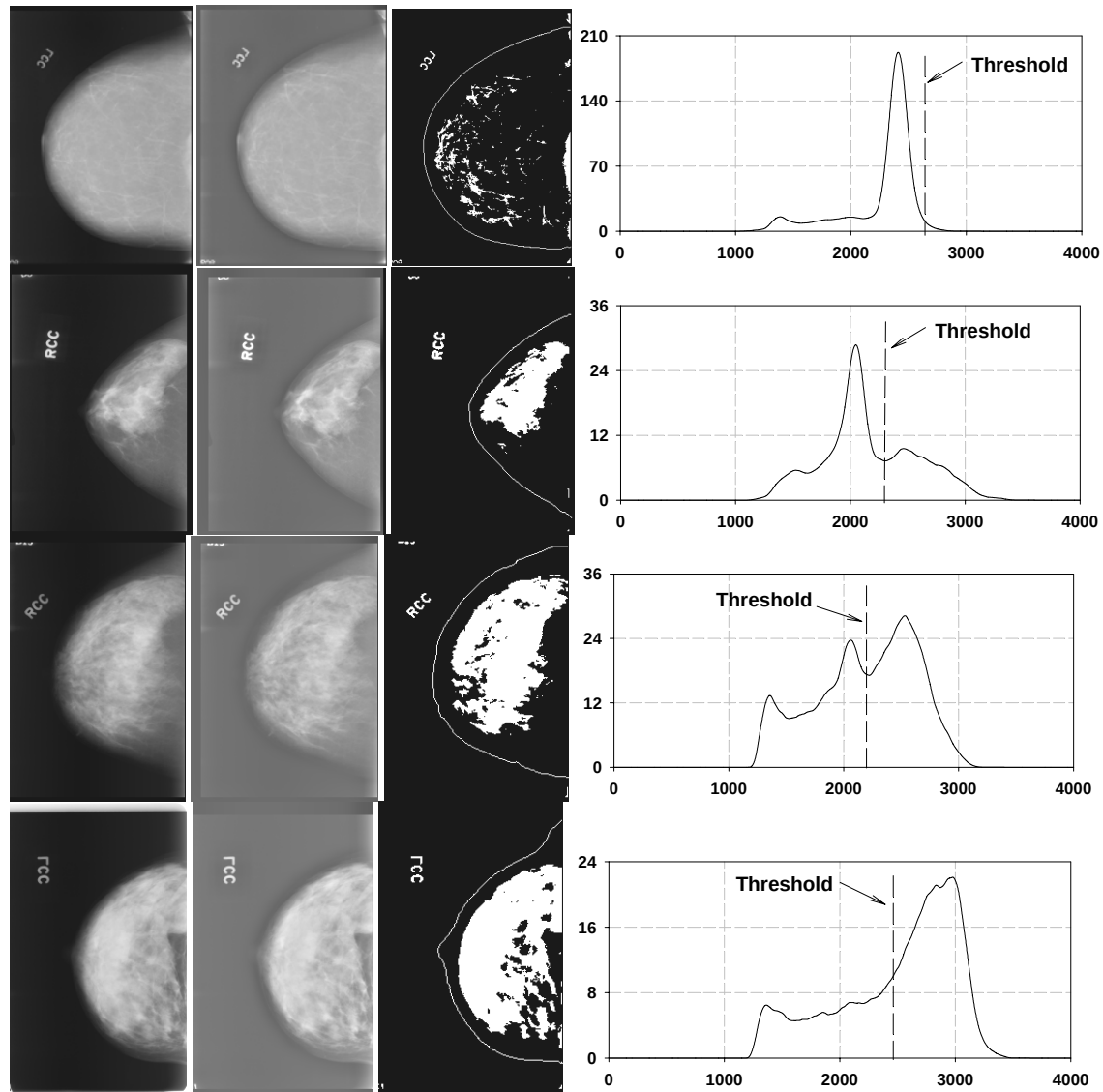


Figure 1. Typical mammograms in the four density classes and the corresponding enhanced and segmented images, histograms and thresholds. The columns from left to right correspond to the original image, enhanced image, segmented image and the histogram. Rows from top to bottom correspond to class one to four.

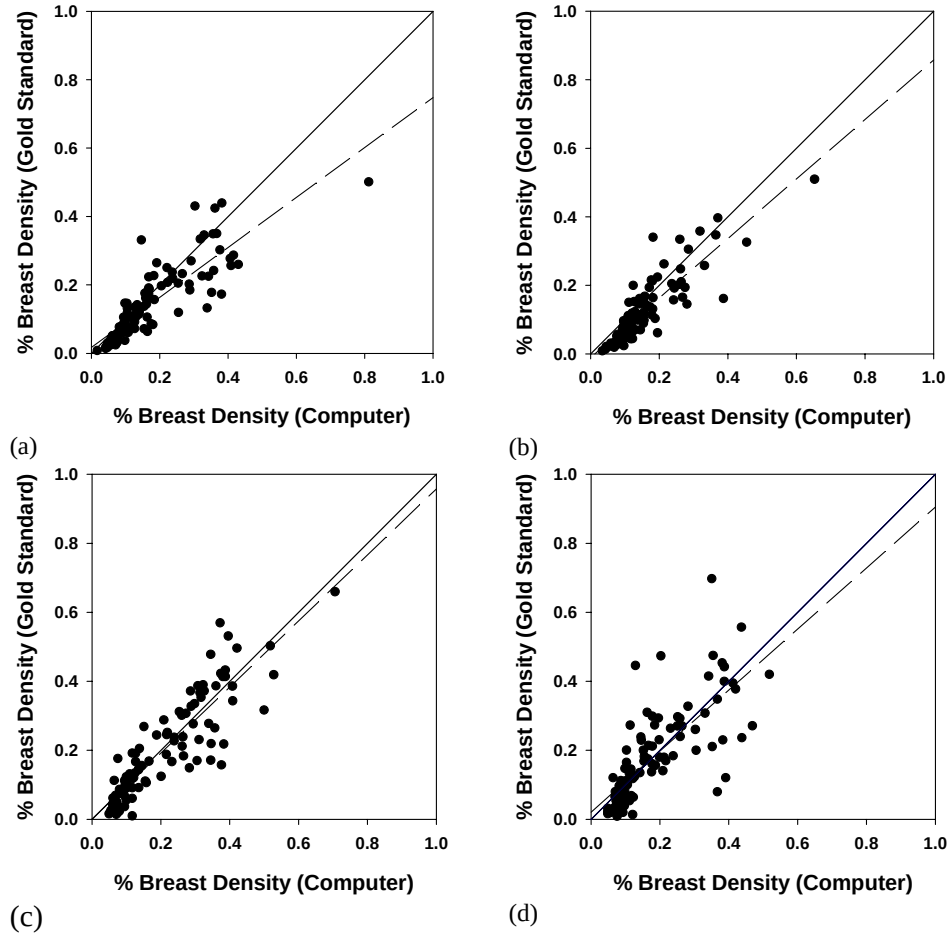


Figure 2. Comparison of the percent dense area between the estimation by the MDEST system and the gold standard. The dashed line represents the linear regression of the data on the plot. (a) FFDM CC view, (b) FFDM MLO view, (c) DFM CC view, (d) DFM MLO view.

Table I. The correlation and RMS difference between the estimated percent dense area by the MDEST system and the gold standard (average of four MQSA radiologists).

Image subsets	FFDM		DFM	
	Correlation	RMS Error	Correlation	RMS Error
CC view	0.850	7.26%	0.885	6.87%
MLO view	0.873	5.70%	0.824	8.16%
All images	0.859	6.52%	0.855	7.54%

on FFDM, and 0.885 and 0.824 on DFM, for CC and MLO views, respectively. For all of the images combining CC- and MLO-views, the correlation is 0.859 and 0.855 on FFDM and DFM, respectively. The RMS difference in the percent dense area between the MDEST estimation and the gold standard is 7.26%, 5.70% and 6.52% on FFDM, and 6.87%, 8.16% and 7.54% on DFM for CC-view alone, MLO-view alone, and combined CC and MLO-views, respectively.

4. Conclusion

Our preliminary study demonstrated that the estimation of mammographic density could be performed efficiently by the automated image analysis tool. The computer-estimated percent dense area had a high correlation with the gold standard obtained from averaging four MQSA radiologists' manual segmentation. The results also demonstrated the feasibility of estimating breast density automatically on FFDM and DFM using the same MDEST system. Further study will be conducted to improve the breast density segmentation accuracy.

Acknowledgements

This work is supported by U. S. Army Medical Research and Materiel Command grant DAMD 17-01-1-0326 and USPHS grant CAD95153. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred.

References

- Boyd, N. F., G. A. Lockwood, J. W. Byng, D. L. Tritchler and M. J. Yaffe. 1998. Mammographic densities and breast cancer risk. *Cancer Epidemiology Biomarkers & Prevention* 7:1133-1144.
- Brisson, J., R. Verreault, A. S. Morrison, D. Tennina and F. Meyer. 1989. Diet, mammographic features of breast tissue, and breast cancer risk. *Am. J. Epidemiology* 130:14-24.
- Oza, A. M. and N. F. Boyd. 1993. Mammographic parenchymal patterns: A marker of breast cancer risk. *Epidemiologic Reviews* 15:196-208.
- Saftlas, A. F., R. N. Hoover, L. A. Brinton, M. Szklo, D. R. Olson, M. Salane and J. N. Wolfe. 1991. Mammographic densities and risk of breast cancer. *Cancer* 67:2833-2838.
- Saftlas, A. F. and M. Szklo. 1987. Mammographic parenchymal patterns and breast cancer risk. *Epidemiologic Reviews* 9:146-174.
- Wei, J., B. Sahiner, L. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, C. Zhou and Z. Ge. 2004. Computer aided detection of breast masses on full-field digital mammograms: false positive reduction using gradient field analysis. *Proc. SPIE Medical Imaging 2004: Image Processing* 5370:992-998.
- Yaffe, M. J., N. F. Boyd, J. W. Byng, R. A. Jong, R. Fishell, G. A. Lockwood, L. E. Little and D. L. Tritchler. 1998. Breast cancer risk and measured mammographic density. *European J. of Cancer Prevention* 7, Suppl. 1:S47-S55.
- Zhou, C., H. P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, B. Sahiner and L. M. Hadjiiski. 2001. Computerized image analysis: Estimation of breast density on mammograms. *Medical Physics* 28:1056-1069.

Computerized pectoral muscle identification on MLO-view mammograms for CAD applications

Chuan Zhou, Lubomir M. Hadjiiski, Chintana Paramagul, Berkman Sahiner, Heang-Ping Chan, Jun Wei
Department of Radiology, The University of Michigan, Ann Arbor, MI 48109-0904

ABSTRACT

Automatic identification of the pectoral muscle on MLO view is an essential step for computerized analysis of mammograms. It can reduce the bias of mammographic density estimation, will enable region-specific processing in lesion detection programs, and also may be used as a reference in image registration algorithms. We are developing a computerized method for the identification of pectoral muscle on mammograms. The upper portion of the pectoral edges was first detected to estimate the direction of the pectoral muscle boundary. A gradient-based directional (GD) filter was used to enhance the linear texture structures, and then a gradient-based texture analysis was designed to extract a texture orientation image that represented the dominant texture orientation at each pixel. The texture orientation image was enhanced by a second GD filter. An edge flow propagation method was developed to extract edges around the pectoral boundary using geometric features and anatomic constraints. The pectoral boundary was finally generated by a second-order curve fitting. 118 MLO view mammograms were used in this study. The pectoral muscle boundary identified on each image by an experienced radiologist was used as the gold standard. The accuracy of pectoral boundary detection was evaluated by two performance metrics. One is the overlap percentage between the computer-identified area and the gold standard, and the other is the root-mean-square (RMS) distance between the computer and manually identified pectoral boundary. For 118 MLO view mammograms, 99.2% (117/118) of the pectoral muscles could be identified. The average of the overlap percentage is 94.8% with a standard deviation of 20.9%, and the average of the RMS distance is 4.3 mm with a standard deviation of 5.9 mm. These results indicate that the pectoral muscle on mammograms can be detected accurately by our automated method.

Keywords: Computer-aided detection, Pectoral muscle trimming, Breast density estimation, Directional gradient filter

1. INTRODUCTION

Breast cancer is one of the leading causes of cancer mortality among women^{1, 2}. At present, the most successful method for the early detection of breast cancer is screening mammography³. It has been demonstrated that an effective computer-aided diagnosis (CAD) system can provide a second opinion to the radiologists and improve the accuracy of detection and characterization of mammographic abnormalities, which, in turn, may reduce unnecessary biopsies. Studies have shown that there is a strong positive correlation between breast parenchymal density on mammograms and breast cancer risk.^{1, 4-6} The relative risk is estimated to be about 4-6 times higher for women whose mammograms have parenchymal densities over 60% of the breast area, as compared to women with less than 5% of parenchymal densities. Mammograms are analyzed visually by radiologists, the qualitative response may vary from radiologist to radiologist due to the subjective nature of visual analysis. We have previously developed a computerized system, mammographic density estimator (MDEST), to estimate breast density automatically on digitized film mammograms.⁷ For each mammogram, the breast region was first segmented by breast boundary detection and, for the mediolateral oblique (MLO) view, with additional pectoral muscle trimming. A gray level threshold was then automatically determined to segment the dense tissue from the breast region. The breast density was estimated as the percentage of the segmented dense area relative to the breast area. Our preliminary study indicated that the computer-estimated mammographic breast density correlated closely with the "reference standard" obtained by averaging five experienced radiologists' manual segmentations and the average bias was much less than that of the radiologists' visual estimation.

Automatic identification of the pectoral muscle is an essential step for computerized analysis of mammograms. Accurate segmentation of the pectoral muscle on MLO-view mammograms can reduce the bias of mammographic

density estimation and improve the performance of our MDEST method. It will enable region-specific processing in lesion detection programs to reduce false negatives. False positives can be reduced if the detected objects in the pectoral muscle area can be selectively suppressed. The identification of the pectoral muscle may also be used as a reference in image registration algorithm for multiple-view analysis of mammograms.

In our preliminary study⁷, the pectoral muscle was trimmed using a gradient-based pectoral edge detection method: the initial edge in the pectoral region was first found as the maximum gradient point by a line-by-line gradient analysis from the chest wall to the breast boundary. An edge validation process was then performed to remove the false pectoral muscle edges using a line fitting method, and a coarse direction of the pectoral edges was estimated from the validated edges. The remaining pectoral edges were extrapolated along the estimated pectoral direction. Finally, a second order curve was fitted to the detected pectoral edges to generate the pectoral boundary. Using the above method, 74.6% of the pectoral muscles were determined by visual judgment to be correctly identified in this preliminary study.

The purpose of this study is to improve the performance of our previously developed pectoral muscle segmentation method. Accurate identification of the pectoral muscle on mammograms is challenging, especially for the improperly positioned MLO-view images and the images containing dense glandular tissues overlapping with the pectoral muscle region. In this work, we developed a two-stage gradient-based texture analysis method to detect the pectoral boundary. In the first stage, linear texture structures were enhanced and the directional gradients were computed using a directional filter. In the second stage, a texture orientation image was derived as the dominant texture orientation at each pixel. A diffusion filter was used to estimate the global direction of the pectoral boundary. An edge flow propagation method was developed to extract the pectoral edges with the guidance of the estimated global direction.

2. MATERIALS AND METHODS

2.1 Materials

In this study, 118 MLO-view mammograms from 103 patients were randomly selected from the patient files in the Radiology Department at the University of Michigan. Data collection was approved by the Institutional Review Board and individual patient informed consent was waived. The mammograms were acquired with Mammography Quality Standards Act (MQSA) approved GE DMR (Milwaukee, Wisconsin) mammography units using Kodak MR2000 screen/film systems. All films were digitized with a LUMISYS 85 laser film scanner with a pixel size of $50\ \mu\text{m} \times 50\ \mu\text{m}$ and 4096 gray levels. The resolution of the mammograms was reduced to $800\ \mu\text{m} \times 800\ \mu\text{m}$ for segmentation of the pectoral muscle.

2.2 Pectoral muscle identification

Figure 1 summarizes the automatic pectoral muscle identification scheme. The interference due to overlapping of the glandular tissue on the pectoral muscle region is first reduced by smoothing the mammogram using an edge preserving anisotropic diffusion filter⁸. Because less glandular tissue appears at the upper region of the pectoral muscle, the upper portion of the pectoral boundary usually remains sharp after smoothing and can be detected robustly by searching the maximum horizontal gradients on the diffused image. The extrapolation of the detected upper pectoral boundary provides a coarse global direction of the pectoral boundary. To refine the entire pectoral boundary, a gradient-based directional (GD) filter was first employed to enhance the linear texture structures on the mammogram. The orientation of the digitized image could be automatically determined by the curvature of the breast boundary. For example, if the image was positioned such that the chest wall was on the right side, it could be assumed that the pectoral boundary is at a direction approximately from the top-left to the bottom-right with less than 45 degree deviation. Therefore, in our study, the kernel of the GD filter was designed as a step function with 45 degree orientation. After the pectoral edge was enhanced by the GD filter, a gradient-based texture analysis⁹ was used to compute an orientation image which represented the dominant texture orientation at each pixel. The orientation image was smoothed using an edge preserving mean shift algorithm¹⁰ that iteratively shifted each pixel to the average of the pixels in its neighborhood. The texture patterns with dominant texture orientations directing from the top-left to the bottom-right, which were more likely to be the pectoral edges, were enhanced by applying a second GD filter to the smoothed orientation image. Candidate edges of the pectoral muscle were detected on the enhanced orientation image using a ridge-tracking

algorithm. The ridges were tracked by searching for the local maximum along the coarse global direction estimated, as described above, by the upper pectoral boundary on the anisotropic diffused image. With the guidance of the estimated global direction of the pectoral boundary and the anatomical constraints, an edge flow propagation algorithm was then used to extract the boundary points of the pectoral muscle by pruning the edges that are less likely to lie on the pectoral boundary. A second order curve fitting was finally used to generate the pectoral muscle boundary. Figure 2 shows examples of the intermediate images of pectoral boundary enhancement and edge tracking corresponding to the various stages shown in the flowchart in Figure 1.

3. RESULTS

An experienced MQSA-radiologist used a graphical user interface to manually draw the pectoral muscle boundary on each MLO-view mammogram, which was then used as the gold standard for the evaluation of the performance of our pectoral muscle detection program.

For each MLO view mammogram, the accuracy of pectoral boundary detection was evaluated by two performance metrics: the percentage of overlap, defined as the ratio of the overlap area between the computer detected pectoral muscle area and the gold standard relative to the gold standard, and the root-mean-square (RMS) distance obtained by calculating the shortest distance point by point between the computer-identified pectoral boundary and the manually marked pectoral boundary. For the data set of 118 MLO view mammograms, 99.2% (117/118) of the pectoral muscles could be identified, the average of the percent overlap area is 94.8% with a standard deviation of 20.9%, the average of the RMS distance is 4.3 mm with a standard deviation of 5.9 mm.

Figure 3 shows some examples of pectoral boundary identification on mammograms. The computer identified pectoral boundaries were shown in white lines and the dark lines show the radiologist's hand drawn boundaries. Figure 3 (a)-(b) show the pectoral boundary can be identified accurately on mammograms with weak pectoral edges (figure 3(a)) and a large area of dense tissues overlapping on the pectoral muscle area (shown in figure 3(b)). Figure 3(c)-(d) show two examples of less accurate pectoral boundaries detected by the computer. Figure 3(e) shows the only case in this data set that the computer failed to detect the boundary.

4. CONCLUSION

The newly developed gradient-based directional filter and the dominant texture orientation estimation method can enhance the pectoral boundary regions. The edge flow propagation method can accurately extract pectoral edges to generate the pectoral boundary. Automatic pectoral muscle identification will provide the foundation for many mammographic image analysis tasks in CAD applications.

5. ACKNOWLEDGMENT

This work is supported by USPHS grant CA 95153 and U.S Army Medical Research and Materiel Command grant DAMD17-02-1-0214. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred.

REFERENCES

1. S. H. Landis, T. Murray, S. Bolden and P. A. Wingo, "Cancer statistics, 1998," *CA Cancer J Clin* **48**, 6-29, 1998.
2. C. Byrne, C. R. Smart, C. Cherk and W. H. Hartmann, "Survival advantage differences by age: Evaluation of the extended follow-up of the breast cancer detection demonstration project," *Cancer* **74**, 301-310, 1994.
3. H. C. Zuckerman, "The role of mammography in the diagnosis of breast cancer," *Breast cancer, diagnosis and treatment*, I. M. Ariel and J. B. Cleary, Eds., 152-172, McGraw-Hill, 1987.

4. J. N. Wolfe, "Breast patterns as an index of risk for developing breast cancer," *Am. J. Roentgenol.* **126**, 1130-1139, 1976.
5. J. N. Wolfe, A. F. Saftlas and M. Salane, "Evaluation of mammographic densities: A case-control study," *AJR* **148**, 1087-1092, 1987.
6. A. F. Saftlas, R. N. Hoover, L. A. Brinton, M. Szklo, D. R. Olson, M. Salane and J. N. Wolfe, "Mammographic densities and risk of breast cancer," *Cancer* **67**, 2833-2838, 1991.
7. C. Zhou, H. P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, B. Sahiner and L. M. Hadjiiski, "Computerized image analysis: Estimation of breast density on mammograms," *Medical Physics* **28**, 1056-1069, 2001.
8. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. on Pattern analysis and machine intelligence* **12**, 629-639, 1990.
9. A. R. Rao and B.G.Schunck, "Computing oriented texture fields," *CVGIP: Graphical models image processing* **53**, 157-185, 1991.
10. D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. on Pattern analysis and machine intelligence* **24**, 603-619, 2003.

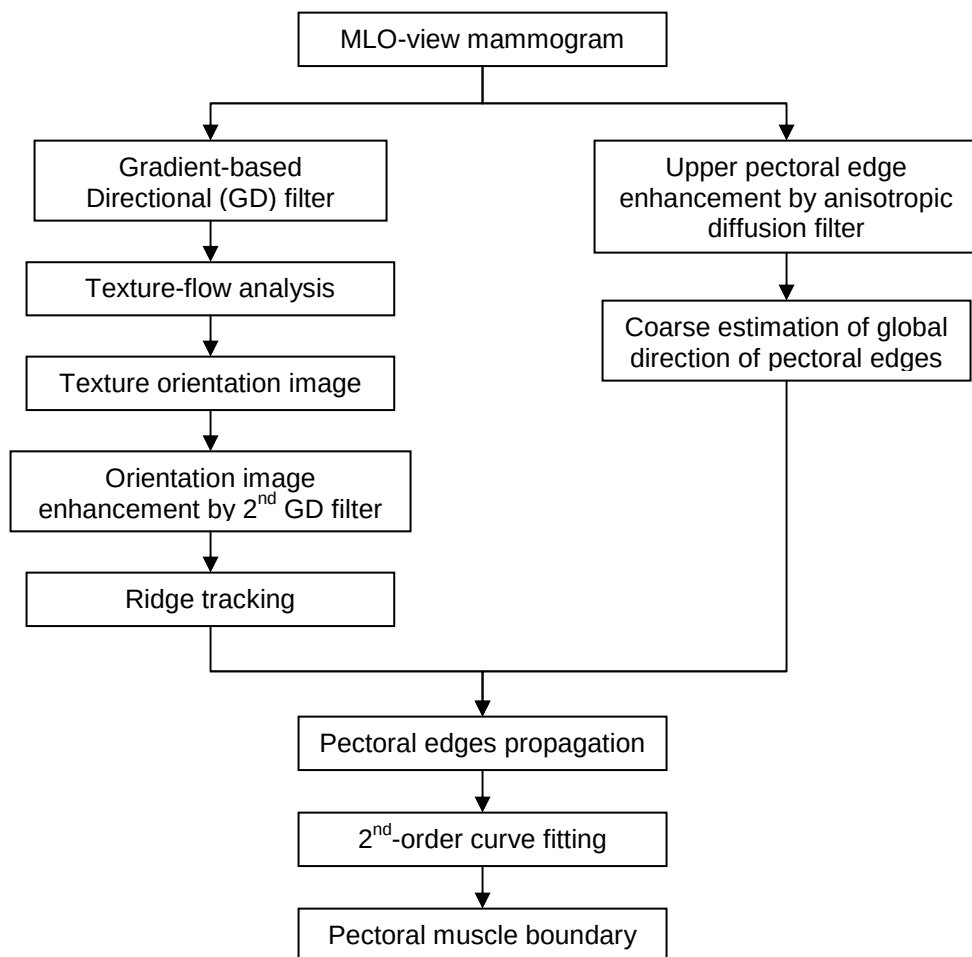


Figure 1. Automated pectoral muscle detection scheme

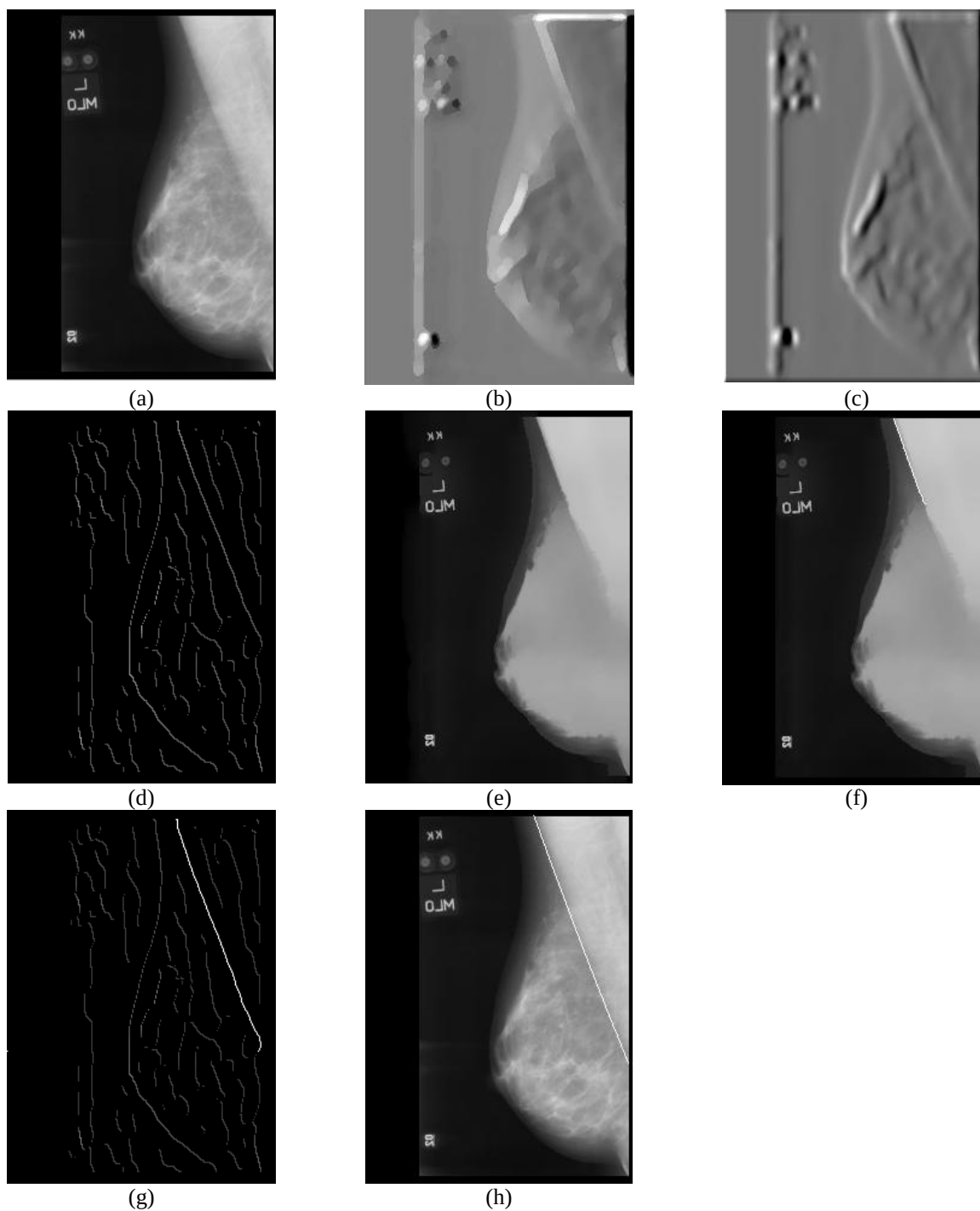


Figure 2. Example of boundary enhancement and segmentation of pectoral muscle. (a) original image; (b) texture orientation image after first GD filter and texture-flow analysis; (c) ridge image enhanced by the 2nd GD filter; (d) tracked ridges; (e) smoothed image using anisotropic diffusion filter; (f) initial pectoral edges detected from the smoothed image in (e) for the estimation of the coarse direction of the pectoral boundary; (g) propagated pectoral edges on the ridge image (c) with the guidance of the coarse direction estimated from the smoothed image shown in (f); (h) the final identified pectoral boundary after 2nd order curve fitting.

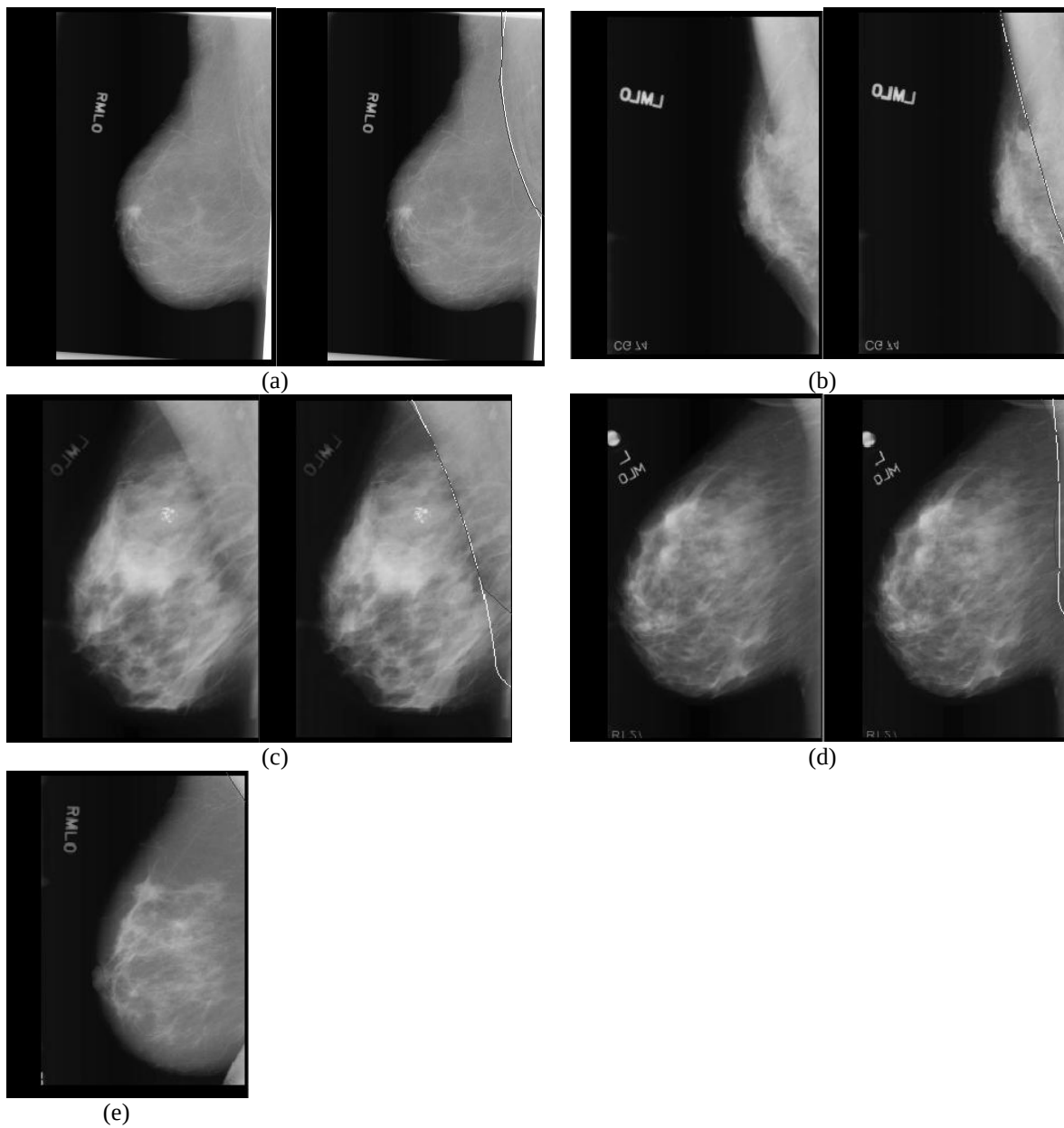


Figure 3. Examples of pectoral boundary segmentation on mammograms. (a)-(b): accurate identification of pectoral boundary; (c)-(d): less accurate identification of pectoral boundary; (e) the only mammogram in our data set that the computer failed to identify the pectoral muscle due to the small portion of the pectoral muscle area within the breast region.

Breast Density Estimation: Correlation of Mammographic Density and MR Volumetric Density

Chan HP, Hadjiiski LM, Roubidoux MA, Helvie MA, Paquerault S, Sahiner B, Chenevert T, Goodsitt MM

Studies have demonstrated a strong correlation between mammographic breast density and breast cancer risk. Mammographic breast density may therefore be used as a surrogate marker for monitoring the response to treatment in studies of breast cancer prevention or intervention methods. In this study, we evaluated the accuracy of using mammograms for estimating breast density by analyzing the correlation between the mammographic areal density and the glandular tissue volume as estimated from MRI.

A data set of fifty patients who had corresponding MR images and mammograms was collected. The coronal MR images provided full volume coverage of both breasts with contiguous image sections. Heavily T1-weighted acquisition was used which produced images with well-separated dense parenchyma and adipose tissue. The glandular tissue regions in the MR sections were outlined interactively by experienced radiologists with a graphical user interface and the percent glandular tissue volume calculated. Mammographic density was estimated with an automated image analysis program and compared to that provided by manual segmentation. The computer algorithm included dynamic range compression, breast boundary tracking, and automatic thresholding based on analysis of the gray level histogram. The dense tissue regions on the mammogram were then segmented and the percent dense area estimated.

The correlation between the computer-estimated percent dense area and radiologists' manual segmentation of mammographic density was found to range from 0.91 to 0.94 in an independent set of mammograms. The analysis of the relationship between mammographic percent dense area and MR percent glandular tissue volume is underway. We will discuss the correlation of the breast fibroglandular-to-adipose ratios estimated from these two approaches.

AUTOMATED ANALYSIS OF MAMMOGRAPHIC BREAST DENSITY FOR BREAST CANCER RISK ESTIMATION

**Heang-Ping Chan, Mark Helvie, Jun Wei, Lubomir Hadjiiski,
Chuan Zhou, Mitchell Goodsitt, Berkman Sahiner, Marilyn Roubidoux**

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

E-mail: chanhp@umich.edu

Mammographic breast density, an indicator of the proportion of fibroglandular vs. fatty tissue in the breast, has been found to have strong correlation with breast cancer risk. Mammographic breast density has therefore been used for monitoring the response in studies of preventive or interventional treatment of breast cancer. Breast density changes during the course of treatment are often estimated visually on mammograms by radiologists; which involves large inter- and intraobserver variations. The goal of this project is to develop an automated image analysis method that can provide a more consistent and reproducible estimate of the percent dense breast area on a mammogram.

An automated computer program has been developed that performs breast density analysis using the following steps: detection of the breast boundary, reduction of the image dynamic range, analysis and classification of the shape of the gray level histogram, adaptive gray level thresholding, and estimation of the percent dense tissue area relative to the breast area. The performance of the algorithm was evaluated by comparing the computer segmentation results to manual segmentation with interactive thresholding by five radiologists.

To further investigate the relationship between the mammographic breast density and the amount of fibroglandular tissue in the breast, the image analysis program was applied to the mammograms of 37 patients who had corresponding magnetic resonance (MR) images of the breasts. The fibroglandular tissue regions in the MR slices were segmented interactively with a user interface, and the percentage of fibroglandular tissue volume in the breast estimated. The correlation between the percent dense area estimated from mammograms and the percent volumetric fibroglandular tissue estimated from MR images was studied.

We found that the correlation between the computer-estimated percent dense area and the average of the five radiologists' manual segmentation was 0.94 and 0.91, respectively, for CC and MLO views, with a mean bias of less than 2%. The percent breast dense area of the CC and MLO views has a correlation of 0.92 and 0.91, respectively, with the percent volumetric fibroglandular tissue on MR images. Mammographic density is therefore highly correlated with the volumetric fibroglandular tissue in the breast, indicating its usefulness as a surrogate for breast density estimation.

The computerized image analysis tool is useful for breast density estimation on mammograms. The automated analysis is expected to contribute to the understanding of the relationship of mammographic density to breast cancer risk, detection, and prognosis, and to the prevention and treatment of breast cancer.

Breast Density Estimation on Mammograms and MR Images: A Tool for Assessment of Breast Cancer Risk

Jun Wei, Heang-Ping Chan, Mark A. Helvie, Lubomir M. Hadjiiski,
Berkman Sahiner, Marilyn A. Roubidoux, Chuan Zhou, Sophie Paquerault,
Thomas Chenevert, Mitchell M. Goodsitt

PURPOSE: Previous studies have found that mammographic breast density is highly correlated with breast cancer risk. We have developed a computerized image analysis tool, Mammography Density ESTimator (MDEST), to estimate the percent dense area on mammograms. In this study, we analyzed the correlation between mammographic percent dense area and percent volumetric fibroglandular tissue on MR images.

METHOD AND MATERIALS: For the estimation of mammographic breast density, MDEST performs the following procedures: detection of the breast boundary, reduction of the image dynamic range, analysis and classification of the gray level histogram, adaptive gray level thresholding, and estimation of the percent dense tissue area relative to the breast area. The performance of MDEST was validated by comparing its segmentation to that with manual interactive thresholding by five radiologists in 260 mammograms. For the estimation of the percent volume of fibroglandular tissue in breast MR images, a semi-automatic method has been developed to segment the fibroglandular tissue from each slice. First, the breast boundary is detected automatically. A deformable model and manual modification are used to correct for incorrectly detected boundaries that usually occur in slices near the chest wall where the breast boundary is not well-defined. Because of the nonuniformity of the breast coil, the signal intensity in the breast region is not uniform across the field of view. A background correction technique that estimates the low frequency background from the gray levels along the breast boundary is developed to reduce the nonuniformity. Finally, manual interactive thresholding of the gray level histogram in the breast region is used to separate the fibroglandular from the fatty region. The tissue volume is calculated by integration over all slices containing the breast. A data set of 54 cases having MR images and corresponding 4-view mammograms was used in this study. The MR images were coronal 3D SPGR T1-weighted pre-contrast images.

RESULTS: The percent volume of fibroglandular tissue had a correlation of 0.92 and 0.91 with the percent dense area obtained on CC-view and MLO-view mammograms, respectively. The percent mammographic dense area slightly overestimates the percent volume with a mean bias of 3%.

CONCLUSIONS: Mammographic density is highly correlated with the volumetric fibroglandular tissue in the breast, indicating its usefulness as a surrogate for breast density estimation and thus for monitoring breast cancer risk.

Comparison of mammographic density estimated on digital mammograms and screen-film mammograms

Heang-Ping Chan, Jun Wei, Chuan Zhou, Mark A. Helvie, Marilyn Roubidoux, Janet Bailey,
Lubomir Hadjiiski, Berkman Sahiner

PURPOSE: To compare breast density estimated on pairs of digital mammogram (DM) and screen-film mammogram (SFM) obtained from the same patients.

METHODS AND MATERIALS: We are comparing image information on DMs and SFMs for radiologist's interpretation and computerized image analysis. One hundred forty-five pairs of DM and SFM (76 CC views and 69 MLO views) were collected with IRB approval from 68 patients. The time interval between the DM and SFM ranged from 0 to 118 days (median=21 days). The SFMs were acquired with GE DMR systems and the DMs were acquired with a GE Senographe 2000D system. Both the DMs and the SFMs were acquired with automated exposure techniques that selected the appropriate target, filter, and kVp. The SFMs were digitized with a laser film scanner. The breast boundaries on the DMs and SFMs were detected automatically by the computer. The mammograms were displayed on a workstation with a graphical user interface that allowed interactive thresholding of the gray level histograms to segment the dense region from the fatty region. The DMs and SFMs were segmented independently in separate sessions so that the observer could not compare the density of the corresponding DM and SFM. Hard copies of the displayed images were available for reference during segmentation. The mammographic density was estimated as the percent dense area relative to the breast area, excluding the pectoral muscle in the MLO views.

RESULTS: The correlation between the mammographic density on SFM and DM was 0.94 and 0.92, the root-mean-square residual was 4.5% and 4.6%, and the average ratio of mammographic density estimated on SFM to that on DM of the same breast was 1.18 and 1.22, respectively, for CC and MLO views. The differences in the percent dense area between the DM and SFM were statistically significant (paired t test: $p < 0.0000001$) for both views. The DMs used harder beams (Mo/Mo 4.5%, Mo/Rh 22.4%, Rh/Rh 73.1%) while the SFMs used softer beams (Mo/Mo 44.2%, Mo/Rh 48.1%, Rh/Rh 7.8%). The peak potential used for DM was 1 to 5 kVp higher than that for SFM in 84% of cases.

CONCLUSION: Breast density on DMs generally appears to be lower than that on SFMs because of the harder beam quality used and image processing applied to the DMs. The lower density may improve the mammographic sensitivity for lesion detection on dense breasts. However, for patients with SFMs and DMs taken over time, comparison of serial mammograms for breast density changes will be problematic.

Computerized mammographic breast density estimation: Expectation-Maximization estimation and neural network classification of breast density

Chuan Zhou, Lubomir M.Hadjiiski, Berkman Sahiner, Heang-Ping Chan, Mark A. Helvie, Jun Wei

PURPOSE:

Our previous study showed the feasibility of a rule-based automatic breast density estimation method. However, the rule-based technique could not classify the very fatty and very dense breasts consistently with high accuracy because some of these breasts have very similar gray level histograms. This study develops a new neural network classifier to improve the performance of rule-based breast density estimation.

METHOD & MATERIALS:

A mammogram is digitized and the pixel size is reduced to 0.8 mm. The breast region is first segmented by an automatic boundary tracking and a pectoral muscle trimming algorithm. An adaptive dynamic range reduction technique is used to reduce the range of the gray levels in the low frequency background and to enhance the separation of the gray levels of the dense and fatty regions. The breast images are first classified by a rule-based method into a class of median dense and a class of combined very dense/fatty breasts based on the characteristics of their gray level histograms. An Expectation-Maximization (EM) algorithm is then applied to the latter class to extract the gray level features. One morphological feature and 12 EM extracted gray level features are input to a feedforward neural network to further classify the mammograms in the combined class into a class of very dense breasts and a class of very fatty breasts. For each class, a gray level threshold is automatically estimated to segment the dense tissue. For comparison, an experienced radiologist provided a manually segmented percent dense area by interactive thresholding.

RESULT:

In this preliminary study, 498 mammograms from 141 patients were used and 243 were classified into the very dense/fatty combined class by the rule-based classifier. With a jackknife method, this class was randomly partitioned into four non-overlapping groups. In each jackknife cycle, three groups were used for training and one group for testing. The overall accuracy for classification of the four test groups into very dense and very fatty breasts reached 99.6% by the neural network, and 84.8% could be reached by our previous rule-based classifier.

CONCLUSION

The results demonstrate the feasibility of training a neural network classifier for the classification of very dense and fatty breasts. The neural network can be trained very well using the morphological feature and the features extracted by the EM algorithm. Combining the rule-based method with the NN classifier, the two-stage classification improved the performance of our previous breast density estimation technique.

**Performance evaluation of an automated breast density estimation system
for digital mammograms and digitized film mammograms**

Zhou C, Chan HP, Wei J, Helvie MA, Roubidoux MA, Paramagul C, Nees A, Hadjiiski LM, Sahiner B

Studies have demonstrated a strong correlation between mammographic breast density and breast cancer risk. We have previously developed a computerized system, mammographic density estimator (MDEST), to estimate breast density automatically on digitized film mammograms (DFM). In this study, we evaluated the performance of the MDEST system on full field digital mammograms (FFDM) and DFMs. The input to the system is a preprocessed dynamic range compressed image, the breast region is first segmented by breast boundary detection. The pectoral muscle is trimmed if it is an MLO view. A rule-based classifier is then used to classify the breast image into one of four classes according to the characteristics of its gray level histogram. A gray level threshold is determined to segment the dense area from the breast region. The breast density is estimated as the percentage of the segmented dense area relative to the breast area. In this study, two-view FFDM and the corresponding DFM from 99 patients with 202 images in each set were used. The dense area on each mammogram was segmented by 4 radiologists using interactive thresholding and their average was used as the “ground truth.” The MDEST system was directly applied to the FFDM and DFM data without any re-training except that the preprocessing filter was modified for FFDMs. We found that the correlation between the estimated percent dense area and the truth was 0.846 (0.813, 0.862) and 0.855 (0.808, 0.861) on FFDM, and 0.880 (0.867, 0.858) and 0.765 (0.766, 0.739) on DFM, for CC and MLO views, respectively. The results demonstrated the feasibility of estimating breast density automatically on FFDM and DFM using the same MDEST system.

Computerized pectoral muscle identification on MLO-view mammograms for CAD applications

Chuan Zhou, Lubomir M. Hadjiiski, Chintana Paramagul, Berkman Sahiner,
Heang-Ping Chan, Jun Wei

(Department of Radiology, The University of Michigan, Ann Arbor, MI 48109-0904)

SUMMARY: Automatic identification of the pectoral muscle on MLO view is an essential step for computerized analysis of mammograms. It can reduce the bias of mammographic density estimation, will enable region-specific processing in lesion detection programs, and also may be used as a reference in image registration algorithms. We are developing a computerized method for the identification of pectoral muscle on mammograms.

The upper portion of the pectoral edges was first detected to estimate the direction of the pectoral muscle boundary. A gradient-based directional bandpass (GDB) filter was used to enhance the linear texture structures, and then a gradient-based texture analysis is designed to extract a texture orientation image that represented the dominant texture orientation at each pixel. The texture orientation image was enhanced by a second GDB filter. An edge flow propagation method was developed to extract edges around the pectoral boundary using geometric features and anatomic constraints. The pectoral boundary was finally generated by a second-order curve fitting. 118 MLO view mammograms were tested in this study. The pectoral muscle boundary identified on each image by an experienced radiologist was used as the gold standard.

The accuracy of pectoral boundary detection was evaluated by two performance metrics. One is the overlap percentage between the computer-identified area and the gold standard, and the other is the root-mean-square (RMS) distance between the computer and manually identified pectoral boundary. For 118 MLO view mammograms, 99.15% (117/118) of the pectoral muscles could be identified. The average of the overlap percentage is 85.9% with a standard deviation of 14.0%, and the average of the RMS distance is 4.31 mm with a standard deviation of 5.94 mm. These results indicate that the pectoral muscle on mammograms can be detected accurately by our automated method.

Computerized pectoral muscle identification on MLO-view mammograms for CAD applications

Chuan Zhou, Lubomir M. Hadjiiski, Chintana Paramagul, Berkman Sahiner,
Heang-Ping Chan, Jun Wei

(Department of Radiology, The University of Michigan, Ann Arbor, MI 48109-0904)

PURPOSE: The pectoral muscle is imaged in most of the medial-lateral oblique (MLO) view mammograms. Automatic identification of the pectoral muscle is an essential step for computerized analysis of mammograms. It can reduce the bias of mammographic density estimation, will enable region-specific processing in lesion detection programs, and also may be used as a reference in image registration algorithms. The goal of this study is to develop an automated method to identify the pectoral muscle on mammograms.

METHOD & MATERIALS: Because of noise and the overlapping structures, the pectoral boundary is usually detected as pieces of edges mixed with a large number of false edges. The pieces of edges have to be pruned and linked to obtain a continuous boundary. We are developing a pectoral boundary detection algorithm based on gradient-based texture orientation analysis. The interference due to overlapping of the glandular tissue on the pectoral muscle region is first reduced by blurring the mammogram using an edge preserving anisotropic diffusion filter. Because less glandular tissue appears at the upper region of the pectoral muscle, the upper portion of the pectoral boundary usually remains sharp after blurring and can be detected robustly by searching the maximum horizontal gradients on the diffused image. The extrapolation of the detected upper pectoral boundary provides a coarse global direction of the pectoral boundary. To explore and refine the entire pectoral boundary, a gradient-based directional bandpass (GDB) filter is first employed to enhance the linear texture structures on the mammogram. A gradient-based texture analysis is then used to compute an orientation image which represents the dominant texture orientation at each pixel. The orientation image is smoothed using an edge preserving mean shift algorithm that iteratively shifts each pixel to the average of the pixels in its neighborhood. A second GDB filter is applied to the smoothed orientation image to enhance the texture patterns that have high similarity in the dominant texture orientations. Candidate edges of the pectoral muscle are detected on the enhanced orientation image using a ridge-tracking algorithm. With the guidance of the estimated global direction of the pectoral boundary and the anatomical constraints, an edge flow propagation algorithm is used to extract the boundary points of the pectoral muscle by pruning the edges that are less likely to lie on the pectoral boundary. A second order curve fitting is used to generate the final pectoral muscle boundary. 118 MLO view mammograms from 103 patients were randomly selected for the evaluation of the algorithm in this study. The mammograms were digitized with 0.05 mm/pixel resolution and reduced to a resolution of 0.8 mm/pixel for pectoral muscle identification. An experienced radiologist used a graphical user interface to manually draw the pectoral muscle boundary on each mammogram, which was then used to define the pectoral muscle region and used as the gold standard for the evaluation of computer performance.

RESULTS: For each MLO view mammogram, the accuracy of pectoral boundary detection was evaluated by two performance metrics: the percentage of computer detected pectoral muscle area overlapped with the gold standard, and the root-mean-square (RMS) distance between the computer-identified pectoral boundary and the manually marked pectoral boundary. For 118 MLO view mammograms, 99.15% (117/118) of the pectoral muscles could be identified, the average of the percent overlap area is 85.9% with a standard deviation of 14.0%, the average of the RMS distance is 4.31 mm with a standard deviation of 5.94 mm.

NEW WORK TO BE PRESENTED: Accurate identification of the pectoral muscle on mammograms is challenging, especially for the improperly positioned MLO-view images and the images containing dense glandular tissues overlapping with the pectoral muscle region. In this work, we developed a two-stage gradient-based texture analysis method to detect the pectoral boundary. In the first stage, linear texture structures were enhanced and the directional gradients were computed using a directional bandpass filter. In the second stage, a texture orientation image was derived as the dominant texture orientation at each pixel. A diffusion filter was used to estimate the global direction of the pectoral boundary. An edge flow propagation method was developed to extract the pectoral edges with the guidance of the estimated global direction.

CONCLUSION: The newly developed gradient-based directional filter and the dominant texture orientation image estimation method can enhance the pectoral boundary regions. The edge flow propagation method can accurately extract pectoral edges to generate the pectoral boundary. Automatic pectoral muscle identification will provide the foundation for many image analysis tasks in CAD applications.

Computerized mammographic breast density estimation on full field digital mammogram and digitized film mammogram

Chuan Zhou, Heang-Ping Chan, Mark A. Helvie, Jun Wei, Jun Ge, Lubomir M. Hadjiiski, Chintana Paramagul, Marilyn A. Roubidoux, Caroline E. Blane, Berkman Sahiner

PURPOSE:

We have previously developed an automatic mammographic density estimator (MDEST) on digitized film mammograms (DFM). In this study, we modified MDEST to estimate breast density on full field digital mammograms (FFDM) and further improved the performance of the MDEST on DFM.

METHOD & MATERIALS:

The breast region is first extracted by breast boundary detection. The pectoral muscle is trimmed if it is an MLO view. An adaptive dynamic range reduction technique is used to reduce the gray level range in the low frequency background. The breast image is classified into one of four classes ranging from fatty to very dense based on the characteristics of their gray level histograms. For each class, an Expectation-Maximization (EM) algorithm is developed to extract gray level features and a rule-based classifier is trained to segment the dense regions from the fatty background. The parameters of the new rule-based method are trained separately for FFDMs and DFMs. The breast density is estimated as the percentage of the segmented dense area relative to the breast area. Two-view FFDMs and the corresponding DFMs from 99 patients with 202 images in each set were used as the test set. The computerized segmentation on the two sets of mammograms is compared to the “gold standard”, which is obtained from interactive thresholding segmentation averaged over 4 MQSA radiologists for each mammogram.

RESULT:

For FFDM, the correlation between the computer estimated percent dense area and the gold standard was 0.94 for CC view, 0.92 for MLO view, and 0.96 for each breast with the percent dense area estimated as the average of two views. The corresponding root-mean-square (RMS) error was 4.2%, 4.4%, and 3.5%, respectively. For DFM, the corresponding correlations were 0.88, 0.86 and 0.92 with RMS error of 7.0%, 7.1% and 5.7%, respectively.

CONCLUSION

The results demonstrate the feasibility of estimating breast density automatically on FFDMs and DFMs using the same MDEST system by only incorporating a new EM estimation step. The adaptability of the new EM method improved the robustness of our breast density estimation technique for mammograms acquired with different imaging systems.