AD_____

Award Number: DAMD17-98-1-8211

TITLE: Analysis of Interval Changes on Mammograms for Computer
Aided Diagnosis

PRINCIPAL INVESTIGATOR: Lubomir Hadjiiski, Ph.D.

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

REPORT DATE: May 2000

TYPE OF REPORT: Annual Summary

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

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 Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE May 2000	3. REPORT TYPE AND DATES COVERED Annual Summary (6 Apr 99 - 5 Apr 00)

4. TITLE AND SUBTITLE
Analysis of Interval Changes on Mammograms for Computer Aided Diagnosis

5. FUNDING NUMBERS
DAMD17-98-1-8211

6. AUTHOR(S)
Lubomir Hadjiiski, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Michigan
Ann Arbor, Michigan 48109-1274

E-MAIL:
lhadjisk@umich.edu

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. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

A multistage regional registration technique was developed for identifying masses on temporal pairs of mammograms. In the first stage, an initial fan-shape search region was defined on the prior mammogram. In the second stage, the location of the fan-shape region was refined by warping, based on an affine transformation and simplex optimization. A new refined search region was defined on the prior mammogram. In the third stage a search for the best match between the lesion template from the current mammogram and a structure on the prior mammogram was carried out within the search region. This technique was evaluated on 124 temporal pairs of mammograms containing biopsy-proven masses. Eighty-seven percent of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations. The average distance between the estimated and the true centroid of the lesions on the prior mammogram was 4.2±5.7 mm. The registration accuracy was improved in comparison with our previous study that used a data set of 74 temporal pairs of mammograms. This improvement gain is mainly from the local affine transformation. This technique can be useful for identification of corresponding lesions on temporal pairs of mammograms.

14. SUBJECT TERMS
Breast Cancer, Computer-aided diagnosis, Screening, Classification, Image Analysis

15. NUMBER OF PAGES
12

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

20010228 078

FOREWORD

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

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

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

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

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

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

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

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

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

L. Hodjiiski   _05/04/2000_
PI - Signature   Date

(4)   Table of Contents

(5)   Introduction

Treatment of breast cancer at an early stage can significantly improve the survival rate of patients. Mammography is currently the most sensitive method for detecting early breast cancer [1, 2], and it is also the most practical for screening. Although general rules for the differentiation between malignant and benign lesions exist, in clinical practice, only 15-30% of cases referred for surgical biopsy are actually malignant. A number of research groups are in the process of developing computer-aided diagnosis (CAD) methods which can provide a second opinion to the radiologist for the detection and classification of breast abnormalities.

Radiologists routinely use several mammograms of different views of a patient with those obtained in previous years for identifying interval changes, detecting potential abnormalities, and in evaluating breast lesions. It is widely accepted that interval changes in mammographic features are very useful for both detection and classification of breast abnormalities. Some existing CAD techniques use information from multiple views of the same breast. Others use previous mammograms for detection. However none incorporates information about the temporal mammographic changes in the breast tissue for classification.

The goal of this project is to evaluate the usefulness of using interval changes to distinguish between normal structures, benign masses, and malignant masses in CAD. The purpose of this study is summarized as follows: 1. Characterize temporal changes in terms of the mammographic features of normal breast structures, as well as benign and malignant masses. 2. Use this information to develop methods for CAD. We hypothesize that the use of temporal changes in mammographic features between current and previous mammograms of the patient will improve the success of CAD technique for classification of masses. It is therefore expected that the use of such temporal information will improve the positive predictive value of mammography by reducing benign biopsies, and hence reduce both cost and patient morbidity.

To accomplish this goal we will first develop and evaluate reliable techniques for the temporal regional registration of mammograms of the same patient. The temporal mammogram registration technique we have developed is a novel approach in which the computer emulates the search method used by many radiologists for finding corresponding structures on mammograms. The method aims at registering a small region containing a suspected mass on the most recent mammogram of the patient with one on a mammogram obtained from a previous year. Our regional registration technique involves three steps: (1) identification of a suspicious structure on the most recent mammogram, (2) initial estimation of the location on a previous mammogram of the region corresponding to the suspicious structure and the definition of a search region which encloses the object of interest on the previous mammogram, and (3) accurate identification of the location of the matched object within the search region. The characteristic features of the two matched lesions then will be automatically extracted and interval changes estimated. This interval change information will be incorporated in an integrated CAD system.

(6)   Body

In the second year (4/6/99-4/5/00) of this grant, we have performed the following studies:

(A)   Database collection and extraction of regions of interest

We continued collecting the data set for this study from the files of patients who had undergone biopsy at the University of Michigan. The mammograms are scanned and the images are saved in our storage device using automated graphic user interface developed in our laboratory. Additionally the film information is recorded in a Microsoft Access database. Temporal pairs of images were obtained. The current mammogram of each temporal pair exhibited a biopsy-proven mass. We scan both cranio-caudal and mediolateral-oblique views. The mammograms were digitized with a LUMISCAN 85 laser scanner at a pixel resolution of 0.05 mm x 0.05 mm and with 12-bit resolution.

While the regional registration technique can be used for determining a corresponding structure or region for any structure (both normal tissues and masses) in the breast, in this study we are analyzing its accuracy on biopsy-proven masses alone. The location of the mass on the current mammogram identified by an MQSA-approved radiologist experienced in breast imaging using an interactive image analysis tool on a UNIX workstation. To provide the ground truth for evaluation of the computerized method, the radiologist manually identifies the corresponding region on the previous mammogram. Bounding boxes enclosing the mass on the current mammogram and the corresponding object on the previous mammogram are also provided by the radiologist for each case. Each mass as well as the corresponding structure on the previous mammogram are rated for its visibility on a scale of 1 to 10, where the rating of 1 corresponded to the most visible category. The size of the mass on the current mammogram as well as the size of the corresponding structure on the previous mammogram are also measured by the radiologist. The parenchymal density is rated based on the BI-RADS lexicon.

(B)   Develop methods for establishing corresponding locations in current and previous mammograms

A multistage regional registration technique was developed for identifying corresponding masses on temporal pairs of mammograms. At the first stage an initial fan-shape search region was defined on the prior mammogram based on the mass location on the current mammogram. At the second stage the location of the search region on the prior mammograms was first refined by maximizing a correlation measure between a template extracted from the current mammogram and the breast structures on the prior mammogram. The affine transformation in combination with simplex optimization was then employed to warp this local region. In the final stage a search for the best match between the lesion template from the current mammogram and a structure on the prior mammogram was carried out within the refined search region. The new registration method [8], [9] allows simplification at the first stage (compared to the method we reported last year [7]) eliminating the need of global breast alignment procedure and at the same time improving the registration results. A more detailed explanation for each of the stages will be presented in the following subsections.

Definition of fan-shape regions

Initially an automatic procedure is used to detect the breast boundary for all of the mammograms. The nipple locations of the two breast images were used as the pivot points in a common reference frame for the alignment.

The location of the mass on the current mammogram is determined in a polar coordinate system with the nipple as the origin. The location is represented as the radial distance from the nipple and the angle between the nipple-mass centroid axis and the *breast periphery* (r, θ). This is a difference from the method reported last year. Only angular scaling is performed (estimation of the angular scale factor s_2) and the position of the mass on the prior mammogram is predicted in a polar coordinate system defined in a similar manner $(r, s_2\theta)$. In this case there is no need for global breast alignment procedure (compared to the method we reported last year [7]). An initial fan-shape search region is then defined on the prior mammogram centered at the predicted location of the mass centroid [7]. The size of the fan-shape region $(r \pm \delta$ and $s_2\theta \pm \epsilon)$ is predefined using a training set so that it will include the mass centroid on the prior mammograms. A fan-shape template centered at the mass is also defined on the current mammogram.

(C)   Develop methods for correlation of breast regions in current and previous mammogram

Warping and alignment

In the second stage the fan-shape search region is refined. By allowing warping of the fan-shape template from the current mammogram, it may provide better compensation for local geometric distortions due to differences in positioning of the breast and in breast compression and may improve the localization of the mass on the prior mammogram. The warping procedure is based on the affine transformation combined with simplex optimization. In the following the warping procedure is explained in greater detail.

Affine transformation

An affine transformation [3] is a linear transformation combining rotation and translation. A two dimensional affine transformation is defined as follows:

$$
\begin{aligned}
x' &= ax + by + c \\
y' &= dx + ey + f
\end{aligned}
\tag{1}
$$

where (x, y) are the original coordinates, (x', y') are the transformed coordinates, and a, b, d, e, c, f are the transformation coefficients. The coefficients a, b, d, e determine a scaling and a rotation, and the coefficients c and f determine a translation. The result of the application of the affine transformation of Eq. (1) in combination with simplex optimization (described below) is shown in Fig. 1. Since the affine transformation is linear, the transformed objects are linearly resized and rotated. This can be observed from the edges of the fan-shape region bounding box (the white box in Fig. 1). After the transformation the edges are still straight lines, however, the corner angles are different from 90 degrees and the length of the lines is also linearly scaled.

Nonlinear Simplex Optimization

The Nelder and Mead [4], [5] nonlinear simplex optimization is used in order to adjust the coefficients a, b, c, d, e and f and to warp the fan-shaped template in order to maximize the

correlation between the template and a breast structure on the prior mammogram. This optimization defines a hyper polygon. For each vertex an error function is calculated. Then the polygon is "rolled" towards the minimum. The movement of the polygon (toward the minimum) is obtained by the reflection in the direction opposite to the vertex with maximal error. Fig. 1 shows the result of application of the affine transformation whose coefficients were obtained by the nonlinear simplex optimization.

Fan-shape
template
(x,y)

Warped
fan-shape
template
(x',y')
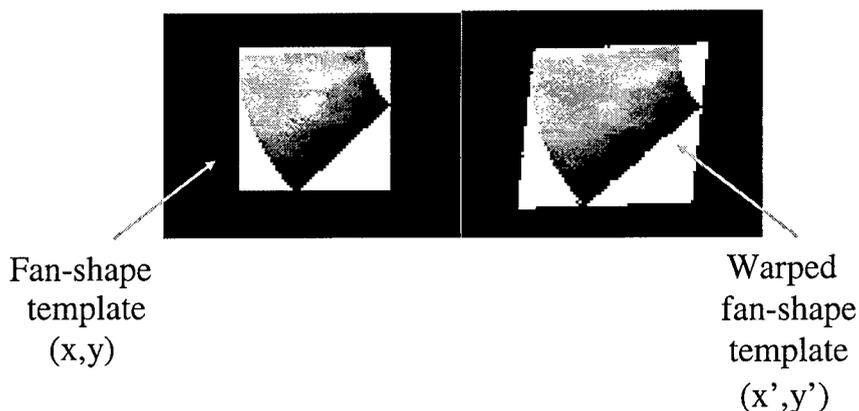
Figure 1. Fan-shaped template and warped fan-shaped template by the affine transformation.

Mass template alignment and identification of corresponding lesion

In this stage a new search region with a reduced size is defined on the prior mammogram. A template containing the mass is extracted from the current mammogram. Then, the mass location on the prior mammogram is determined by maximizing the correlation between the template and a structure within the search region.

Once a corresponding structure is found on the previous mammogram for a suspicious object on the current mammogram, it can be used for an interval change analysis within a CAD scheme, as we have shown in an independent study [6]. If the search procedure in the fan-shaped region does not yield a corresponding region, then the suspicious object on the current mammogram can be considered as a newly developed density. Objects for which no corresponding object can be found on the previous mammogram can be analyzed with methods designed for single images in an overall CAD scheme.

(D)   Develop quantitative measures for assessing regional registration technique

The accuracy of the multistage regional registration was analyzed in terms of two measures. The first measure is the overlap area between the estimated and the true lesions on the prior mammogram. The fractions of registered temporal pairs that could provide an accuracy of over 50% area overlap and over 75% area overlap were examined. The second measure is the average Euclidean distance between the centroids of the estimated and true lesion locations

(E)   Evaluate accuracy of regional registration technique

The regional registration technique has been evaluated with a data set of 124 temporal pairs of mammograms.

Initial Estimates and Search Regions

The Euclidean distance between the initial estimate of the centroid location of the corresponding structure on the previous mammogram and the center of the bounding box provided by the radiologist was estimated. For the 124 temporal image pairs used in this data set, the average Euclidean distance error of the initial estimate was 8.5 mm (std. dev. 5.4 mm). Based on observation of the radial deviation errors and the angular deviation errors, a search region defined by $\varepsilon = 0.25 + 5/r$ radians and $\delta = 20$ mm, where r is the radial distance from the nipple, was used for the evaluation of the local search criteria used in Stage 3 of regional registration. These results show improvement compared to the results obtained previously ($\varepsilon = 0.35 + 5/r$, and error 9.8 mm (std. dev. 6.0 mm)) .

Local Search Criteria and Final Estimates

In this study 87% of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations. The average distance between the estimated and the true centroids of the lesions on the prior mammogram was 4.2 ± 5.7 mm with a maximum of 31.6 mm. These results are presented in Table 1 and Table 2. For the 87% of the temporal pairs with 50% overlap, the average distance between the estimated and the true centroids of the lesions on the prior mammogram was 2.4 ± 2.1 mm with a maximum of 10.2 mm.

Table 1. The area overlap between the true and the estimated masses on the prior mammogram.

Pairs	50% overlap	75% overlap
Number	108	101
%	87%	82%

Table 2. The distance between the true and the estimated centroids of the mass on the prior mammogram.

	Overall	50% overlap	75% overlap
Mean distance	4.2 mm	2.4 mm	2.2 mm
Standard. Deviation.	5.7 mm	2.1 mm	1.9 mm
Max. distance	31.6 mm	10.2 mm	10.2 mm

(F)   Further development of the regional registration technique

We will continue to improve the regional registration technique. In the first step, an automated method will be developed to detect the nipple location in the breast image. The method will be based on both the change of tangential direction and the change in the tissue density along the breast border. In the third step we have investigated the use of the density-weighted contrast enhancement (DWCE) technique to improve the localization of the corresponding mass on the prior mammogram. After the location of the fan-shaped region was refined by affine transformation and simplex optimization, the DWCE technique was used to segment dense structures within the search region. A search for the best match between the lesion template from the current mammogram and a structure on the prior mammogram was performed within the DWCE segmented densities. The

preliminary results of this study are promising. The DWCE segmentation improved the accuracy of matching by directing the search to the dense structures and thus reducing the chance of mismatch. The average Euclidean distance between the computer estimate of the corresponding structure and the radiologist-identified location and their standard deviation were both reduced. We will present the preliminary results on this improved method at the upcoming World Congress on Medical Physics and Biomedical Engineering, Chicago, July 23 - 28, 2000 [9].

(G)   Initial development of the mass segmentation and feature extraction methods

We already started and will continue the development of automated method to extract and analyze features extracted from corresponding masses on a temporal pair of mammograms. Regions of interest containing the corresponding masses were identified on the current and prior mammograms of the temporal pair. The masses were automatically segmented using an active contour model. Run Length Statistics (RLS) features, spiculation features, and mass size were extracted from each mass. An additional difference RLS features were obtained by subtracting the RLS features of the prior mass from those of the current mass for each temporal pair. The feature space for each temporal pair consisted of the RLS and spiculation features from both the prior and the current mammograms and the difference RLS features. Stepwise feature selection with simplex optimization was used to select the optimal feature subset. A linear discriminant classifier (LDA) was used to merge the selected features for classification of malignant and benign masses. A leave-one-case-out training and testing resampling scheme was used for feature selection and classification. The preliminary results of this study are promising. The size of the mass is not a useful feature for difficult cases because many benign masses grow over time. The difference RLS and prior spiculation features are useful for identification of malignancy in temporal pairs of mammograms. Further studies are underway to improve the technique and to evaluate the performance on a larger data set. We have submitted an abstract on this improved method for presentation at the upcoming RSNA meeting [10].

(7)   Conclusion

During this year, we have continued the development of the regional registration technique. The warping procedure based on an affine transformation in the local alignment stage reduces the size of the search region. Eighty-seven percent of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations. The average distance between the estimated and the true centroids of the lesions on the prior mammogram was 4.2 ± 5.7 mm. When the threshold was set to 75% area overlap, 82% of the temporal pairs could still exceed the threshold. The registration accuracy of the current method has been improved in comparison with that of our previous method [7], although the data set was increased from 74 pairs to 124 pairs. This improvement is obtained mainly from the second stage affine transformation and simplex optimization. This result indicates that our technique is a promising approach for identification of corresponding lesions on temporal pairs of mammograms and thus may be used as a basis for analysis of interval change on mammograms. We will continue to enlarge the data set and improve the registration method in future years. Further study is underway to develop a feature matching method to improve lesion localization within the search region. We will continue the development of automated method to extract and analyze features extracted from corresponding masses on a temporal pair of mammograms for analysis of the temporal changes.

(8)   References

[1]   H. C. Zuckerman, "The role of mammography in the diagnosis of breast cancer," in *Breast Cancer, Diagnosis and Treatment*, edited by I. M. Ariel and J. B. Cleary (McGraw-Hill, New York, 1987), pp. 152-172.

[2]   C.C. Boring, T. S. Squires, T. tong and S. Montgomery, "Cancer statistics 1994", CA-A Cancer Journal for Clinicians 44, 7-26, 1994.

[3]   L. Quan, T. Kanade, "Affine structure from line correspondance with uncalibrated affine cameras", *IEEE Trans. Pat. Anal. Machine Intel.*, Vol. 19, No. **8**, Aug. 1997.

[4]   S.S. Rao, *Optimization: Theory and Applications*", Wiley Eastern Limited, 1979.

[5]   F.A. Lootsma, (ed). *Numerical methods for non-linear optimization*, Academic Press, New York, 1972.

[6]   S. Sanjay-Gopal, H. P. Chan, B. Sahiner, N. Petrick, T. Wilson, M. Helvie, "Evaluation of interval change in mammographic features for computerized classification of malignant and benign masses," Radiology **205p**, 216, 1997.

[7]   S.S. Gopal, H.P. Chan, T.E. Wilson, M.A. Helvie, N. Petrick, B. Sahiner, "A regional registration technique for automated interval change analysis of breast lesions on mammograms", *Medical Physics,* 1999, 26:2669-2679.

[8]   L. Hadjiiski, H.P. Chan, B. Sahiner, N. Petrick, M.A. Helvie, S.S. Gopal, "Automated identification of breast lesions in temporal pairs of mammograms for interval change analysis", Presented at the *85th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Nov.28-Dec.3, 1999, Chicago, Illinois. Radiology 1999; 213(P): 229-230.

[9]   L. Hadjiiski, H.P. Chan, B. Sahiner, N. Petrick, M.A. Helvie, S. Paquerault, C. Zhou, "Interval Change Analysis in Temporal Pairs of Mammograms Using a Local Affine Transformation", Poster Presentation at *SPIE International Symposium on Medical Imaging*, San Diego, California, February 12-18, 2000. To appear in *Proc. SPIE Medical Imaging 2000.*

[9]   L. Hadjiiski, N. Petrick, H.P. Chan, B. Sahiner, M. A. Helvie, C. Zhou, M. Gurcan, S. Paquerault, "Regional registration of masses on current and prior mammograms using DWCE segmentation", To be presented at *The World Congress on Medical Physics and Biomedical Engineering*, Chicago, July 23 - 28, 2000.

[10] L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvie, M. Gurcan, "Computer-Aided Classification of Malignant and Benign Breast Masses by Analysis of Interval Change of Features in Temporal Pairs of Mammograms", Submitted for presentation at the *86th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, 2000.

(9) Appendix

1. Key research accomplishments in current year as a result of this grant

- Database collection and extraction of regions of interest.
- Development of methods for establishing corresponding locations in current and previous mammograms.
- Development of methods for correlation of breast regions in current and previous mammograms.
- Develop quantitative measures for assessing regional registration technique.
- Evaluation of the accuracy of regional registration technique.
- Further development of the regional registration technique.
- Initial development of the mass segmentation and feature extraction methods.

2. Publications in current year as a result of this grant

[1] S.S. Gopal, H.P. Chan, T.E. Wilson, M.A. Helvie, N. Petrick, B. Sahiner, "A regional registration technique for automated interval change analysis of breast lesions on mammograms", *Medical Physics,* 1999, 26:2669-2679.

[2] L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvie, "Classification of Malignant and Benign Masses Based on Hybrid ART2LDA Approach", *IEEE Transactions on Medical Imaging*, Vol. 18, No. 12, Dec. 1999, pp. 1178-1187.

[3] L. Hadjiiski, H.P. Chan, B. Sahiner, N. Petrick, M.A. Helvie, S.S. Gopal, "Automated identification of breast lesions in temporal pairs of mammograms for interval change analysis", Presented at the *85th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Nov.28-Dec.3, 1999, Chicago, Illinois. Radiology 1999; 213(P): 229-230.

[4] L. Hadjiiski, H.P. Chan, B. Sahiner, N. Petrick, M.A. Helvie, S. Paquerault, C. Zhou, "Interval Change Analysis in Temporal Pairs of Mammograms Using a Local Affine Transformation", Poster Presentation at *SPIE International Symposium on Medical Imaging*, San Diego, California, February 12-18, 2000. To appear in *Proc. SPIE Medical Imaging 2000.*

[5] L. Hadjiiski, N. Petrick, H.P. Chan, B. Sahiner, M. A. Helvie, C. Zhou, M. Gurcan, S. Paquerault, "Regional registration of masses on current and prior mammograms using DWCE segmentation", To be presented at *The World Congress on Medical Physics and Biomedical Engineering*, Chicago, July 23 - 28, 2000.

[6] L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvie, M. Gurcan, "Computer-Aided Classification of Malignant and Benign Breast Masses by Analysis of Interval Change of Features in Temporal Pairs of Mammograms", Submitted for presentation at the *86th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, 2000.

3. Copies of publications are enclosed with this report.

A regional registration technique for automated interval change analysis of breast lesions on mammograms

S. Sanjay-Gopal, Heang-Ping Chan,[a] Todd Wilson, Mark Helvie, Nicholas Petrick, and Berkman Sahiner
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109-0030

Analysis of interval change is a useful technique for detection of abnormalities in mammographic interpretation. Interval change analysis is routinely used by radiologists and its importance is well-established in clinical practice. As a first step to develop a computerized method for interval change analysis on mammograms, we are developing an automated regional registration technique to identify corresponding lesions on temporal pairs of mammograms. In this technique, the breast is first segmented from the background on the current and previous mammograms. The breast edges are then aligned using a global alignment procedure based on the mutual information between the breast regions in the two images. Using the nipple location and the breast centroid estimated independently on both mammograms, a polar coordinate system is defined for each image. The polar coordinate of the centroid of a lesion detected on the most recent mammogram is used to obtain an initial estimate of its location on the previous mammogram and to define a fan-shaped search region. A search for a matching structure to the lesion is then performed in the fan-shaped region on the previous mammogram to obtain a final estimate of its location. In this study, a quantitative evaluation of registration accuracy has been performed with a data set of 74 temporal pairs of mammograms and ground-truth correspondence information provided by an experienced radiologist. The most recent mammogram of each temporal pair exhibited a biopsy-proven mass. We have investigated the usefulness of correlation and mutual information as search criteria for determining corresponding regions on mammograms for the biopsy-proven masses. In 85% of the cases (63/74 temporal pairs) the region on the previous mammogram that corresponded to the mass on the current mammogram was correctly identified. The region centroid identified by the registration technique had an average distance of 2.8 ± 1.9 mm from the centroid of the radiologist-identified region. These results indicate that our new registration technique may be useful for establishing correspondence between structures on current and previous mammograms. Once such a correspondence is established an interval change analysis could be performed to aid in both detection as well as classification of abnormal breast densities. © *1999 American Association of Physicists in Medicine.* [S0094-2405(99)00612-4]

Key words: image registration, computer-aided diagnosis, computer vision, interval change, breast cancer

I. INTRODUCTION

Mammography is currently the most effective method for early breast cancer detection.[1,2] A variety of computer-aided diagnosis (CAD) techniques have recently been developed to detect mammographic abnormalities and to distinguish between malignant and benign lesions.[3-8] Knowledge from diverse areas such as signal and image processing, pattern recognition, computer vision, artificial intelligence, and neural networks has been used to develop algorithms to be implemented within a CAD scheme. Varying degrees of success for these approaches have been reported in the literature. One common feature of most of these CAD techniques is that they use a single mammogram for analysis. However, some malignancies may only manifest as a new density on mammograms without associated calcifications or masses, others distinguish themselves from benign lesions only by their relatively rapid changes in sizes. Therefore, radiologists routinely use several mammographic views along with mammograms obtained in previous years for detecting and evaluating breast lesions and for identifying interval changes. The importance of interval change analysis in mammographic interpretation has been established in clinical practice.[9,10] It can be expected that analysis of changes in mammographic features between current and previous mammograms of the patient will also be an important component of a CAD system for both the detection and the classification tasks. The ability for automated analysis of interval changes would further the ability of CAD to offer an objective second opinion. This improvement, in turn, could increase the positive predictive value of mammography, reduce the number of benign biopsies, and hence reduce both cost and patient morbidity.

While a number of CAD schemes use only a single mammogram, the simultaneous use of more than one mammogram has been under investigation for some time. Several researchers have used views of the contra-lateral breast for detecting masses and developing densities. For instance, Yin

et al.[11,12] have utilized architectural asymmetry between the right and left breasts to detect masses. While it is widely accepted that interval changes in mammographic features are very useful for both detection and classification of breast abnormalities, the development of CAD techniques to use this information has achieved limited success.[13–18] Sallam and Bowyer[13] have proposed a warping technique for mammogram registration. They manually obtained control points and calculated a mapping function for mapping each point on the current mammogram to a point on the previous mammogram. The mapping function was obtained based on local affine transformations, as well as interpolation and surface fitting techniques. A drawback of this technique is the need for manual demarcation of control points. Brzakovic *et al.*[14] have investigated a three-step method for comparison of most recent and previous mammograms. They first registered two mammograms using the method of principal axis, and partitioned the current mammogram using a hierarchical region-growing technique. The breast regions in the two mammograms were aligned with respect to each other by means of translation, rotation, and scaling. Although the technique was evaluated on a total of 64 images obtained from eight cases, this work mainly aimed toward detecting cancerous changes in breast tissue and, therefore, no quantitative analysis of registration accuracy was presented. Vujovic and co-workers[15,16] have proposed a multiple-control-point technique for mammogram registration. They first determined several control points independently on the current and previous mammograms based on the intersection points of prominent anatomical structures in the breast. A correspondence between these control points was established based on a search in a local neighborhood around the control point of interest. In a more recent publication,[17] they have evaluated their approach for establishing the correspondence between control points extracted from two mammograms using 29 temporal image pairs, and presented a qualitative evaluation based on an observer study. They have demonstrated that 91% of 103 computer-matched control points were in agreement with those matched by a radiologist. An important assumption of their work was that the distances between the control points did not change significantly between the two mammograms. However, this assumption is not necessarily a valid one. Variations in compression could potentially cause a large variation in the relative distances between the control points. Furthermore, the control points representing the intersections of elongated structures do not always have correspondences on the two mammograms. Most of these points are two-dimensional projection image of structures at different depths of an elastic and compressible three-dimensional breast. The projected intersection points can thus vary from image to image and are not invariant lankmarks. As noted by the authors, the potential control points are not points that are naturally selected by a radiologist when examining mammograms. Hence, the significance of these points is debatable.

An important factor that may limit the success of the above-mentioned techniques is that the extraction of any meaningful information from previous mammograms first re-

quires a common frame of reference between the current and previous mammograms. Several complicating factors confound obtaining such a frame of reference. These factors include differences in breast compression and positioning between the current and previous mammograms, differences in the imaging technique between the two examinations, and changes in breast structure, size, and tissue density between the two images with patient age. As a result, the mammographic appearance of breast tissue on the current and previous mammograms of the same patient may vary considerably. Although these variabilities have not been quantified experimentally, they can be observed easily from most mammograms. Conventional registration techniques work well for applications involving rigid objects. Because of the elasticity of the breast tissue, the absence of obvious landmarks, and the large variability in the relative positions of the breast tissues projected onto the mammogram from one examination to the other, these techniques may not be optimal for registration of breast images.

In mammographic interpretation, a radiologist routinely compares the current mammogram with previous mammograms (if available) of the same view in order to detect changes in mammographic features. For example, if a mass is detected in the current mammogram, the radiologist searches for that mass in the previous mammogram to determine if this is a new or developing density. If the corresponding mass is found on the previous mammogram, then the radiologist compares the current and previous mass size and estimates if the mass has increased in size. To facilitate these comparisons, we plan to develop automated methods to detect the interval changes as a part of a computer-aided diagnostic system. As a first step, we have developed a novel method for automatic registration of lesions on temporal pairs of mammograms. In our approach, the computer emulates the search method used by many radiologists for finding corresponding structures on mammograms. The method aims at registering a small region containing a suspected mass on the most recent mammogram of the patient with one on a mammogram obtained from a previous year. Our regional registration technique involves three steps: (1) identification of a suspicious structure on the most recent mammogram, (2) initial estimation of the location on a previous mammogram of the region corresponding to the suspicious structure and the definition of a search region which encloses the object of interest on the previous mammogram, and (3) accurate identification of the location of the matched object within the search region. After the two matched lesions are identified, their characteristic features can be automatically extracted and interval changes estimated. In the present study, we focused on the development and the evaluation of the regional registration technique, rather than to solve the entire interval change analysis problem. The subsequent steps in the interval change analysis are beyond the scope of this study.

In the following sections we will provide a detailed description of our regional registration technique for temporal registration of mammograms and the results of a quantitative evaluation using a data set of 74 temporal image pairs. Although we evaluated a semiautomated version of the tech-
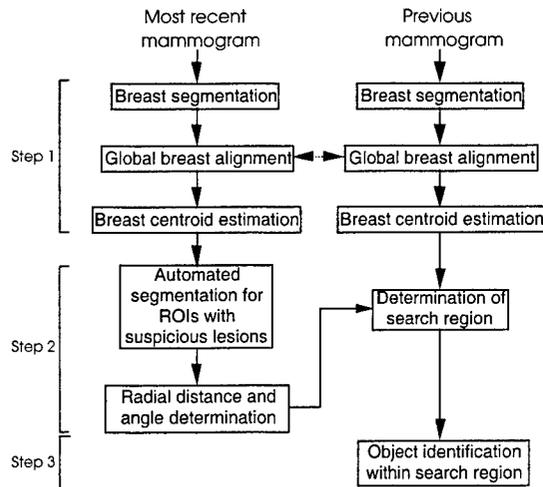
FIG. 1. Regional registration technique for determining an object on the previous mammogram which corresponds to a suspicious object on the most recent or current mammogram.

FIG. 2. Global breast alignment based on the mutual information between the two breast regions. N_c—nipple location in current mammogram, N_p—nipple location in previous mammogram, N—nipple location for both current and previous mammograms after translating them to the common frame of reference. The previous mammogram is rotated until the mutual information between the two mammograms is maximized.

nique in this preliminary study, it can be fully automated by incorporating a nipple detection step so that no user interaction will be required.

II. MATERIALS AND METHODS

A. Regional registration and mammogram correspondence

As the term indicates, regional registration is a local rather than a global registration technique. It is a multistep procedure and utilizes computer-detected objects in the most recent (hereafter termed current) mammogram. In the context of this paper, a current mammogram is either the latest mammogram of the patient, or the latest mammogram before biopsy. The detected objects could be either true masses (benign or malignant) or false positives (normal breast structures). Regional registration then finds a matching object on a previous mammogram. The three major steps in regional registration are illustrated in Fig. 1 and details of the technique are described below.

In the first step of regional registration, the breast region is segmented from the background on both the current and the previous mammograms. For this purpose we have used a breast boundary detection algorithm previously developed in our laboratory.[19,20] This algorithm could successfully track the breast boundaries in over 90% of the 1000 mammograms in a previous study. It performed reliably on all the images in our database. After extracting the breast border from the mammogram, the location of the nipple is estimated on both the current and the previous mammograms. Any automated method[21,22] can be used for finding the nipple location. However, in this study, the nipple location was manually identified by a radiologist for all images in our data set. The breast border and the nipple location now form the basis of a global breast alignment (GBA) procedure illustrated in Fig. 2. Since the sizes and the orientations of the two images could vary between the current and previous mammograms, a common frame of reference is needed. The GBA procedure has been

devised specifically to provide such a frame of reference. We first define a new frame of reference with the nipple location on the current mammogram (N_c) as the origin. The previous mammogram is translated so that its nipple location (N_p) aligns with the origin in the common frame of reference as shown in Fig. 2. Using the origin as the pivot point, we rotate the previous mammogram to align the breast regions in the two images.

We have evaluated two different methods for estimation of the optimum rotation angle. The first method is based on maximization of the overlap area, and the second method is based on maximization of the mutual information (MI)[23,24] between the two segmented breast regions. To determine the MI, we first rescale the breast portion of both mammograms to a 0–255 gray scale. For a given rotation angle θ, the two-dimensional (2D) histogram $h_\theta(i,j)$ of the gray levels for the corresponding pixels on the current mammogram and the previous mammogram is constructed. Here i refers to the gray level on the current mammogram and j refers to the gray level on the previous mammogram rotated by an angle θ. The probability density of the gray scale co-occurrences is estimated from the 2D histogram as

$$f_\theta(i,j) = \frac{h_\theta(i,j)}{\sum_{m,n} h_\theta(m,n)}, \tag{1}$$

where $0 \leq i,j \leq 255$, $0 \leq m,n \leq 255$. The mutual information (MI_θ) between the two images for a specific rotation angle θ is computed as

$$\text{MI}_\theta = \sum_{i,j} f_\theta(i,j) * \log_2 \left\{ \frac{f_\theta(i,j)}{\sum_m f_\theta(i,m) \sum_n f_\theta(n,j)} \right\}. \tag{2}$$
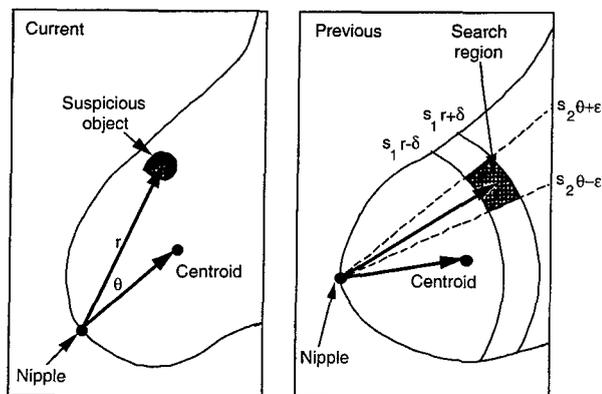
FIG. 3. Polar coordinate system defined using the nipple location and the nipple–centroid axis. The search region for finding a matching object on the previous mammogram is shown as the shaded region.

The above-mentioned procedure is repeated for several rotation angles and the angle θ_{max} which provides the maximum mutual information is chosen for global breast alignment of the previous mammogram and the current mammogram. Note that while the area overlap method for GBA uses the binary image after segmentation, the MI-based method uses the original gray scale image. The effects of the two methods on the accuracy of regional registration will be discussed later in Sec. IV. Once the two images are aligned in the common frame of reference, the centroid of the breast region is estimated, and the nipple–centroid axis is defined for both mammograms. For comparison we also show in Sec. III regional registration results based on computing the centroids of the two breast regions without global breast alignment. The nipple–centroid axis forms the basis for the second step of regional registration.

In the second step, suspicious regions are automatically segmented from the breast region on the current mammogram. This can be accomplished by using a density-weighted contrast enhancement (DWCE) technique[25] previously developed in our laboratory. While the use of the DWCE technique is not critical for regional registration, it does help automate the entire procedure. Alternatively, a radiologist can manually identify a suspicious object or a region of interest on the current mammogram and the regional registration technique can be used to identify a corresponding region on the previous mammogram. Once suspicious objects have been identified on the current mammogram, the centroid of each object is estimated. A polar coordinate system is then defined using the nipple as the origin and the nipple–centroid axis as the 0° axis on both images. This is illustrated in Fig. 3. The location of the centroid of a suspicious object on the current mammogram is determined as (r, θ). We then compute two scale factors—the radial scale factor s_1 and the angular scale factor s_2. These scale factors have been devised to provide a first-order correction for factors such as breast compression differences between the current and previous mammograms, differences in image magnification and size, and changes in overall breast shape between the two images. The radial scale factor s_1 is estimated as the ratio of

the nipple–centroid distances on the previous and current images. The angular scale factor s_2 is estimated as the ratio of the angular width of the breast on the previous image at radius $s_1 r$ to that on the current image at radius r. The initial estimate of the corresponding location of the suspicious object on the previous mammogram is then obtained as $(s_1 r, s_2 \theta)$.

Using the initial estimate of the centroid of the object on the previous mammogram, we can define a fan-shaped search region bounded by $s_1 r \pm \delta$ and $s_2 \theta \pm \epsilon$ as illustrated in Fig. 3. The object found on the current mammogram is then used as a template to search for a matching object in the search region on the previous mammogram. The size of the search region (defined by δ and ϵ) depends on the variability between mammograms obtained from one examination to the other. Since it is difficult to predict the variability of an elastic and deformable object such as the breast by any analytical method, we have determined this variability experimentally from the mammograms in our data set. The variation in compression can cause a change in the relative locations of various breast structures on these images as well as a rotation of the breast boundary with respect to the fixed image coordinates. By relating the position of a breast structure to the corresponding nipple–centroid axis, and by performing a search in the corresponding search region, we can reduce the effect of this variability. In this study we have estimated the size of the search region required to enclose all corresponding objects on the previous mammogram using ground truth objects identified on the previous mammograms by a radiologist. The distance of the initial estimate of the center of the search region from the centroid of the ground truth object was also estimated.

The third and final step in the regional registration procedure involves a systematic search to identify a corresponding structure within the fan-shaped search region on the previous mammogram. In this study we have evaluated two different search criteria. The first criterion is based on gray scale template matching. A rectangular gray scale template centered on the mass centroid is extracted from the current mammogram. The choice of the size of the template region can affect the accuracy of the registration technique. The minimum required size of a rectangular template is, of course, a rectangular region which encloses the mass exactly. However, one can also include a small portion of the background region in the template. We have analyzed the performance of our algorithm using two different sizes for this template. The first includes a 1-pixel-wide background region all around the boundary of the suspicious object while the second includes a 5-pixel-wide background region. For each pixel (i,j) in the fan-shaped region on the previous mammogram, a region of interest (ROI) centered on the pixel and of the same size as the mass template is extracted. We denote the (m,n)th pixel in the gray scale template extracted from the current mammogram as $p(m,n)$ and that from the ROI obtained from the fan-shaped region as $q_{i,j}(m,n)$. A correlation measure defined as

$$C_{i,j} = \frac{\Sigma_{m,n}(p(m,n)-\bar{p})(q_{i,j}(m,n)-\bar{q})}{\sqrt{(\Sigma_{m,n}(p(m,n)-\bar{p})^2)(\Sigma_{m,n}(q_{i,j}(m,n)-\bar{q})^2)}} \quad (3)$$

is then obtained for each pixel (i,j) within the search region on the previous mammogram. Here the summation is performed over the mass template, and $\bar{p}$ and $\bar{q}$ denote the average pixel values in the template and ROI, respectively. The correlation values in the search region are then smoothed by a 3×3 averaging kernel to reduce fluctuations. The final estimate of the location of the mass centroid on the previous mammogram is obtained as the location corresponding to maximum correlation. The second search criterion is based on maximizing the mutual information between the mass template and the ROI extracted from within the search region. The MI approach is similar to that described earlier for alignment of the breast regions, except that the regions to be matched are limited to the size of the mass template.

Once a corresponding structure is found on the previous mammogram for a suspicious object on the current mammogram, it can be used for an interval change analysis within a CAD scheme, as we have shown in an independent study.[26] If the search procedure in the fan-shaped region does not yield a corresponding region, then the suspicious object on the current mammogram can be considered as a newly developed density. Objects for which no corresponding object can be found on the previous mammogram can be analyzed with methods designed for single images in an overall CAD scheme. Note that in this study the search techniques are structured in a way to always determine a matching object. Search criteria to identify new densities will be developed in future studies.

B. Image acquisition and data set

The data set for this study consisted of 127 images obtained from the files of 34 patients who had undergone biopsy at the University of Michigan. From these 127 mammograms, 74 temporal pairs of images were obtained. The current mammogram of each temporal pair exhibited a biopsy-proven mass. All previous mammograms in the 74 temporal pairs contained a mass, a structure, or a density which the radiologist could match to the mass detected in the corresponding current image. Since some patient files contained a sequence of mammograms over three years, the number of temporal pairs was larger than half the number of

images. The 74 temporal image pairs were comprised of 43 cranio-caudal views and 31 mediolateral-oblique views.

The mammograms of 20 temporal pairs were digitized with a LUMISYS DIS-1000 laser scanner at a pixel resolution of 0.1 mm×0.1 mm and with 12 bit resolution. The digitizer was calibrated so that the gray values were linearly and inversely proportional to the optical density (OD) within the range of 0.1–2.8 OD units, with a slope of −0.001 OD/pixel value. Outside this range, the slope of the calibration curve decreased gradually. The OD range of this digitizer was 0–3.5. The mammograms of the remaining 54 temporal pairs were digitized with a LUMISCAN 85 laser scanner at a pixel resolution of 0.05 mm×0.05 mm and with 12 bit resolution. This digitizer was calibrated so that the gray values were linearly and inversely proportional to the OD within the range 0–4 OD units, with a slope of −0.001 OD/pixel value. All images were subsequently reduced to 0.8 mm resolution by averaging adjacent 8×8 pixels (20 pairs) or 16×16 pixels (54 pairs). Since the same digitizer was used for digitizing all films of the same case, the differences in the digitizers would have no effect on the analysis of each image pair. Given the small differences between the two laser digitizers and the large differences in the imaging technique and in the breast appearance from one case to another, it could be expected that the use of cases collected with the two different digitizers would not affect the evaluation of the registration technique.

While the regional registration technique can be used for determining a corresponding structure or region for any structure (both false positives and masses) in the breast, in this study we have analyzed its accuracy on biopsy-proven masses alone. The location of the mass on the current mammogram was identified by an MQSA-certified radiologist experienced in breast imaging. The radiologist manually identified the corresponding region on the previous mammogram and the nipple location on both the current and the previous mammograms using an interactive image analysis tool on a UNIX workstation. For each current mammogram, the boundary of the mass was manually delineated by the radiologist using an image display program developed in our laboratory. A bounding box enclosing the corresponding object on the previous mammogram was provided by the radiologist for each of the masses. Each mass as well as the corresponding structure on the previous mammogram was rated for its visibility on a scale of 1–10, where the rating of
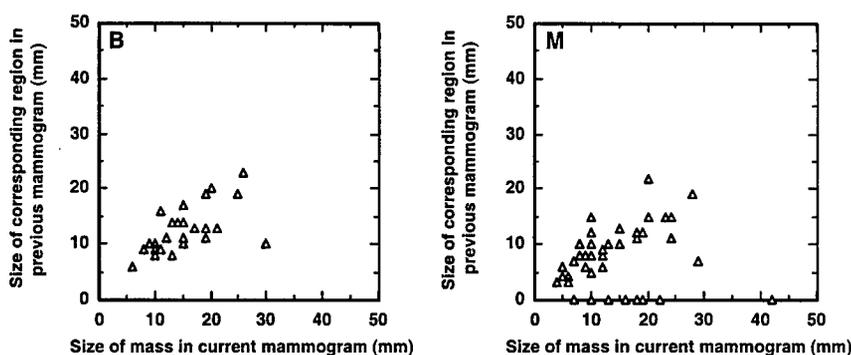
FIG. 4. Distribution of the size of the mass on the current mammogram with respect to the size of the corresponding structure on the previous mammogram as estimated by an experienced breast radiologist for benign (B) and malignant (M) cases in the data set.
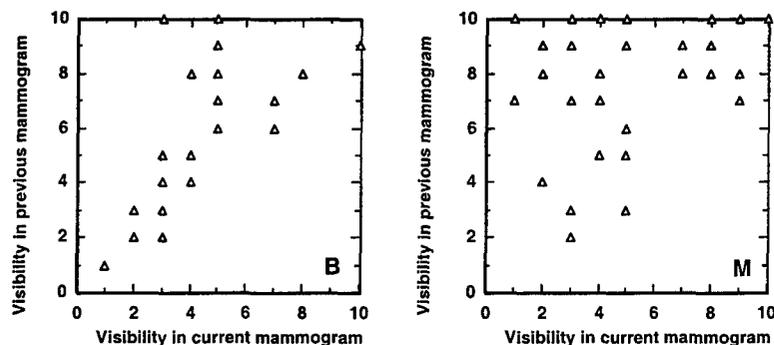
FIG. 5. Distribution of the visibility of the mass on the current mammogram with respect to the visibility of a corresponding structure on the previous mammogram as rated by an experienced breast radiologist for benign (B) and malignant (M) cases. In this rating scale the visibility of the masses decreases from 1 to 10 with 10 being the least visible. The total number of points in these two graphs is less than the total number of mammogram pairs in our database, because mammogram pairs with the same rating appear as a single point.

1 corresponded to the most visible category. The size of the mass on the current mammogram as well as the size of the corresponding structure on the previous mammogram was also provided by the radiologist. For previous mammograms on which the radiologist could not identify a distinct mass, the "mass" size was given a size of 0 mm. The parenchymal density was rated based on the BIRADS lexicon. The distributions of the size and visibility ratings for benign and malignant cases in this data set are shown in Figs. 4 and 5.

C. Evaluation of registration accuracy

The bounding box enclosing the corresponding object on the previous mammogram provided by the radiologist was used as the "ground truth" to evaluate the accuracy of the regional registration technique. We have used two different measures for assessing registration accuracy. The first measure quantifies whether the corresponding region is correctly identified by the registration algorithm. This measure is computed simply as the number of cases in which the estimated centroid location of the mass on the previous mammogram is inside the bounding box provided by the radiologist. The second measure quantifies the error in the estimate of the corresponding region on the previous mammogram and is defined as the Euclidean distance between the estimated centroid of the corresponding region and the center of the bounding box provided by the radiologist. Together these two measures answer the questions: (a) does regional regis-

tration work? (b) how well does the technique perform in matching structures between the current and previous mammograms? In Sec. III we provide the results of regional registration with and without global breast alignment and using both correlation and mutual information as the search criterion in step 3.

III. RESULTS

To provide the reader with a qualitative idea of algorithm performance we first illustrate the intermediate results at various stages of the algorithm. Then the results of each of the three steps of the algorithm are presented with an analysis of the dependence of the performance on various algorithm parameters. Also presented is an analysis of the accuracy of regional registration using the error measures defined in Sec. II C. In the following sections, the term "initial estimate" refers to the estimate of the center of the search region in step 2 of regional registration. The term "final estimate" refers to the outcome of the search procedure adopted in step 3 and represents the overall result of regional registration.

A. Intermediate results of regional registration

Figures 6–8 show an example of the intermediate and final results of applying the regional registration technique to a temporal pair of mammograms. The original digitized mammograms—current and previous—with the automati-
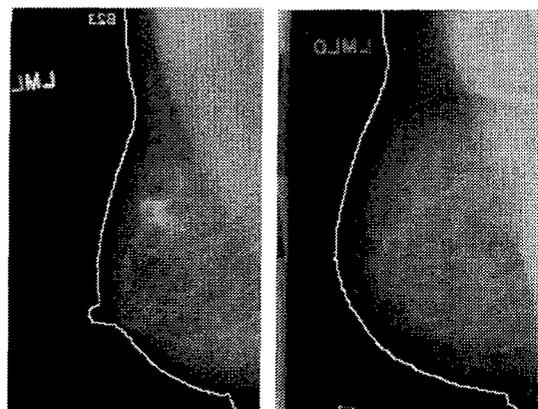
FIG. 6. Left—most recent or current mammogram. Right—previous mammogram. The breast images are superimposed with the breast borders detected by a breast boundary tracking algorithm.
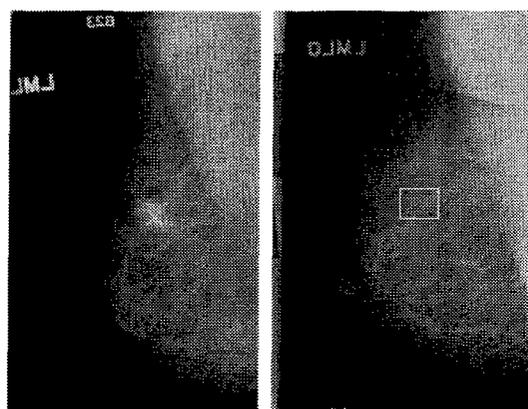
FIG. 7. Left—location of the mass on the current mammogram. Right—radiologist-identified region on previous mammogram corresponding to the mass on the current mammogram.
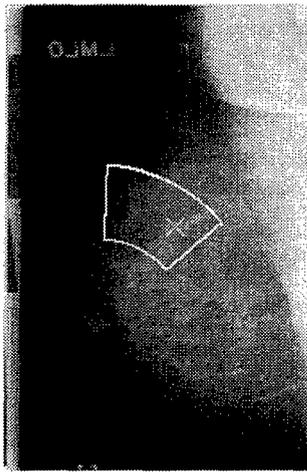
FIG. 8. The fan-shaped search region on the previous mammogram. The initial computer estimate of the centroid location of the region corresponding to the mass is at the center of the search region. The final estimate of the centroid of the corresponding region (indicated by X) is obtained by using the correlation criterion within the fan-shaped search region.

cally tracked breast boundaries superimposed, are shown in Fig. 6. The location of the mass on the current mammogram is shown in Fig. 7 along with the corresponding radiologist-identified region on the previous mammogram. Figure 8 shows the fan-shaped search region on the previous mammogram estimated in step 2 of regional registration. The initial estimate is at the center of this search region which is to be used in step 3 for localization of the corresponding mass. The centroid location of the corresponding object estimated by the algorithm using the correlation measure as the search criterion is also shown in Fig. 8.

B. Initial estimates and search regions

Figure 9 shows histograms of the Euclidean distance between the initial estimate of the centroid location of the corresponding structure on the previous mammogram and the center of the bounding box provided by the radiologist. For the 74 temporal image pairs used in this data set, the average Euclidean distance error of the initial estimate was 10.5 mm (std. dev. 6.4 mm) without the GBA procedure and 9.8 mm (std. dev. 6.0 mm) with the GBA procedure. The overall accuracy was 46% in both cases, i.e., in 34 of the 74 temporal image pairs the initial estimate was inside the ground-truth bounding box. Based on observation of the radial deviation errors and the angular deviation errors (defined in Sec. IV) in Figs. 10 and 11, a search region defined by ϵ

$=0.35+5/r$ rad and $\delta=20$ mm with GBA ($\delta=25$ mm for no GBA), where r is the radial distance from the nipple, was used for the evaluation of the local search criteria used in step 3 of regional registration.

C. Local search criteria and final estimates

Figure 12 shows the histograms of the Euclidean distance errors of the final estimate of the corresponding structure using the correlation measure as the search criterion. Table I summarizes the results along with the average Euclidean distance errors and standard deviations using both the correlation and the mutual information search criteria and with and without the GBA procedure. The average Euclidean distance errors and deviations for the cases where the final estimate is inside the ground-truth region identified by the radiologist and the cases where it is outside are also listed separately. Regional registration incorporating the GBA procedure and using correlation as a search criterion has an accuracy of 85%. In 63 of the 74 temporal image pairs, the final estimate of the location of the corresponding region was inside the radiologist-identified ground-truth region. The use of mutual information as a search criterion yielded an accuracy of 74% (55 out of 74 temporal pairs). The average Euclidean distance error for regional registration incorporating GBA and correlation was 4.7 mm (std. dev. 5.8 mm) for all 74 temporal pairs and 2.8 mm (std. dev. 1.9 mm) in 85% (63/74) of the temporal pairs. Use of mutual information as a search criterion in step 3 results in values of 7.2 mm (std dev. 8.6 mm) and 3.0 mm (std. dev. 2.0 mm), respectively, for the same quantities.

IV. DISCUSSION

A. Initial estimates and search regions

From the histograms of Fig. 9, we observe that the use of the GBA procedure results only in a marginal improvement in the initial estimate, if the Euclidean distance error is the only measure considered. However, the GBA procedure has a significant effect in reducing the size of the search region required for regional registration. In order to compute the required sizes (δ and ϵ in Fig. 3) of the search region, we computed two quantities—the radial distance deviation and the angular deviation—using the initial estimate obtained from step 2 for the 74 temporal image pairs. The radial distance deviation is defined as the absolute difference between $s_1 r$ and r_c, where r_c is the radial distance of the center of the ground-truth region from the nipple location on the pre-
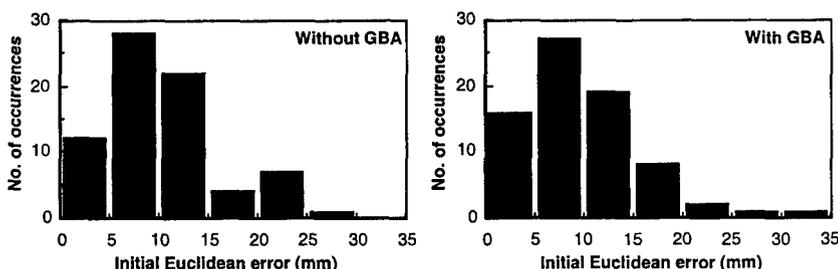
FIG. 9. Histograms of Euclidean distance between the initial estimate of the centroid location of the corresponding object and the center of the radiologist-identified object on the previous mammogram with and without GBA.

FIG. 10. Histograms of radial distance deviation between the initial estimate of the centroid location of the corresponding object and the center of the radiologist-identified object on the previous mammogram with and without GBA.
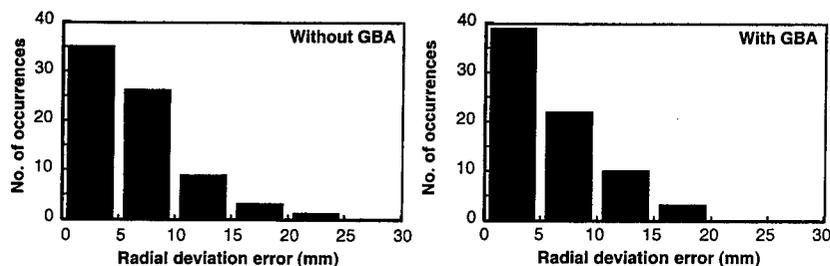
vious mammogram. The histograms of radial distance deviations for the 74 temporal image pairs with and without the GBA procedure are shown in Fig. 10. An important observation is that a δ value of 25 mm is needed to include the centers of the ground-truth structures if the GBA procedure is not used in step 1. The use of the GBA procedure results in a decrease in the value of δ to 20 mm. This decrease helps significantly increase the overall accuracy of the regional registration as discussed below.

In Fig. 11 the angular deviation of the initial estimate is plotted against the radial distance of the centers of the ground-truth regions on the previous mammogram. The angular deviation ϵ is defined as $s_2\theta - \theta_c$ where θ_c is the angle between the nipple-ground-truth center vector and the nipple-centroid axis. In an earlier study[27] using both false positives and masses, we have observed that the value of ϵ needed to include the center of the ground-truth region decreases with distance from the nipple, i.e., increases with

distance from the chest wall. This may be attributed to the increased deformability of the breast tissue closer to the nipple compared to the tissue closer to the chest wall. This indicates that a possible approach to take into account this variability is to incorporate a variable ϵ, one which is inversely proportional to the radial distance r from the nipple. For the data set in this study, we have investigated several forms for this dependence all of which fit under the general model

$$\epsilon = \epsilon_{th} + K/r.$$

Here ϵ_{th} and K are two constants which affect the form of the dependency. Based on our observation of the angular deviations for the entire data set of 74 temporal pairs we have chosen $\epsilon_{th} = 0.35$ rad and $K = 5$ rad-mm. As can be seen from Fig. 11, with these values of ϵ_{th} and K, all of the centers of the ground-truth regions are within the search region. Therefore, a search region defined by $\epsilon = 0.35 + 5/r$ rad, and $\delta = 20$ mm (if GBA was applied) or $\delta = 25$ mm (if GBA was not applied) was used for evaluation of the local search criteria used in step 3 of regional registration.
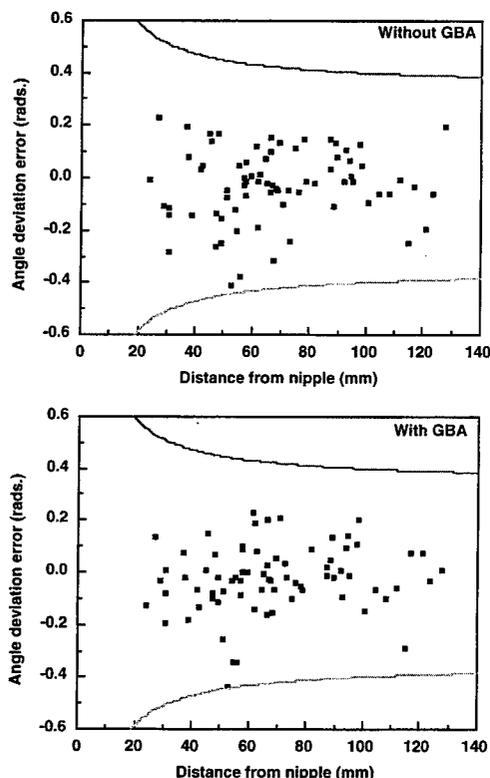
FIG. 11. Angular deviation between the initial estimate of the centroid location of the corresponding object and the center of the radiologist-identified object on the previous mammogram with and without GBA. Also shown are the bounding lines defined using $\epsilon = 0.35 + 5/r$ rad.
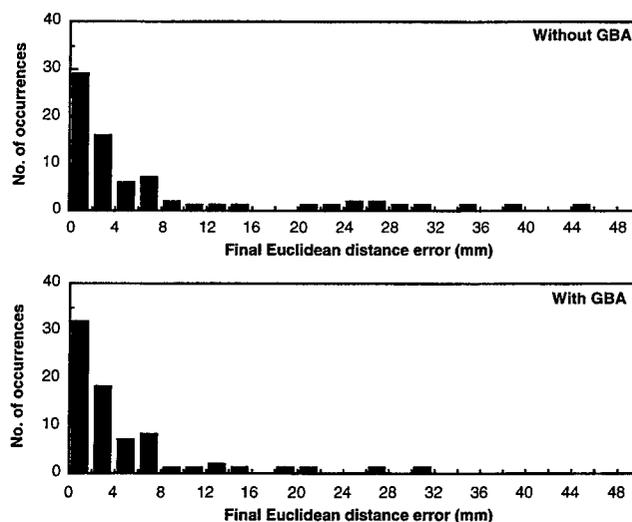
FIG. 12. Histograms of Euclidean distance error for corresponding regions estimated by regional registration using the correlation measure in step 3 with and without GBA. This error is defined as the Euclidean distance between the centroid location of the estimated corresponding region and the center of the radiologist-identified ground-truth corresponding region on the previous mammogram.

TABLE I. Accuracy of regional registration using correlation measure and mutual information measure in step 3 with and without global breast alignment (GBA) and using a 1-pixel-wide background region for the template from the current mammogram. Correct estimates are the cases where the estimated centroid location was within the bounding box of the radiologist-identified object location.

Method	Accuracy	Overall average error (mm)	Average error (mm) for correct estimates	Average error (mm) for incorrect estimates
Correlation without GBA	77% (57/74)	7.4±10.2	2.8±2.0	22.9±11.5
Mutual information without GBA	68% (50/74)	8.8±10.5	3.0±2.0	20.7±11.1
Correlation with GBA	85% (63/74)	4.7±5.8	2.8±1.9	15.7±8.3
Mutual information with GBA	74% (55/74)	7.2±8.6	3.0±2.0	19.4±8.9

B. Local search criteria and final estimates

We have evaluated the use of correlation and mutual information as the local search criteria. From Table I we observe that the GBA procedure results in a higher accuracy irrespective of the search criterion. While the use of mutual information as a search criterion performs reasonably well by itself (74% accuracy with an average error of 7.2 mm) the use of correlation measure was observed to result in more accurate registration. For the images in this data set, the correlation measure outperformed the mutual information measure irrespective of whether the breast centroids were computed with or without the GBA procedure.

A few observations on the 11 cases where the final estimate was outside the radiologist-identified ground-truth corresponding region are in order. In 7 of the 11 cases although the radiologist did provide a region corresponding to the mass on the current mammogram, the corresponding structure on the previous mammogram was very subtle (visibility rating 8 or higher) with indistinct boundaries. The radiologist could only estimate the region where the mass would develop rather than the mass itself, so the truth was uncertain. In one of the remaining 4 cases, the mass was an architectural distortion in the current mammogram. In a second (benign) case the mass shape had changed considerably. Upon consultation of the pathology report, the radiologist concluded that the mass was a benign cyst which had been aspirated in the previous year resulting in a substantial change in its shape. In the third case, the proximity of the mass to the chest wall resulted in it being incompletely imaged in the previous year compared to the current year. In such cases the correlation measure of a neighboring breast structure would tend to be higher than that of the corresponding structure. In the fourth case, an overlap of two vessels was identified as corresponding to the mass on the current mammogram while the region corresponding to the mass was observed to be extremely subtle. In almost all of the 11 cases the proximity

of the corresponding region to a dense structure combined with the subtle nature of the structure on the previous mammogram render the correlation measure ineffective in establishing correspondence. However, in clinical practice, these masses will likely be categorized as a newly developed density. Criteria to distinguish a newly developed density will be investigated in further studies.

C. GBA: Area overlap vs mutual information

For the images used in this study, the result of the GBA procedure based on maximizing the area overlap between the breast regions in the two images of a temporal pair is comparable to that based on maximizing the mutual information. However, our observation is that the mutual information criterion is preferable to the area overlap criterion. The area overlap measure suffers from the drawback that if the breast region in one of the mammograms is uniformly smaller than that in the other, i.e., the breast edge in one is completely within the breast edge in the other, then there is no unique rotation angle at which the area overlap is maximized. Although the range of rotation angles over which local maxima of the area overlap occur is small, the resulting estimate of the rotation angle for GBA may be suboptimal. The use of mutual information, however, results in a single unique rotation angle at which MI is maximized. In any case, as discussed earlier, the use of the GBA procedure before computing the breast centroid results in a reduction in the size of the search region. A smaller search region reduces the likelihood that the mass template is matched to an incorrect structure and, therefore, increases the accuracy and reduces the Euclidean distance error.

D. Template size, scale factors, and computation times

The size of the background region in the gray scale template extracted from the current mammogram affects registration accuracy. For the 74 temporal pairs in this data set, the best performance was observed when a 1-pixel-wide background region was included all around the boundary of the mass template. A 5-pixel-wide background region resulted in a decrease in accuracy and an increase in the average Euclidean distance error. The accuracy progressively decreased and the Euclidean distance error increased with an increase in the size of the background region in the template. Figure 13 shows the distributions of the radial and angular scale factors for the images used in this study. The radial scale factor s_1 ranged from 0.94 to 1.05 for this data set. Use of s_1 reduced the size of the search area by decreasing the required value for δ. The angular scale factor s_2 was very close to 1 in all cases and did not seem to make any major difference for the images in this data set. On a final note the computation time required for regional registration incorporating correlation was on the average 2 s without GBA and 4 s with GBA on a UNIX workstation (DEC AlphaStation 600 series).
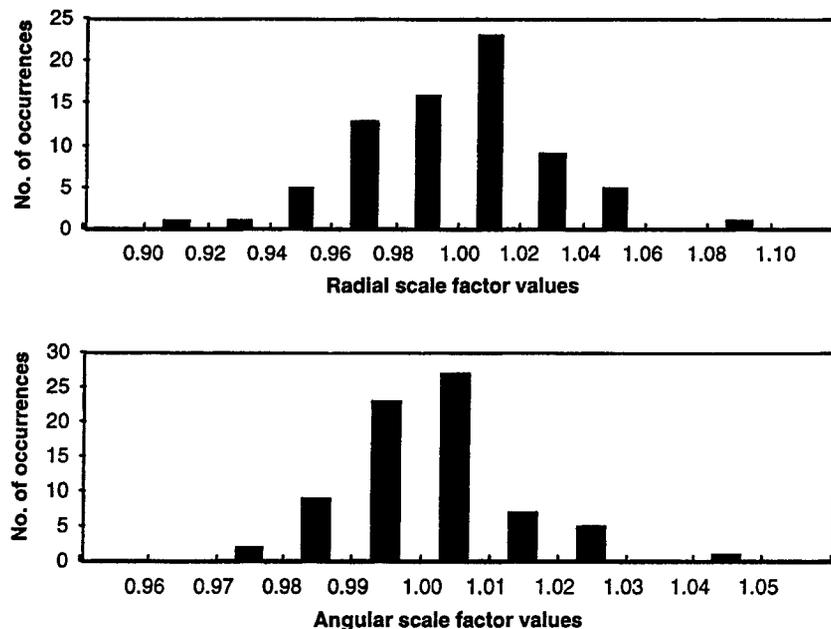
FIG. 13. Histograms of the radial scale factor and the angular scale factor for 74 temporal image pairs. The radial scale factor s_1 is estimated as the ratio of the nipple–centroid distances on the previous and current images. The angular scale factor s_2 is estimated as the ratio of the angular width of the breast on the previous image at radius $s_1 r$ to that on the current image at radius r.

V. CONCLUSIONS

Radiologists are interested in determining any local changes in breast tissue over time which may indicate a developing cancer. We have developed a novel regional registration technique for temporal registration of mammograms. This technique could become an important component of a CAD scheme for mammographic analysis. Unlike other techniques found in the literature, our regional registration technique does not depend on the identification of landmark structures or control points on the mammograms. It is based on a search technique that many radiologists use and has proven to be successful in mammographic interpretation. After corresponding objects are found, they can be analyzed for interval changes in a CAD scheme. Our preliminary results indicate that the regional registration technique is promising in identifying corresponding regions from temporal mammographic pairs. In 85% (63/74) of the cases the regional registration technique correctly identified the corresponding region in the previous mammogram. For these 63 cases, it is highly encouraging to note that the estimated location of the region corresponding to the mass in the current mammogram was less than 3 mm on the average from radiologist-identified corresponding locations.

Areas for future work include the development of an automated technique for identifying the nipple location on the mammograms, investigation of other local search criteria such as Fourier descriptors and shape-invariant moments to be used in the fan-shaped search region, adaptive methods for determining the size of the search region, criteria for identifying newly developed densities, application of regional registration to false positives as well as masses, and studies with a large data set to investigate the robustness of the regional registration technique. It may be noted that the regional registration technique may also be applicable to other related registration problems, such as the registration of left and right mammograms.

ACKNOWLEDGMENTS

[a])Electronic mail: chanhp@umich.edu

[1] S. A. Feig and R. E. Hendrick, "Risk, Benefit, and Controversies in Mammographic Screening," in *Syllabus: A Categorical Course in Physics Technical Aspects of Breast Imaging*, edited by A. G. Haus and M. J. Yaffe (Radiological Society of North America, Oak Brook, IL, 1993).

[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 (N.Y.) **74**, 301–310 (1994).

[3] Y. Wu, K. Doi, M. L. Geiger, and R. M. Nishikawa, "Computerized detection of clustered microcalcifications in digital mammograms: Applications of artificial neural networks," Med. Phys. **19**, 555–560 (1992).

[4] H. P. Chan *et al.*, "Improvement in radiologists' detection of clustered microcalcifications: The potential of computer-aided diagnosis," Invest. Radiol. **25**, 1102–1110 (1990).

[5] S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," IEEE Trans. Med. Imaging **8**, 377–386 (1989).

[6] J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computerized image analysis," IEEE Trans. Med. Imaging **12**, 664–669 (1993).

[7] W. P. Kegelmeyer, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," Radiology **191**, 331–337 (1994).

[8] H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, M. M. Goodsitt, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space," Phys. Med. Biol. **40**, 857–876 (1995).

[9] L. W. Bassett, B. Shayestehfar, and I. Hirbawi, "Obtaining previous mammograms for comparison: Usefulness and costs," Am. J. Roentgenol. **163**, 1083–1086 (1994).

[10] E. A. Sickles, "Periodic mammographic follow-up of probably benign lesions: Results in 3183 consecutive cases," Radiology **179**, 463–468 (1991).

[11] F. F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images," Med. Phys. **18**, 955–963 (1991).

[12] F. F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique," Med. Phys. **21**, 445–452 (1994).

[13] M. Sallam and K. Bowyer, "Detecting abnormal densities in mammograms by comparison with previous screenings," in *Digital Mammography '96*, edited by K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt (Elsevier, Amsterdam, 1996).

[14] D. Brzakovic, N. Vujovic, M. Neskovic, P. Brzakovic, and K. Fogarty, "Mammogram analysis by comparison with previous screenings" in Ref. 13.

[15] N. Vujovic, D. Brzakovic, and K. Fogarty, "Detection of cancerous changes in mammograms using intensity and texture measures," Proc. SPIE **2434**, 37–47 (1995).

[16] N. Vujovic, P. Bakic, and D. Brzakovic, "Detection of potentially cancerous signs by mammogram followup," in Ref. 13.

[17] N. Vujovic and D. Brzakovic, "Establishing the correspondence between control points in pairs of mammographic images," IEEE Trans. Image Process. **6**, 1388–1399 (1997).

[18] W. K. Zouras, M. L. Giger, P. Lu, D. E. Wolverton, C. J. Vyborny, and K. Doi, "Investigation of a temporal subtraction scheme for computerized detection of breast masses in mammograms," in Ref. 13.

[19] A. R. Morton, "Design of an x-ray beam equalization filter for mammographic imaging," M.S. thesis, Department of Environmental and Industrial Health, University of Michigan, 1996.

[20] A. R. Morton, H. P. Chan, and M. M. Goodsitt, "Automated model-guided breast segmentation algorithm," Med. Phys. **23**, 1107–1108 (1996).

[21] A. J. Mendez, P. G. Tahoces, M. J. Lado, M. Souto, J. L. Correa, and J. J. Vidal, "Automatic detection of breast border and nipple in digital mammograms," Comput. Methods Programs Biomed. **49**, 253–262 (1996).

[22] R. Chandrasekhar and Y. Attikiouzel, "A simple method for automatically locating the nipple on mammograms," IEEE Trans. Med. Imaging **16**, 483–494 (1997).

[23] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," IEEE Trans. Med. Imaging **16**, 187–198 (1997).

[24] A. Maintz, E. Meijering, and M. Viergever, "General multimodal elastic registration based on mutual information," Proc. SPIE **3338**, 144–154 (1998).

[25] N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and tissue classification," Med. Phys. **23**, 1685–1696 (1996).

[26] S. Sanjay-Gopal, H. P. Chan, B. Sahiner, N. Petrick, T. Wilson, and M. Helvie, "Evaluation of interval change in mammographic features for computerized classification of malignant and benign masses," Radiology **205(P)**, 216 (1997).

[27] S. Sanjay-Gopal, H. P. Chan, N. Petrick, T. Wilson, B. Sahiner, M. Helvie, and M. Goodsitt, "A regional registration technique for automated analysis of interval changes of breast lesions," Proc. SPIE **3338**, 118–131 (1998).

Classification of Malignant and Benign Masses Based on Hybrid ART2LDA Approach

Lubomir Hadjiiski,* *Member, IEEE*, Berkman Sahiner, *Member, IEEE*,
Heang-Ping Chan, Nicholas Petrick, *Member, IEEE*, and Mark Helvie

Abstract—A new type of classifier combining an unsupervised and a supervised model was designed and applied to classification of malignant and benign masses on mammograms. The unsupervised model was based on an adaptive resonance theory (ART2) network which clustered the masses into a number of separate classes. The classes were divided into two types: one containing only malignant masses and the other containing a mix of malignant and benign masses. The masses from the malignant classes were classified by ART2. The masses from the mixed classes were input to a supervised linear discriminant classifier (LDA). In this way, some malignant masses were separated and classified by ART2 and the less distinguishable benign and malignant masses were classified by LDA. For the evaluation of classifier performance, 348 regions of interest (ROI's) containing biopsy proven masses (169 benign and 179 malignant) were used. Ten different partitions of training and test groups were randomly generated using an average of 73% of ROI's for training and 27% for testing. Classifier design, including feature selection and weight optimization, was performed with the training group. The test group was kept independent of the training group. The performance of the hybrid classifier was compared to that of an LDA classifier alone and a backpropagation neural network (BPN). Receiver operating characteristics (ROC) analysis was used to evaluate the accuracy of the classifiers. The average area under the ROC curve (A_z) for the hybrid classifier was 0.81 as compared to 0.78 for the LDA and 0.80 for the BPN. The partial areas above a true positive fraction of 0.9 were 0.34, 0.27 and 0.31 for the hybrid, the LDA and the BPN classifier, respectively. These results indicate that the hybrid classifier is a promising approach for improving the accuracy of classification in CAD applications.

Index Terms— Computer-aided diagnosis, hybrid classifier, mammography, neural networks.

I. INTRODUCTION

MAMMOGRAPHY is the most effective method for detection of early breast cancer [1]. However, the specificity for classification of malignant and benign lesions from mammographic images is relatively low. Clinical studies

have shown that the positive predictive value (i.e., ratio of the number of breast cancers found to the total number of biopsies) is only 15% to 30% [2]–[4]. It is important to increase the positive predictive value without reducing the sensitivity of breast cancer detection. Computer-aided diagnosis (CAD) has the potential to increase the diagnostic accuracy by reducing the false-negative rate while increasing the positive predictive values of mammographic abnormalities.

Classifier design is an important step in the development of a CAD system. A classifier has to be able to merge the available input feature information and make a correct evaluation. Commonly used classifiers for CAD include linear discriminants (LDA) [5], [6] and backpropagation neural networks (BPN) [7]–[9] which have been shown to perform well in lesion classification problems [10]–[22]. These classifiers are generally designed by supervised training. However, these types of classifiers have limitations dealing with the nonlinearities in the data (in case of LDA) and in generalizability when a limited number of training samples are available (especially BPN). Another classification approach is based on unsupervised classifiers, which cluster the data into different classes based on the similarities in the properties of the input feature vectors. Therefore, unsupervised classifiers can be used to analyze the similarities within the data. However, it is difficult to use them as a discriminatory classifier [29], [30]. They also have limited generalizability when the training sample set is small.

We propose here a hybrid unsupervised/supervised structure to improve classification performance. The design of this structure was inspired by neural information processing principles such as self organization, decentralization and generalization. It combines the adaptive resonance theory network (ART2) [26], [27] and the LDA classifier as a cascade system (ART2LDA). The self-organizing unsupervised ART2 network automatically decomposes the input samples into classes with different properties. The ART2 network has been found to perform better compared to conventional clustering techniques in terms of learning speed and discriminatory resolution for the detection of rare events in many classification tasks [28]–[30]. The supervised LDA then classifies the samples belonging to a subset of classes that have greater similarities. By improving the homogeneity of the samples, the classifier designed for the subset of classes may be more robust.

The ART2LDA design implements both structural and data decomposition. Decomposition is a powerful approach that can reduce the complexity of a problem. Both structural decom-

position and data decomposition can improve classification accuracy [23] as well as model accuracy [24]. However, decomposition can also reduce the prediction accuracy due to overfitting the training data. We will demonstrate in this paper that the proposed hybrid structure can reduce the overfitting problem and improve the prediction capabilities of the system. The performance of the hybrid ART2LDA classifier will be compared with those of an LDA alone or a BPN classifier.

The rest of the paper is organized as follows. In Section II the ART2 unsupervised network is described. A hybrid ART2LDA classifier is introduced in Section III. Section IV describes the data set used in this study. The results are presented in Section V. Section VI contains discussion of these results. Finally, Section VII concludes this investigation.

Fig. 1.   Structure of the ART2 network.

II. ART2 UNSUPERVISED NEURAL NETWORK

The ART2 is a self-organizing system that can simulate human pattern recognition. ART2 was first described by Grossberg [25] and a series of further improvements were carried out by Carpenter, Grossberg, and coworkers [26]–[28]. The ART2 network clusters the data into different classes based on the properties of the input feature vectors. The members within a class have similar properties. The process of ART2 network learning is a balance between the plasticity and stability dilemma. Plasticity is the ability of the system to discover and remember important new feature patterns. Stability is the ability of the system to remain unchanged when already known feature patterns with noise are input to the system. The balance between plasticity and stability for the ART2 training algorithm allows fast learning [28], i.e., rare events can be memorized with a small number of training iterations without forgetting previous events. The more conventional training algorithms, such as back propagation [7]–[9], perform slow learning, i.e., they tend to average over occurrences of similar events and require many training iterations.

The structure of the ART2 system is shown in Fig. 1. It consists of two parts: the ART2 network and the learning stage. Suppose that there are n input features x_i ($i = 1, \cdots, n$) and k classes in the ART2 network. When a new vector is presented to the input of the ART2 network, an activation value p_j for class j is calculated as

$$p_j = \sum_{i=1}^{n} x_i w_{ij}, \qquad j = 1, \cdots, k \qquad (1)$$

where w_{ij} is the connection weight between input i and class j. The activation value is a measure of the membership of the particular input feature vector to class j. The higher the value p_j is, the better the input vector matches class j. The maximum value p_r is selected from all p_j ($j = 1, \cdots, k$) to find the best class match. Furthermore, in order to balance the contribution to the activation value from all feature components, the input feature values applied to the ART2 system are scaled between zero and one [30]. This normalization will allow detection of similar feature patterns even when the magnitudes of the input feature components are very different.

The learning stage of the ART2 system can influence the weights of the selected class or the complete ART2 network
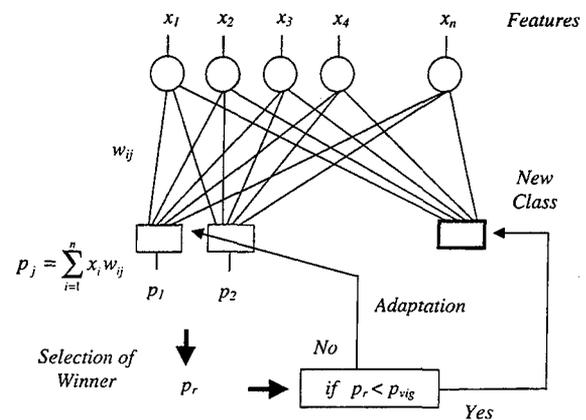
structure by adding a new class. An additional parameter, the vigilance, is used to determine the type of learning [26]. The vigilance parameter p_{vig} is a threshold value that is compared to the maximum activation value p_r. If p_r is larger than p_{vig} then the input vector is considered to belong to class r. The adaptation of the weights connected with class r is performed as follows:

$$w_{ir}^{new} = w_{ir}^{old} + \eta(x_i - w_{ir}^{old}), \qquad \text{for } i = 1, \cdots, n \qquad (2)$$

where η is a learning rate. The adaptation of the class r weights (2), aims at maximization of the p_r value for the particular input vector. In an iterative manner the weights are adjusted so that the activation values produced for similar input vectors will be maximum only for the class to which they belong and these maximum activation values will be higher than p_{vig}.

If the maximum activation value p_r is smaller than p_{vig}, it is an indication that a novelty has appeared and a new class will be added to the ART2 structure. The new weights connecting the input with the new class $(k + 1)$ are initialized with the scaled input feature values of this novelty. In such a way, the activation value p_{k+1} will be maximum $(p_r = p_{k+1})$ higher than p_{vig} when computed for this novelty in further training iterations. The value of the vigilance parameter p_{vig} determines the resolution of ART2. It can be chosen in the range between zero and one. In the case that p_{vig} is relatively small, only very different input feature vectors will be distinguished and separated in different classes. If p_{vig} is relatively large, the input feature vectors that are more similar will be separated into different classes. The value of p_{vig} is selected differently depending on the particular application.

III. ART2LDA CLASSIFIER

Despite the good performance of ART2 for efficient clustering and detection of novelties, the fast learning approach can cause problems associated with the generalization capability of the system and the correct classification of unknown cases. Supervised classifiers such as linear discriminants or backpropagation neural network classifiers can have better generalization capability than ART2, because they are trained by averaging over similar event occurrences. However, the learning process in these traditional learning algorithms tends

to erase the memory of previous expert knowledge when a new type of expertise is being learned. Therefore, these classifiers do not have as good an ability to correctly classify rare events as ART2 [28], [29].

In order to improve the accuracy and generalization of a classifier, we propose to design a hybrid classifier that combines the unsupervised ART2 network and a supervised LDA classifier. This hybrid classifier (ART2LDA) utilizes the good resolution capability of ART2 and the good generalization capability of LDA. The ART2 first analyzes the similarity of the sample population and identifies a subpopulation that may be separated from the main population. This will improve the performance of the second-stage LDA if the subpopulation causes the sample population to deviate from multivariate normal distributions for which LDA is an optimal classifier. Therefore, the ART2 serves as a screening tool to improve the homogeneity of the sample distributions by classifying outlying samples into separate classes.

The ART2LDA hybrid classifier can be described as

$$y_{AL} = g(f_2(x))f_1(x) + 1 - g(f_2(x)) \tag{3}$$

where x is the input vector, $f_1(\cdot)$ is the LDA classifier, $f_2(\cdot)$ is the ART2 classifier, and $g(\cdot)$ is a binary membership function, which labels the classes identified by ART2 to be one of the two types: malignant class or mixed class. A particular class is defined as malignant if it contains only malignant members. It is defined as mixed if it contains both malignant and benign members. The membership function is defined as follows:

$$g(c) = \begin{cases} 0, & \text{if } c \text{ is a malignant class} \\ 1, & \text{if } c \text{ is a mixed class.} \end{cases} \tag{4}$$

The type of a given class is determined based on ART2 classification of the training data set.

The structure of the ART2LDA classifier is shown in Fig. 2. The ART2 classifies the input sample x into either a malignant or a mixed class. Depending on the class type the function $g(\cdot)$ determines whether the LDA classifier will be used. If x is classified into a mixed class, the final classification will be obtained based on the LDA classifier. However, if x is classified by ART2 into a malignant class, then the mass will be considered malignant, without using the LDA classifier. Therefore, in the ART2LDA structure, the ART2 is used both as a classifier and a supervisor. This can be seen in (3). The first term in (3), $g(f_2(x))f_1(x)$, is the LDA classifier multiplied by the ART2 control part $g(f_2(x))$. The second term in (3), $(1 - g(f_2(x)))$, gives the classification result of the ART2 stage. If $f_2(x)$ is a malignant class, then $g(f_2(x)) = 0$, the LDA stage is eliminated, and the classifier output y_{AL} is equal to 1. On the other hand, if $f_2(x)$ is a mixed class, then $g(f_2(x)) = 1$, the ART2 term is eliminated, and the final classification is determined by the LDA classifier $(y_{AL} = f_1(x))$.

IV. METHODS

A. Data Set

The mammograms used in this study were randomly selected from the files of patients who had undergone biopsies
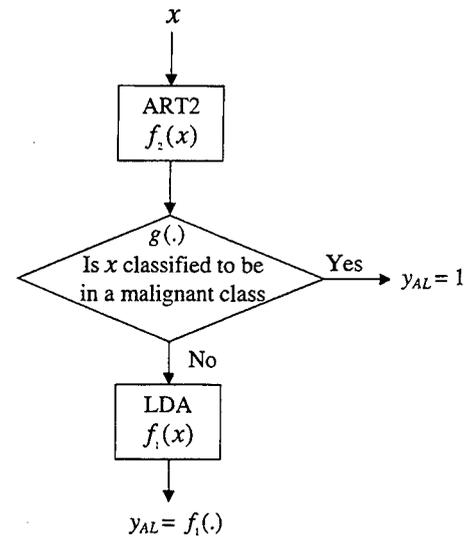
Fig. 2. Structure of the ART2LDA classifier.

at the University of Michigan. The criterion for inclusion of a mammogram in the data set was that the mammogram contained a biopsy-proven mass. The data set contained 348 mammograms with a mixture of benign ($n = 169$) and malignant ($n = 179$) masses. On each mammogram, a region of interest (ROI) containing the mass was identified by a radiologist experienced in breast imaging. The visibility of the masses was rated by the radiologist on a scale of 1 to 10, where the rating of 1 corresponds to the most visible category. The distributions of the visibility rating for both the malignant and benign masses are shown in Fig. 3. The visibility ranged from subtle to obvious for both types of masses. It can be observed that the benign masses tend to be more obvious than the malignant ones. Additionally the likelihood of malignancy for each mass was estimated based on its mammographic appearance. The radiologist rated the likelihood of malignancy on a scale of 1 to 10, where 1 indicated a mass with the most benign appearance. The distribution of the malignancy rating of the masses is shown in Fig. 4.

The data set can be considered as representative of the patient population that is sent for biopsy under current clinical criteria. Some characteristics of many malignant and benign masses can be visually distinguished by radiologists. However, there is also a nonnegligible fraction of malignant masses that are very similar to benign masses (the low malignancy rating region in Fig. 4). The estimated likelihood of malignancy of malignant and benign masses that are sent for biopsy basically overlaps over the entire range. This is consistent with the fact that in order not to miss malignant masses radiologists must recommend biopsy for even very low suspicion lesions.

Three hundred and five of the mammograms were digitized with a LUMISYS DIS-1000 laser scanner at a pixel resolution of 100 μm $\times$ 100 μm and 4096 gray levels. The digitizer was calibrated so that gray level values were linearly and inversely proportional to the optical density (OD) within the range of 0.1 to 2.8 OD units, with a slope of -0.001 OD/pixel value. Outside this range, the slope of the calibration curve decreased gradually. The OD range of the digitizer was 0
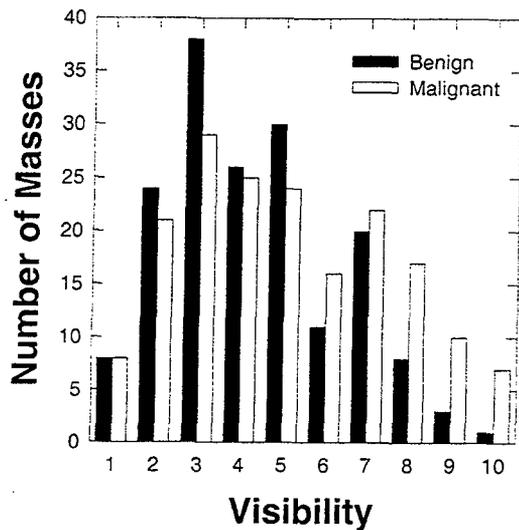
Fig. 3. The distribution of the visibility ranking of the masses in the dataset. The ranking was performed by an experienced breast radiologist (1: very obvious, 10: very subtle).
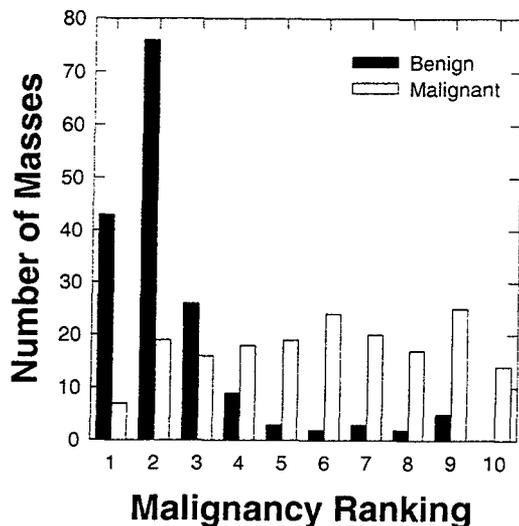
Fig. 4. The distribution of the malignancy ranking of the masses in the dataset. The ranking was performed by an experienced breast radiologist (1: very likely benign, 10: very likely malignant).

to 3.5. The remaining 43 mammograms were digitized with a LUMISCAN 85 laser scanner at a pixel resolution of 50 μm × 50 μm and 4096 gray levels. The digitizer was calibrated so that gray level values were linearly and inversely proportional to the OD within the range of 0 to 4 OD units, with a slope of −0.001 OD/pixel value. In order to process the mammograms digitized with these two different digitizers, the images digitized with LUMISCAN 85 digitizer were averaged with a 2 × 2 box filter and subsampled by a factor of two, resulting in 100 μm images.

In order to validate the prediction abilities of the classifier, the data set was partitioned randomly into training and test subsets on a 3:1 ratio, under the constraints that both the malignant and the benign samples were split with the 3:1 ratio and that the images from the same patient were grouped into the same (training or test) subset. These constraints caused the subsets to deviate from an exact 3:1 ratio. The data set was repartitioned randomly ten times. On average, 73% of the samples were grouped into the training set and 27% into the test set. The training and test results from the ten partitions were averaged to reduce their variability.

B. Feature Extraction

A rectangular ROI was defined to include the radiologist-identified mass with an additional surrounding breast tissue region of at least 40 pixels wide from any point of the mass border. A fully automated method was then used for segmentation of the mass from the breast tissue background within the ROI. The rubber band straightening transform (RBST) was previously developed [12] to map a band of pixels surrounding the mass onto the Cartesian plane (a rectangular region). In the transformed image, the border of mass appears approximately as a horizontal edge and spiculations appear approximately as vertical lines. The transformation of the radially oriented textures surrounding the mass margin to a more uniform orientation facilitates the extraction of texture features.

The texture features used in this study were calculated from spatial gray-level dependence (SGLD) matrices [10]–[12], [31], and run-length statistics (RLS) matrices [32] computed from the RBST images. The (i, j)th element of the SGLD matrix is the joint probability that gray levels i and j occur in a direction at a distance of θ pixels apart in an image. Based on our previous studies [10], a bit depth of eight was used in the SGLD matrix construction, i.e., the four least significant bits of the 12-bit pixel values were discarded. Thirteen texture measures, including correlation, energy, difference entropy, inverse difference moment, entropy, sum average, sum entropy, inertia, sum variance, difference average, difference variance, and two types of information measure of correlation were used. These measures were extracted from each SGLD matrix at ten different pixel pair distances ($d = 1, 2, 3, 4, 6, 8, 10, 12, 16$ and 20) and in four directions ($0°$, $45°$, $90°$, and $135°$). Therefore, a total of 520 SGLD features were calculated for each image. The definitions of the texture measures are given in the literature [10]–[12], [31]. These features contain information about image characteristics such as homogeneity, contrast, and the complexity of the image.

RLS texture features were extracted from the vertical and horizontal gradient magnitude images, which were obtained by filtering the RBST image with horizontally or vertically oriented Sobel filters and computing the absolute gradient value of the filtered image. A gray level run is a set of consecutive, collinear pixels in a given direction which have the same gray level value. The run length is the number of pixels in a run [32]. The RLS matrix describes the run length statistics for each gray level in the image. The (i, j)th element of the RLS matrix is the number of times that the gray level i in the image possesses a run length of j in a given direction. In our previous study, it was found experimentally that a bit depth of five in the RLS matrix computation could provide good texture characteristics [12].

Five texture measures, namely, short run emphasis, long run emphasis, gray level nonuniformity, run length nonuniformity,

and run percentage were extracted from the vertical and horizontal gradient images in two directions, $\theta = 0°$ and $\theta = 90°$. Therefore, a total of 20 RLS features were calculated for each ROI. The formal definition of the RLS feature measures can be found in [32].

A total of 540 features (520 SGLD and 20 RLS) were therefore extracted from each ROI.

C. Feature Selection

In order to reduce the number of the features and to obtain the best feature set to design a good classifier, feature selection with stepwise linear discriminant analysis [33] was applied. At each step of the stepwise selection procedure one feature is entered or removed from the feature pool by analyzing its effect on the selection criterion. In this study, the Wilks' lambda (the ratio of within-group sum of squares to the total sum of squares [34]) was used as a selection criterion. The optimization procedure used a threshold F_{in} for feature entry and a threshold F_{out} for feature removal. On a feature entry step, the features not yet selected are entered into the selected feature pool one at a time, the significance of the change in the Wilks' lambda caused by this feature is estimated based on F statistics. The feature with the highest significance is entered into the feature pool if its significance is higher than F_{in}. On a feature removal step, the features which have already been selected are analyzed one at a time from the selected feature pool and the significance of the change in the Wilks' lambda is estimated. The feature with the least significance is removed from the selected feature pool if the significance is less than F_{out}. Since the appropriate values of F_{in} and F_{out} are not known *a priori*, we examined a range of F_{in} and F_{out} values and chose the appropriate thresholds in such a way that a minimum number of features were selected to achieve a high accuracy of classification by LDA for the training sets. More details about the stepwise linear discriminant analysis and its application to CAD can be found in [10]–[12].

D. Performance Analysis

To evaluate the classifier performance, the training and test discriminant scores were analyzed using receiver operating characteristic (ROC) methodology [35]. The discriminant scores of the malignant and benign masses were used as decision variables in the LABROC1 program [36], which fit a binormal ROC curve based on maximum likelihood estimation. The classification accuracy was evaluated as the area under the ROC curve, A_z. For the ART2LDA classifier, the discriminant scores of all case samples classified in the two stages are combined. All masses classified into the malignant group by the ART2 stage were assigned a constant positive discriminant score higher than or equal to the most malignant discriminant score obtained from the LDA stage .

The performance of ART2LDA was also assessed by estimation of the partial area index $(A_z^{(0.9)})$ and compared with the corresponding performance index of the LDA and BPN classifiers. The partial area index $(A_z^{(0.9)})$ is defined as the area that lies under the ROC curve but above a sensitivity threshold of 0.9 (TPF$_0$ = 0.9) normalized to the total area above TPF$_0$,

TABLE I
NUMBER OF SELECTED FEATURES FOR THE TEN DATA GROUPS WITH THE CORRESPONDING F_{IN} AND F_{OUT} PARAMETERS

Data Group No.	Number of selected features	F_{in}	F_{out}
1	12	1.8	1.6
2	15	2.4	2.2
3	13	2.4	2.2
4	18	2.4	2.2
5	14	2.4	2.2
6	14	2.1	1.8
7	13	2.4	2.2
8	18	1.8	1.6
9	14	2.4	2.2
10	14	2.4	2.2

(1-TPF$_0$). The partial $A_z^{(0.9)}$ indicates the performance of the classifier in the high-sensitivity (low false negative) region which is most important for clinical cancer detection task. In addition, the performance of the LDA stage of the ART2LDA classifier was evaluated by the estimation of the area under the ROC curve, denoted as A_z (LDA), for the case samples passed onto the LDA classifier.

V. RESULTS

In this section the ART2LDA classification results for malignant and benign masses will be presented and compared with those of the LDA or BPN classifiers. The important point in this study is the fact that the test subset is truly independent of the training subset. Only the training subset is used for feature selection and classifier training, and only the test subset is used for classifier validation. In order to validate the prediction abilities of the classifier, ten different partitions of the training and test sets were used. A different ART2LDA classifier was trained using each training set and the corresponding set of selected features. The classification result was estimated as the average performance for the ten partitions.

For a given partition of training and test sets, feature selection was performed based on the training set alone. The feature selection results for the ten different training groups are shown in Table I. The average number of selected features was 14. An average of two RLS features and twelve SGLD features were selected for each of the training sets which represented 10% of all RLS features and 2.3% of all SGLD features, respectively. Both types of features (RLS and SGLD) are necessary in order to obtain good classification. The most often selected RLS features for the ten training sets were: horizontal short run emphasis (four times), horizontal long run emphasis (six times), vertical run length nonuniformity (three times), horizontal run length nonuniformity (three times). The most often selected SGLD texture measures for the ten training sets were: inverse difference moment (eight times), information measure of correlations one and two (19 times), difference average (nine times), and correlation (ten times). For a given texture measure, features at different angles or distances may be selected, but these features are usually highly correlated so
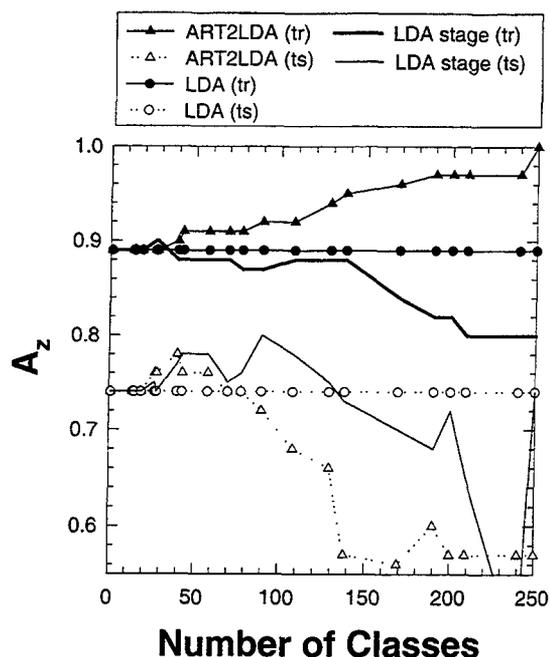
Fig. 5. ART2LDA and LDA classification results for training and test sets from data group three as a function of the generated number of classes. Additionally the results for the LDA stage from the ART2LDA classifier are plotted.

Fig. 6. ART2LDA and LDA classification results for training and test sets from data group one as a function of the generated number of classes. Additionally the results for the LDA stage from the ART2LDA classifier are plotted.

that they can be considered to be similar and counted together as described above.

A. ART2LDA Classification Results

For the ART2LDA classifier, the number of selected features determines the dimensionality of the input vector of the ART2 classifier and the dimensionality of the LDA classifier. By applying different values for the vigilance parameter, ART2 classifiers with different number of classes were obtained. In this study, the vigilance parameter p_{vig} was varied from 0.9 to 0.99, resulting in a range of 10 to 240 classes. The overall performance of the ART2LDA classifier was evaluated for different numbers of ART2 classes because different subset of the samples were separated and classified by ART2 when p_{vig} was varied. In Fig. 5, the classification results for the ART2LDA are compared to the results from LDA alone for the training and test set partition three. The classification accuracy, A_z, was plotted as a function of the number of ART2 classes. For this training and test set partition, when the number of classes was between 20 and 60, the ART2LDA classifier improved the classification accuracy for the test set in comparison to LDA. As the number of classes increased to greater than 60, the A_z value increased for the training data set, but decreased for the test data set and was lower than that of the LDA alone. The two solid lines in Fig. 5 show the A_z values for the LDA stage in the ART2LDA classifier for both the training and test sets. It can be observed that the test A_z for the LDA stage is higher than the A_z for the LDA classifier alone, but not as high as A_z obtained by ART2LDA when the number of classes is small.

In Fig. 6 the classification results of LDA and ART2LDA for the partition one training and test sets are shown. In this case it appeared that in the test set there were two large malignant outliers which degraded the LDA performance. Only 15 classes at the ART2 stage in the ART2LDA was enough to cluster the outliers into a separate malignant class and to improve the performance of the LDA stage and the overall result. The rest of the outliers required more ART2 classes before they were clustered into separate classes and correctly classified as malignant. This is the reason for the similar behavior of the classifiers for partitions three and one in the range of 40 to 70 classes as seen in Figs. 5 and 6. When the number of classes was less than 70, the test A_z for the LDA stage (A_z(LDA)) was higher than the LDA alone, but not as high as the A_z for ART2LDA with less than 30 classes (Fig. 6). The best A_z values for the test data sets of the ten training and test partitions are presented in Table II and Fig. 7. The ART2LDA classifier achieved higher A_z values than the LDA alone in nine of the ten partitions. The average A_z is 0.81 for ART2LDA and 0.78 for LDA alone. The standard deviations of the A_z values for the ten groups range from 0.03 to 0.05 for the ART2LDA classifier and from 0.04 to 0.05 for the LDA classifier.
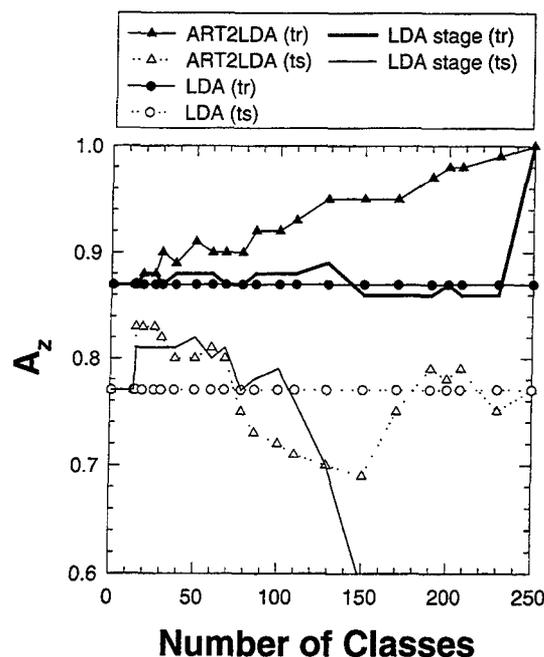
The performance of ART2LDA was also assessed by estimation of the partial area under the ROC curve $A_z^{(0.9)}$ at a TPF higher than 0.9. The results are presented in Table III and Fig. 7. In the lower part of Fig. 7, the $A_z^{(0.9)}$ values of the test set for the corresponding ten partitions of training and test sets are presented. The average test $A_z^{(0.9)}$ value is 0.34 for the ART2LDA and 0.27 for LDA. For nine of the ten partitions, the $A_z^{(0.9)}$ value was improved at the high-sensitivity operating region (TPF > 0.9) of the ROC curve.

The classifier performance was also evaluated when the ART2LDA classifiers were designed using a fixed number

TABLE II
CLASSIFIERS PERFORMANCE FOR THE TEN TEST SETS. THE A_z
VALUES REPRESENT THE TOTAL AREA UNDER ROC CURVE

Data Group No.	LDA	ART2LDA	BPN	ART2LDA(1)
1	0.77	0.83	0.85	0.80
2	0.78	0.80	0.82	0.77
3	0.74	0.78	0.77	0.78
4	0.77	0.77	0.75	0.77
5	0.77	0.78	0.76	0.77
6	0.80	0.83	0.82	0.81
7	0.80	0.81	0.82	0.77
8	0.77	0.80	0.74	0.75
9	0.77	0.80	0.81	0.80
10	0.86	0.89	0.84	0.89
Mean	0.78	0.81	0.80	0.79
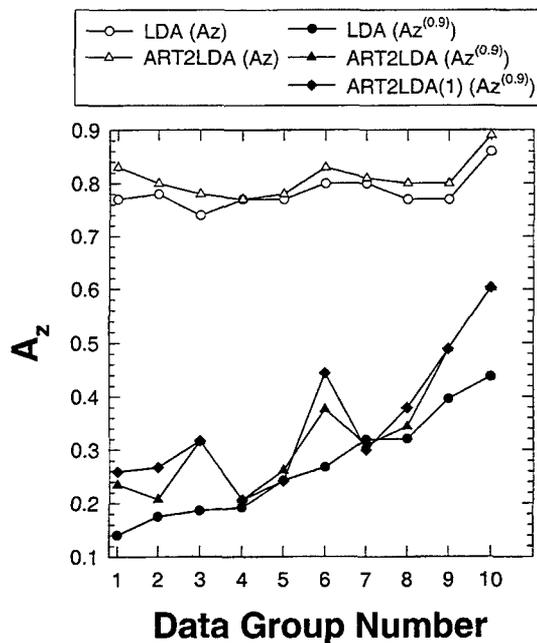
Fig. 7. Average A_z classification results for the 10 test sets. The top graphs represent the ART2LDA and LDA A_z values for the total area under the ROC curve. The bottom graphs represent the ART2LDA, ART2LDA(1) and LDA A_z values for the partial area of the ROC curve above the true positive fraction of 0.9.

TABLE III
CLASSIFIERS RESULTS FOR THE TEN TEST SETS. THE A_z
VALUES REPRESENT THE PARTIAL AREA OF THE ROC CURVE
ABOVE THE TRUE POSITIVE FRACTION OF 0.9 $(A_z^{(0.9)})$

Data Group No.	LDA	ART2LDA	BPN	ART2LDA(1)
1	0.14	0.23	0.31	0.26
2	0.17	0.21	0.28	0.27
3	0.19	0.32	0.27	0.32
4	0.19	0.21	0.19	0.21
5	0.24	0.26	0.32	0.24
6	0.27	0.38	0.27	0.44
7	0.32	0.31	0.38	0.30
8	0.32	0.34	0.25	0.38
9	0.40	0.49	0.40	0.49
10	0.44	0.60	0.38	0.60
Mean	0.27	0.34	0.31	0.35

of ART2 classes. The A_z, and $A_z^{(0.9)}$ results, averaged over the ten test partitions, are presented in Table IV. The average A_z with the ART2LDA classifier, compared to that of LDA alone, was again improved between 15 and 40 classes. The maximum average A_z of 0.80 was achieved between 20 and 40 classes. The average $A_z^{(0.9)}$ results are improved for all

TABLE IV
AVERAGE A_z AND AVERAGE $A_z^{(0.9)}$ CLASSIFICATION RESULTS FOR THE TEN TEST
SETS. CLASSIFIERS WERE DESIGNED USING A FIXED NUMBER OF ART2 CLASSES

No. of classes	LDA	ART2LDA					
		15	20	30	40	50	60
A_z	0.78	0.80	0.80	0.80	0.80	0.78	0.77
$A_z^{(0.9)}$	0.27	0.30	0.31	0.33	0.33	0.31	0.31

ART2LDA classifiers presented in Table IV. The maximum average $A_z^{(0.9)}$ value is 0.33 and it remains constant between 30 and 40 classes.

An alternative way to evaluate the performance of a classifier is its classification accuracy when a decision threshold for malignancy is selected based on the training set. For instance, a decision threshold may be selected such that all positive samples from the training set are classified correctly i.e., at a sensitivity of 100%. The ART2LDA with this decision threshold is referred to as ART2LDA(1). For a given training and test partitioning, ART2LDA classifiers with different number of classes in the ART2 stage were obtained (Figs. 5 and 6). For each of these models the decision threshold for a sensitivity of 100% was selected from the training set and the corresponding ART2LDA(1) classifier was obtained. Then the ART2LDA(1) classifier (with a specific number of classes in the ART2 stage) that correctly classified the maximum number of malignant masses in the test set is selected. By using all samples of the test set, the A_z value is calculated for the corresponding ART2LDA model. The A_z values for the ART2LDA(1) classifiers for the test sets of the ten data partitionings are shown in Tables II and III. For five of the partitions the overall A_z value for ART2LDA(1) is higher than that of LDA alone (Table II). The average A_z value was 0.79. The partial areas above the TP fraction of 0.9, $A_z^{(0.9)}$, for the ten test data sets obtained by the ART2LDA(1) classifier are also shown in Fig. 7. The ART2LDA(1) achieved the highest average $A_z^{(0.9)}$ value of 0.35 compared to ART2LDA and LDA (Table III).

B. BPN Classification Results

A multilayer perceptron back-propagation neural network with a single hidden layer and a single output node was used for comparison with the ART2LDA classifier. The number of selected features determined the number of input nodes to the BPN. The same ten training/test set partitions (as in the case of ART2LDA) were used for the training and validation of the BPN classifiers. BPN's with their number of hidden nodes ranging from two to ten were evaluated to obtain the best architecture. Back-propagation training was used. Each of the BPN's was trained for up to 18 000 training epochs. At every 1000 epochs the neural network weights were saved and the classification result for the corresponding test set was evaluated. This design procedure was repeated for each of the ten training/test groups. For each group, the best test result among all the BPN architectures (different number of hidden nodes) and all the training epochs examined was selected. The average test A_z over the ten groups for the BPN was 0.80, compared to 0.81 for ART2LDA (Table II). The standard deviations of the A_z values for the ten groups range from 0.04 to 0.05 for the BPN. The average partial $A_z^{(0.9)}$ for the BPN

was 0.31, compared to 0.34 for ART2LDA (Table III). The A_z and $A_z^{(0.9)}$ of the ART2LDA classifier were higher than those of the BPN in six of the ten training/test groups.

VI. DISCUSSION

In the present study, a new classifier (ART2LDA) was designed and applied to the classification of malignant and benign masses. The results indicated that the ART2LDA classifier had better generalizability than an LDA classifier alone. The ART2 classifier grouped the case samples that were different from the main population into separate classes. The minimum number of classes needed to start the clustering of outliers into separate classes depended on how different the outliers were from the rest of the sample population. For the ten different partitions of training and test sets used in this study, the minimum number varied between 13 and 15 classes. When the number of ART2 classes was less than this minimum number of classes, the ART2 classifier generated only mixed malignant-benign classes and all samples were transferred to the LDA stage. In that case, the ART2LDA was equivalent to the LDA classifier alone. When a higher number of classes were generated, an increased number of cases that might be considered outliers of the general data population was removed (clustered in separate classes). For the ten training sets used in this study, the malignant outliers were gradually removed when the number of classes increased. The training accuracy increased when the number of classes increased and A_z could reach the value of 1.0. However, a large number of ART2 classes led to overfitting the training sample set and poor generalization in the test set. The classification accuracy of ART2 for the test set tended to decrease when the number of classes was greater than about 70. The large number of classes also led to a reduction in the generalizability of the second-stage LDA; the training of LDA with a small number of samples would again result in overfitting the training set, and poor generalizability in the test set. This effect was observed when more than 60 or 70 classes were generated by ART2 (see Figs. 5 and 6).

The classification accuracy of ART2LDA increased initially with an increased number of classes and then decreased after reaching a maximum. The correct classification of the outliers by the ART2 in combination with an improvement in the classification by the LDA resulted in the increased accuracy. When the number of ART2 classes was further increased, the effects of overfitting by the ART2 and the LDA became dominant and the prediction ability of the ART2LDA decreased. In some cases the second-stage LDA prediction was much worse than the ART2. In other cases the ART2 could not generalize well. The generation of a high number of classes is therefore impractical and unnecessary both from a computational and a methodological point of view.

For the optimal number of classes (usually less than 50 for the data sets used) the A_z value for the second-stage LDA in the ART2LDA was better than an LDA classifier alone, but it was not as good as the overall A_z from the ART2LDA. It is evident that the ART2 was a useful classifier for improvement of the second-stage classification.

When the partial area of the ROC curve above the true positive fraction (TPF) of 0.9 $(A_z^{(0.9)})$ was considered as a measure of classification accuracy, the advantage of ART2LDA over LDA alone became even more evident. By removing and correctly classifying the outliers, the accuracy of the classification was increased at the high sensitivity end of the curve.

The classifier performance was evaluated when the ART2LDA classifiers were designed using a fixed number of ART2 classes. The results showed improved performance of the ART2LDA in a range between 20 and 40 ART2 classes. Both the average A_z and the average $A_z^{(0.9)}$ reached a maximum within this region, and the maximum average A_z and the average $A_z^{(0.9)}$ values remained unchanged between 30 and 40 classes. These results indicated that the performance of a hybrid ART2LDA classifer was robust and stable and could be potentially useful in real clinical applications.

We have performed statistical tests with the CLABROC program to estimate the significance in the differences between the A_z values from the ART2LDA, the LDA alone, and the BPN, as well as in the differences in the partial $A_z^{(0.9)}$ from the three classifiers. The statistical tests were performed for each individual data set partition because the correlation among the data sets from the different partitions precludes the use of student's paired t test with the ten partitions. We found that the differences in both cases did not reach statistical significance because of the small number of test samples and thus the large standard deviation in the A_z values. However, the consistent improvements in A_z and $A_z^{(0.9)}$ by the ART2LDA (9 out of 10 data set partitions in both cases for LDA and six out of ten data set partitions in both cases for BPN) suggest that the improvement was not by chance alone, and that the accuracy of a classification task could be improved by the use of an ART2 network. In addition, one advantage of the ART2LDA is that the training process is more efficient than that of the BPN, especially when there is a subset of outlying samples. In such a case, the BPN will require a large number of training epochs to minimize the error function.

ART2LDA can be trained to classify the sample cases into more than two classes, such as a class of normal tissue regions in addition to malignant and benign masses. There will be an increase in the complexity of training and a larger training sample size will be desired, but these requirements will be comparable for the different classifiers. In a clinical situation, if the classification task is performed on all computer-detected lesions, the classifier has to distinguish the falsely detected normal tissue from malignant or benign lesions. However, it may be noted that a classifier that can distinguish only malignant and benign masses is applicable to the scenario that the radiologist identifies a suspicious lesion on the mammogram and would like to have a second opinion about its likelihood of malignancy before making a diagnostic decision. Therefore, the development of a classifier that can differentiate malignant and benign masses is the research of interest for many investigators.

Similarly, ART2 can be trained to discover and remove a pure benign mass class. The approach will be similar to the task of classifying and removing the pure malignant classes,

as described in this study. However, our approach of removing the malignant classes will reduce the chance of misclassification of malignant masses. In breast cancer detection, the cost of false-negative (missed cancer) is very high. Therefore, our goal in classifier design is to be conservative. By removing the malignant classes in the first stage, any misclassification to these classes will be regarded as malignant. The remaining classes will be classified again with the second-stage classifier so malignant masses will be less likely to be missed.

The problem of classification of malignant and benign masses has been studied by many investigators. Rangayyan *et al.* [15] used Mahalanobis distance classifer (a modification of an LDA classifier) and the leave-one-out method to evaluate the classification of 54 masses. Fogel *et al.* [16] compared LDA and BPN classifiers using the leave-one-out method and 139 masses (malignant and benign classification). Highnam *et al.* [17] used a morphological feature called a halo to classify 40 masses as malignant and benign. Huo *et al.* [22] employed BPN and a rule-based classifier to classify 95 masses using the leave-one-out evaluation method. Sahiner *et al.* [12] used an LDA classifier and the leave-one-out method to classify 168 masses. An important difference between the classifier designed in this study and the previous studies in the CAD field is the method of feature selection. In the above mentioned studies [12], [15]–[17], [22] and several other published studies [18]–[21] the features were selected from the entire data set first, and then the data set was partitioned into training and test sets. This meant that at the feature selection stage of the classifier design, the entire data set was used as a training set. Depending on the distribution of the features and the total number of samples used, the test results in these studies might be optimistically biased [37]. In our current study, the entire data set was initially partitioned into training and test sets and then feature selection was performed only on the training set. This method will result in a pessimistic estimate of the classifier performance when the training set is small [37]. However, it will provide a more conservative but realistic estimation of the classifier performance in the general patient population. We can expect that the performance would be improved if the classifier in this study were designed using a large data set. Since our main purpose in this study was to compare the ART2LDA classifier with the commonly used LDA and BPN, we did not attempt to quantify how pessimistic our results were in this study.

The most important contribution of this paper is to introduce a new approach that utilizes a two-stage unsupervised–supervised hybrid classifier. We believe that the hybrid approach will improve classification when the sample distribution contains subpopulations that may be difficult for a single classifier to classify. It will be useful for similar classification tasks although different classifiers may be used in each stage of the hybrid structure.

VII. CONCLUSION

A new classifier combining an unsupervised ART2 and a supervised LDA has been designed and applied to the classification of malignant and benign masses. A data set consisting of 348 films (179 malignant and 169 benign) was randomly partitioned into training and test subsets. Ten different random partitions were generated. For each training set, texture features were extracted and feature selection was performed. An average of features were selected for each group. A hybrid ART2LDA classifier, an LDA, and a BPN were trained by using each of the ten training sets. The A_z value under the ROC curve for the test sets, averaged over the ten partitions, was higher for ART2LDA ($A_z = 0.81$) compared to those of the LDA alone ($A_z = 0.78$) and of the BPN ($A_z = 0.80$). A greater improvement was obtained when the partial ROC area above a true-positive fraction of 0.9 was considered. The average partial A_z for ART2LDA was 0.34, as compared to 0.27 for LDA and 0.31 for BPN. Additionally, for the ART2LDA classifiers that correctly classified the maximum number of malignant masses in the test sets with decision threshold defined with the training set, the average partial A_z was 0.35. These results indicate that the hybrid classifier is a promising approach for improving the accuracy of classifiers for CAD applications.

REFERENCES

[1] H. C. Zuckerman, "The role of mammography in the diagnosis of breast canser," in *Breast Canser, Diagnosis and Treatment*, I. M. Ariel and J. B. Cleary, Eds. New York: McGraw-Hill, 1987, pp. 152–172.

[2] D. B. Kopans, "The positive predictive value of mammography," *Amer. J. Roentgenol.*, vol. 158, pp. 521–526, 1992.

[3] D. D. Adler and M. A. Helvie, "Mammographic biopsy recommendations," *Curr. Opin. Radiol.*, vol. 4, pp. 123–129, 1992.

[4] M. Moskowitz, "Impact of a priory medical detection on screening for breast cancer," *Radiology*, vol. 184, pp. 619–622, 1989.

[5] P. A. Lachenbruch, *Discriminant Analysis*. New York: Hafner, 1975.

[6] R. O. Duda, and P.E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[7] P. J. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Harvard Univ., Cambridge, MA, 1974.

[8] D. Rumelhart, G. E. Hinton, and R. J. Williams, in D. E. Rumelhart, Ed., *Parallel and Distributed Processing*. Cambridge, MA: MIT Press, 1986, vol. 1, p. 318.

[9] J. Herz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation*. Reading, MA: Addison-Wesley, 1991.

[10] H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: Linear discriminat analysis in texture feature space," *Phys. Med. Biol.*, vol. 40, pp. 857–876, 1995.

[11] D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograms: Multiresolution texture analysis," *Med. Phys.*, vol. 22, pp. 1501–1513, 1995.

[12] B. Sahiner, H. P. Chan, N. Petick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mamograms: The rubber band sraightening transform and texture analysis," *Med. Phys.*, vol. 25, no. 4, pp. 516–526, Apr. 1998.

[13] B. Sahiner, H. P. Chan, D. Wei, N. Petick, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue," *Med. Phys.*, vol. 23, no. 10, pp. 1671–1683, Oct. 1996.

[14] H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, K. L. Lam, D. D. Adler, and M. M. Goodsitt, "Computerized classification of malignant

and benign microcalsifications on mammograms: Texture analysis using an artificial neural network," *Phys. Med. Biol.*, vol. 42, pp. 549–567, 1997.

[15] R. M. Rangayyan, N. M. El-Farmawy, J. E. Desautels, and O. A. Alim, "Measures of acutance and shape for classification of breast tumors," *IEEE Trans. Med. Imag.*, vol. 16, pp. 799–810, Dec. 1997.

[16] D. B. Fogel, E. C. Wasson, E. M. Boughton, V. W. Porto, and P. J. "Angeline, linear and neural model for classifying breast masses," *IEEE Trans. Med. Imag.*, vol. 17, pp. 485–488, June 1998.

[17] R. P. Highnam, J. M. Brady, and B. J.Shepstone, "A quantitative feature to aid diagnosis in mammography," in *Proc. Digital Mammography'96*, pp. 201–206.

[18] Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C. E. Metz, "Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer," *Radiology*, vol. 187, pp. 81–87, 1993.

[19] V. Goldberg, A. Manduca, D. L. Evert, J. J. Gisvold, and J. F. Greenleaf, "Improvements in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence," *Med. Phys.*, vol. 19, pp. 1475–1481, 1992.

[20] J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computerized image analysis," *IEEE Trans. Med. Imag.*, vol. 12, pp. 664–669, Dec. 1993.

[21] M. F. McNitt-Gray, H. K. Huang, and J. W. Sayre, "Feature selection in the pattern classification problem of digital chest radiograph segmentation," *IEEE Trans. Med. Imag.*, vol. 14, pp. 537–547, Sept. 1995.

[22] Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," *Acad. Radiol.*, vol. 5, pp. 155–168, 1998.

[23] M. Jordan, and R. A. Jacobs, "Hierarchical mixture of experts and EM algorithm," *Neural Comput.*, vol. 6, pp. 181–214, 1994.

[24] L. Hadjiiski and P. Hopke, "Design of large scale models based on multiple neural network approach," *Intelligent Engineering Systems Through Artificial Neural Networks.* ASME, 1997, vol. 7, pp. 61–66.

[25] S. Grossberg, "Adaptive pattern classification and universal recoding. I. Parallel development and coding of neural feature detectors," *Biolog. Cybern.*, vol. 23, no. 3, pp. 121–134, 1976.

[26] G. A. Carpenter and S. Grossberg, "ART 2: Self-organization of stable category recognition codes for analog input patterns," *Appl. Opt.*, vol. 26, no. 23, 1, pp. 4919–4930, Dec. 1987.

[27] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition," *Neural Networks*, vol. 4, no. 4, pp. 493–504, 1991.

[28] G. A. Carpenter and S. Grossberg, "Integrating symbolic and neural processing in a self-organizing architeture for pattern recognition and prediction," in *Artificial Intelligence and Neural Networks: Steps toward Principled Integration.* New York: Academic, 1994.

[29] G. A. Carpenter and N. Markuzon, "ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases," *Neural Networks*, vol. 11, no. 2, pp. 323–336, Mar. 1998.

[30] Y. Xie, P. K. Hopke, and D. Wienke, "Airborne particle classification with a combination of chemical composition and shape index utilizing an adaptive resonance artificial neural network," *Environ. Sci. Technol.*, vol. 28, no. 11, pp. 1921–1928, 1994.

[31] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, pp. 610–621, Nov. 1973.

[32] M. M. Galloway, "Texture analysis using gray level run length," *Comput. Graph. Image Processing*, vol. 4, pp. 172–179, 1975.

[33] M. J. Norusis, *SPSS Professional Statistics 6.1.* Chicago, IL: SPSS, 1993.

[34] M. M. Tatsuoka, "Multivariate Analysis," *Techniques for Educational and Psychological Research.* New York: Macmillan, 1988.

[35] C. E. Metz, "ROC methodology in radiographic imaging," *Invest. Radiol.*, vol. 21, pp. 720–733, 1986.

[36] C. E. Metz, J. H. Shen, and B. A. Herman, "New methods for estimating a binomial ROC curve from continuously distributed test results," presented at the *1990 Annu. Meeting American Statistical Association*, Anahaim, CA, 1990.

[37] B. Sahiner, H. P. Chan, N. Petrick, R. Wagner, and L. Hadjiiski, "The effect of sample size on feature selection in computer-aided diagnosis," *Proc. SPIE*, vol. 3661, pp. 499–510, 1999.

1999 Scientific Program

Radiological Society of North America

85th Scientific Assembly and Annual Meeting

A CENTURY
OF RESEARCH & EDUCATION

RSNA 1999

RSNA | Radiological Society
of North America
Founded in 1915

American Association
of Physicists in Medicine

Scientific Sessions

Tuesday Morning • Room S401AB

Physics (Image Processing: Computer-aided [Diag]nosis - Mammography and Pulmonary)

[Jo]int sponsorship with the American Association of [Phy]sicists in Medicine

[PRESI]DING: **Carey E. Floyd, Jr, PhD,** Durham, NC

[Com]puter Code: H01 • 1½ hours

[Fo]r credit, relinquish attendance voucher at end of session.

10:30 AM

[Comp]uter-aided Diagnosis (CAD) of Images Obtained on Full-Field Digital [Mamm]ography

[R.M. Ni]shikawa, PhD, Chicago, IL • M.L. Giger, PhD • L.M. Yarusso, BS • M.A. [Kuprinski,] BS • A.E. Baehr, MS • L.A. Venta, MD • et al

[PURPO]SE: To demonstrate computer-aided diagnosis methods on full-[d]igital mammograms.

[METHO]D AND MATERIALS: We have developed computerized detection [techni]ques to assist radiologists detect breast cancer on digitized screen-[film ma]mmograms. To date, 18,500 screening cases have been digitized [using t]wo different detection schemes – one for masses and the other for [detect]ed microcalcifications – have analyzed the images. We have per-[formed f]ollow-up analyses on the first 10,000 cases. Sixty-seven women in [this] cohort developed breast cancer. The computer was able to detect [approxim]ately 65% of these cancers at a false-positive rate of 2.0 false [and 0.9] false clusters per image. In 17 mammographically-negative [cases the] computer was able to detect the cancer in 8 of them. We are now [applying the] computerized detection method to 150 full-field digital [mammog]rams and comparing the results to the corresponding screen-film [mammog]rams.

[F]rom an assessment of initial processing of the full-field digital [mammogr]ams, our single view mass detection algorithm selects regions of [interest that] correspond to suspect mass lesions. The influence of periph-[eral enhan]cement filtering to correct for the breast thickness is also [studied. In]formation on the characteristic curve, modulation transfer [function and] noise properties of the FFDM unit need to be incorporated [into CA]D methods for optimal performance, particularly for the [calcificat]ion detection scheme.

[CONCLUSION]S: CAD methods initially developed on digitized screen/[film mammo]grams can be used in the analysis of full-field digital [mammogram]s. (RMN and MLG are shareholders in R2 Technology, Inc., [Los Altos, CA.]) [See also scientific exhibit 0180BR.]

[10:3]9 AM

[565 •] Characterization of Solid Breast Masses Using Three-[Dimensional Ultr]asound Images

[B. Sahiner, PhD, A]nn Arbor, MI • H. Chan, PhD • G.L. LeCarpentier, PhD • N.A. [Petrick, PhD • J.A.] Roubidoux, MD • P.L. Carson, PhD • et al

[PURPOSE: To inv]estigate the feasibility of using texture features extracted [from three-dim]ensional (3-D) ultrasound image volumes for discriminat-[ing mali]gnant and benign solid breast masses.

[METHOD AND] MATERIALS: Our preliminary data set included 3-D [images from] 34 patients obtained with a GE Logic 700 scanner [set to array] at 9MHz. All recruited patients had been scheduled [for biopsy of] a solid or indeterminate breast mass. Eighteen of the masses [were found to be mal]ignant and 16 were found to be benign. Parallel image [slices at fixed in]tervals were obtained by translating the transducer [in the ima]ge plane and acquiring the translation distance with a [position encoder. An] experienced breast radiologist then identified each [mass with a] 3-D ellipsoid, dynamically positioned and shaped [over the select]ed 3-D volume. A segmentation technique based on [the] developed in order to improve the ellipsoidal approxi-[mation. S]patial gray level dependence features were extracted [from] regions of interest (ROIs) on each slice that contained [the mass. One R]OI was the interior of the segmented lesion while the •

second and third ROIs were disk-shaped regions above and below the segmented lesion. Features extracted from the 2-D slices of a mass were averaged to obtain a 3-D feature. A total of 24 features were extracted from each of the lower and upper disk-shaped regions and a total of 48 features were extracted from the interior of the mass. The accuracy of each feature was evaluated using receiver operating characteristic analysis.

RESULTS: The information measure of correlation (IMC) texture feature provided the best separation between malignant and benign masses. The best classification results with 3-D features (A_z=0.85) were obtained by using the IMC feature extracted from the interior of the mass defined by either the active contour model or the ellipsoid. The lower disk-shaped region also provided effective features (A_z=0.84) when the active contour model was used. The classification result with the IMC feature extracted from the best single slice was also A_z=0.85.

CONCLUSIONS: Our preliminary results indicate that texture analysis of solid breast masses on 3-D ultrasound images can provide effective features for classifying malignant and benign lesions. Since the best 2-D slice for texture analysis may not be known a-priori, 3-D analysis may be more useful than 2-D analysis. We are currently in the process of improving the active contour model, investigating additional features, and expanding the data set.

566 • 10:48 AM

Computerized Detection of Mass Lesions in Digital Mammography Using Radial Gradient Index Filtering

M.A. Kupinski, BS, Chicago, IL • M.L. Giger, PhD • A.E. Baehr, MS

PURPOSE: To evaluate the ability of radial gradient index (RGI) filtering to detect mass lesions in single-view mammograms.

METHOD AND MATERIALS: We have developed a new filtering method that computes the maximum RGI value throughout the digital mammogram. This processed feature image is subsequently thresholded to determine potential mass lesion sites, which are input to a feature analysis stage. The features characterizing each suspicious area are extracted and merged with an artificial neural network (ANN), the output of which reports the likelihood of being an actual lesion. The performance of this technique will be presented on a screening database of 60 patients with non-palpable lesions.

RESULTS: Preliminary studies show that the RGI detection method yielded an 81% reduction in the false positives at a sensitivity of 80% over the previous method of bilateral subtraction. The ANN consisted of 3 input nodes, 5 hidden nodes and a single output node. This corresponded to a performance of 80% sensitivity at 2 false-positives per image.

CONCLUSIONS: The new RGI pre-processing gives a substantial improvement in the sensitivity and specificity over the previous bilateral subtraction method. This is expected to enhance the CAD performance on digital mammograms. (MLG is a Stockholder, R2 Technology, Inc., Los Altos, CA.)

567 • 10:57 AM

Automated Identification of Breast Lesions in Temporal Pairs of Mammograms for Interval Change Analysis

L.M. Hadjiiski, PhD, Ann Arbor, MI • H. Chan, PhD • B. Sahiner, PhD • N.A. Petrick, PhD • M.A. Helvie, MD • S. Sanjay-Gopal, PhD

PURPOSE: To develop a registration technique for automated identification of corresponding lesion on a previous mammogram. This technique is the basis for interval change analysis of breast lesions in computer-aided diagnosis applications.

METHOD AND MATERIALS: A multistage regional registration technique that combines global and local alignment procedures was developed for identifying masses on temporal pairs of mammograms. In the first stage, the breast images from the current and previous mammograms were globally aligned by maximizing a mutual information measure. An initial fan-shape search region was then estimated on the previous mammogram based on the mass location on the current mammogram. In the second stage, the location of the fan-shape search region was locally refined by maximizing a correlation measure between a template of the fan-shape region defined on the current mammogram and the breast structures on the previous mammogram. In the third stage a search for the best match between the lesion template from the current mammogram and a structure on the previous mammogram was carried out within this search region. Fine alignment of the lesion within a small neighborhood of the matched region was then performed based on simplex optimization. A set of 74 temporal pairs of mammograms containing biopsy-proven masses was used to examine the performance of this approach. The true lesion locations were identified by an experienced radiologist on all mammograms. The accuracy of the multistage regional registration was analyzed by evaluating the area of overlap between the estimated and the true lesions on the previous mammogram. The average distance between the centroids of the estimated and true lesion locations was also calculated.

RESULTS: In this study 89% of the estimated lesion locations result in an area overlap of at least 50% with the true lesion locations. The average distance between the estimated and the true centroid of the lesions on the previous mammogram was 4.9±4 mm.

CONCLUSIONS: The multistage regional registration technique is useful for identification of corresponding lesions on temporal pairs of mammograms. Further studies are underway to improve the technique and to evaluate its accuracy on a larger data set.

568 • 11:06 AM
Knowledge-based Computer-aided Detection of Masses on Digitized Mammograms: A Preliminary Study

Y. Chang, PhD, Pittsburgh, PA • J.L. King, MS • J. Drescher, BS • X. Wang, MD, PhD • B. Zheng, PhD • W.F. Good, PhD

PURPOSE: _The purpose of this work was to investigate a computer-aided detection (CAD) scheme for the identification of masses on digitized mammograms using a knowledge-based approach.

METHOD AND MATERIALS: _The methods of the CAD scheme included a learning process to establish a knowledge base and an identification process to determine if a suspicious region is likely to be "positive" for a mass. The main idea behind this scheme was to investigate a CAD scheme using a knowledge-based approach so that the scheme is largely independent from other rule-based approaches. During knowledge-based learning, a set of masses that were retrospectively verified by radiologists was quantitatively characterized to establish a knowledge base (i.e., "known" truth). During knowledge-based identification, "similarity" measures among a suspicious region and all the "known" masses were computed. And, a "likelihood" measure as derived from the "similarity" measures was used to compare the suspicious region and all the "known" masses to determine the state of the suspicious region ("positive" or "negative"). To facilitate this study, 300 verified masses, and 300 false-positive regions as identified by our previous rule-based CAD scheme were randomly chosen from a large clinical database. Receiver-operating characteristic (ROC) analysis was then performed to evaluate the scheme performance in distinguishing between true- and false-positive regions.

RESULTS: _Using a leave-one-out validataion method, the knowledge-based CAD scheme achieved an area under the ROC curve of 0.77. With 90% sensitivity of the true-positive masses, the CAD scheme achieved 56% of false-positive identifications (or 44% reduction).

CONCLUSIONS: _The knowledge-based approach can yield a significant reduction of false-positive regions while maintaining reasonable sensitivity of true-positive masses. Such a knowledge-based CAD scheme can potentially be used to improve the performance of rule-based CAD schemes.

569 • 11:15 AM
Computerized Classification of Mass Lesions on Images from a Medium-field Digital Mammography Unit

M.M. Maloney, BS, Chicago, IL • Z. Huo, PhD • M.L. Giger, PhD • L.A. Venta, MD • C.J. Vyborny, MD, PhD

PURPOSE: To evaluate the performance of a computerized method, which was developed on digitized screen/film mammograms, in the task of classifying mammographic lesions in images obtained on a medium-field digital mammography unit.

METHOD AND MATERIALS: A computer-aided diagnosis classification method, which outputs an estimate of the likelihood of malignancy, was developed using digitized screen/film mammograms. The method includes automated segmentation, automated feature extraction, and automated classification using artificial neural networks. Previously, the method yield an Az of 0.90 and a partial Az of 0.40 in a round-robin analysis of the training database (65 cases of digitized screen/film mammograms) and an Az of 0.82 and a partial Az of 0.43 in a validation analysis on an independent database (110 cases of digitized screen/film mammograms). In the current study, the performance of the computer is being evaluated on 100 cases obtained on a medium-field digital mammography unit instead of digitized screen/film mammograms.

RESULTS: In the initial analysis on 20 digital cases, without further optimization or calibration from the prior digitized screen/film mammogram training, individual features of biopsy-proven lesions yielded Az values ranging from 0.75 to 0.86 in the task of distinguishing between malignant and benign lesions. Further improvement is expected after calibrating the computer classification method for the digital acquisition system.

CONCLUSIONS: Our computerized method, which was developed on digitized screen/film mammograms, performs well also on digital images obtained on a medium-field digital mammography unit. (M.L. Giger is a shareholder in R2 Technology, Inc. Los Altos, CA. C. J. Vyborny is a shareholder in R2 Technology, Inc., Los Altos, CA.)

570 • 11:24 AM
Segmentation of Suspicious Microcalcification Clusters in Mammograms

M.A. Gavrielides, MS, Durham, NC • J.Y. Lo, PhD • R. Vargas-Voracek, PhD • C.E. Floyd, Jr, PhD

PURPOSE: To develop a computer-aided diagnosis (CAD) scheme for the automated segmentation of suspicious microcalcification clusters in digital mammograms.

METHOD AND MATERIALS: The scheme consisted of three main processing steps. First, the breast region was segmented and its high frequency content was enhanced using unsharp masking. In the second step, individual microcalcifications were segmented using local histogram analysis on overlapping subimages. For this step, eight histogram features were extracted for each subimage and were used as input to a fuzzy rule-based classifier which identified subimages containing microcalcifications and assigned appropriate local thresholds to segment any microcalcifications within them. The final step clustered the segmented microcalcifications and extracted a set of cluster features, namely number of microcalcifications, average distance between microcalcifications and average number of times pixels in the cluster were segmented in the second step. Fuzzy rules incorporating the cluster features were designed to remove any nonsuspicious clusters, defined as those with typically benign characteristics. A database of 98 images, with 48 images containing one or more microcalcification clusters, provided training and testing sets to optimize the parameters of the method and evaluate its performance respectively.

RESULTS: The results showed a true positive rate of 93.2% and an average of 0.73 false clusters per image. Comparison of our results with other reported segmentation results on the same database showed comparable sensitivity and an improved false positive rate. Our results show that histogram analysis using the selected set of histogram features and the designed fuzzy rule-based classifier can provide an accurate technique for segmenting microcalcifications. Moreover, using the proposed set of cluster features and by focusing on segmenting suspicious clusters, we were able to design a set of decision rules that was efficient in reducing the number of false positive clusters.

CONCLUSIONS: The performance of our CAD scheme is encouraging for its further development as an automatic tool for efficient and accurate diagnosis of breast cancer. [See also scientific exhibit 0184BR.]

571 • 11:33 AM
Computerized Detection of Pulmonary Nodules on Digital Chest Radiographs Using a Contralateral Subtraction Technique

Q. Li, PhD, Chicago, IL • S. Katsuragawa, PhD • R.M. Engelmann, MS • H. MacMahon, MD • K. Doi, PhD

PURPOSE: We are developing a computer-aided diagnostic (CAD) scheme to assist radiologists in the detection of pulmonary nodules in chest radiographs. A major problem in the current CAD scheme is its high false positive rate, mainly due to ribs and rib crossings. Therefore, we have implemented a novel contralateral subtraction technique, in which symmetric skeletal structures such as ribs can be eliminated to reduce false positives, and asymmetric subtle lesions can be enhanced.

METHOD AND MATERIALS: With our CAD scheme, a chest image was first enhanced using a difference image technique for selection of initial nodule candidates. Various image features for nodule candidates were then determined from the difference image, original image and the gradient enhanced image. Rule-based and ANN methods were applied to eliminate some false positives based on the image features. To further reduce false positives, a contralateral subtraction image was obtained by subtraction of a right/left reversed "mirror" image from the original chest image. The misregistration artifacts in the subtraction image were reduced by correction of asymmetric ribs in two lungs using generalized Hough transform and snake model techniques, and by accurate registration of mirror image to the original image using an elastic matching and warping technique. The contrast and relative standard deviation were determined at the corresponding locations of nodule candidates in the subtraction image, to distinguish between true nodules and false positives.

RESULTS: In a pilot study, the CAD scheme was applied to 100 radiographs which consists of 50 normals and 50 abnormals, each with a solitary pulmonary nodule in the lung periphery. A total of 43 nodules and 343 false positives were detected by the original CAD scheme. By examining the features of nodule candidates on the contralateral subtraction images, 188 (54.8%) false positives were eliminated with a reduction of two true positives. The integrated CAD scheme with contralateral subtraction technique achieved a sensitivity of 82% at 1.55 false positives per image.

CONCLUSIONS: The contralateral subtraction technique can eliminate most rib shadows, enhance low-contrast lesions, and thus significantly improve the performance of the CAD scheme for pulmonary nodule detection. (HM and KD are shareholders of R2 Technology, Inc., Los Altos, CA.)

SPIE's International Symposium on

Medical Imaging

12-17 February 2000
Town and Country Hotel
San Diego, California USA

Program Updates

Final Summaries

3979-87, Sunday / Monday Poster Sessioin
Interval Change Analysis in Temporal Pairs of Mammograms Using a Local Affine Transformation

Lubomir Hadjiiski, Heang-Ping Chan, Berkman Sahiner, Nicholas Petrick, Mark A. Helvie, Sophie Paquerault, Chuan Zhou (Department of Radiology, The University of Michigan, Ann Arbor, MI 48109-0904)

The aim of this study is to evaluate the local affine transformation in a computer aided diagnosis technique used for interval change analysis of mammograms. A multistage regional registration technique was developed for identifying masses on temporal pairs of mammograms. In the first stage, the breast images from the current and previous mammograms were globally aligned. An initial fan-shape search region was defined on the previous mammogram. In the second stage, the location of the fan-shape region was refined by warping, based on the affine transformation and simplex optimization. A new refined search region was defined on the previous mammogram. In the third stage, a search for the best match between the lesion template from the current mammogram and a structure on the previous mammogram was carried out within the search region. This technique was evaluated on 107 temporal pairs of mammograms containing biopsy-proven masses. Eighty-eight percent of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations. The average distance between the estimated and the true centroid of the lesions on the previous mammogram was 4.3 ± 4.9 mm. Local warping based on affine transform improves the accuracy of the identification.

3979-88, Sunday / Monday Poster Session
Segmentation of the Fractured Foot CT Image: A Fuzzy Rule-based Approach

Shoji Hirano, Yutaka Hata, Nobuyuki Matsui, Yoshihiro Ando, Makoto Ishikawa (Himeji Institute of Technology, Himeji, 671-2201, Japan)(YA, MI Ishikawa Hospital, Himeji, 671-0221, Japan)

Purpose: Extraction and 3D visualization of each fragment and normal bone in the fractured foot CT images. **Method:** Segmentation boundary is determined by fuzzy inference with two types of knowledge acquired from orthopedic surgeons. Knowledge of joint is used to determine the boundary of adjacent normal bones. It gives higher degree to the articular cartilage according to local structure (parallelity) and intensity distribution around a joint part. Knowledge of fragment is used to find a contact place of fragments. It evaluates EDM of the contact place and gives higher degree to the narrow part. Each of the knowledge is represented by fuzzy if-then rules, which can provide degrees for segmentation boundary. By evaluating the degrees in region growing process, a whole foot bone is decomposed into each of anatomically meaningful bones and fragments. **Breakthroughs:** With the knowledge of joint and fragment, we can effectively give higher degrees on the essential boundary, suppressing generation of useless boundary caused by the internal cavities in the bone. **Results:** An experiment was done on CT images of the foot with depressed fracture on the calcaneus.Each of the normal bones and fragments were correctly segmented. **Conclusions:** Fuzzy expert system could lead successful segmentation of the CT images. The 3D visualization would expedite understandings of spatial positions and shapes of the fragments.

3979-89, Sunday / Monday Poster Session
Quantitative Analysis of Internal Texture for Classification of Pulmonary Nodules in Three-dimensional Thoracic images

Yoshiki Kawata, Noboru Niki, Hironobu Ohmatsu, Masahiko Kusumoto, Ryutaro Kakinuma, Kiyoshi Mori, Hiroyuki Nishiyama, Kenji Fguchi, Masahiro Kaneko, Noriyuki Moriyama (Dept. of Optical Science, Univ. of Tokushima, Tokushima, 770-9506, Japan) (NO, MK, RK, MK National Cancer Center Hospital, Tokyo, Japan) (NM National Cancer Center Hospital East, Chiba, Japan) (KE National Shikoku Cancer Center Hospital, Ehime, Japan) (KM Tochigi Cancer Center, Tochigi, Japan) (HN The Social Health Insurance Medical Center, Tokyo, Japan)

We are developing a computerized feature extraction and a classification method to analyze malignant and benign pulmonary nodules in three-dimensional (3-D) thoracic images. This paper focuses on an approach for characterizing the internal texture which is one of important clues for differentiating between malignant and benign nodules. In this approach, each voxel is described in terms of shape index derived from curvatures on the voxel. The voxels inside the nodule are aggregated via shape histogram to quantify how much of shape category is present in the nodule. Topological features were introduced to characterize the morphology of the cluster constructed from a set of voxels with the same shape category. Properties such as curvedness and CT density were also built into the representation. We evaluated the effectiveness of topological and histogram features extracted from 3-D pulmonary nodules for classification of malignant and benign internal structures. We also compared the performance of the computerized classification to that of the experienced physician. The classification performance based on the combined feature space reached the performance of the experienced physician. Our results demonstrate the feasibility of using topological and histogram features for analyzing internal texture to assist physicians in making diagnostic decisions.

3979-90, Sunday / Monday Poster Session
Fuzzy Approach to Brain Region Classification in Talairach Space

S. Kemeny, S. G. Erberich (Dept. of Neurology, Univ. Hospital, Univ. of Technology, RWTH Aachen, Germany) (SK, SE, Interdisciplinary Center for Clinical Research, Univ. Hospital, Univ. of Technology, RWTH Aachen, Germany)

Functional mapping of the human brain using PET or fMRI has become a widely used method in the neurosciences. Both techniques identify cortical activations by statistical means and overlay the results on standardized templates like the stereotaxic space defined by Talairach and Tournoux. The interpretation of the results involves the correct identification of the underlying anatomy, usually by looking up the relevant coordinates and finding the corresponding anatomical labels. This linear assignment of a coordinate to a corresponding anatomical structure is problematic, because most of the anatomical structures of a real brain do not have clearly identifiable borders. A given coordinate could be part of an anatomical region while a neighboring voxel just one millimeter away would belong to an other region, although there is no morphological difference between both. Furthermore, rigid categorization of normalized data from single subjects ignores individual variety of anatomy. To overcome these pitfalls we developed a computer based atlas with a "fuzzy" identification of anatomy. The program allows the user to import Talairach coordinates and to export the corresponding anatomical names. In contrast to a rigid database, we defined zones and the program knows about anatomical "hotspots", the nearer a coordinate lies to the center of this "hotspot" the more likely it belongs to that region. Coordinates adjacent to two or more hotspots get two or more anatomic definitions, together with a probability score for each. The algorithm used to find the relevant hotspot is similar to the function when one describes the attraction of a mass to a larger mass based on their gravity, depending on size and position of the center and surrounding centers as well.

Interval change analysis in temporal pairs of mammograms using a local affine transformation

Lubomir Hadjiiski, Heang-Ping Chan, Berkman Sahiner, Nicholas Petrick, Mark A. Helvie,

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

ABSTRACT

The aim of this study is to evaluate the use of a local affine transformation for computer-aided interval change analysis in mammography. A multistage regional registration technique was developed for identifying masses on temporal pairs of mammograms. In the first stage, the breast images from the current and prior mammograms were globally aligned. An initial fan-shape search region was defined on the prior mammogram. In the second stage, the location of the fan-shape region was refined by warping, based on an affine transformation and simplex optimization. A new refined search region was defined on the prior mammogram. In the third stage a search for the best match between the lesion template from the current mammogram and a structure on the prior mammogram was carried out within the search region. This technique was evaluated on 124 temporal pairs of mammograms containing biopsy-proven masses. Eighty-six percent of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations. The average distance between the estimated and the true centroid of the lesions on the prior mammogram was 4.4 ± 5.9 mm. The registration accuracy was improved in comparison with our previous study that used a data set of 74 temporal pairs of mammograms. This improvement gain is mainly from the local affine transformation.

Keywords: Computer-Aided Diagnosis, Interval Changes, Affine Transform, Correlation, Mutual Information, Mammography, Malignancy

1. INTRODUCTION

Mammography is currently the most effective method for early breast cancer detection[1,2]. Analysis of interval changes is an important method used by radiologists to detect developing malignancy in mammographic interpretation[3,4]. A variety of computer-aided diagnosis (CAD) techniques have been developed to detect mammographic abnormalities and to distinguish between malignant and benign lesions. We are studying the use of CAD techniques to assist radiologists in interval change analysis.

Sallam et al.[5] have proposed a warping technique for mammogram registration based on manually obtained control points. A mapping function was calculated for mapping each point on the current mammogram to a point on the prior mammogram. Brzakovic et al.[6] have investigated a three-step method for comparison of most recent and prior mammograms. They first registered two mammograms using the method of principal axis, and partitioned the current mammogram using a hierarchical region-growing technique. Translation, rotation, and scaling were then used for registration of the partitioned regions. Vujovic et al.[7] have proposed a multiple-control-point technique for mammogram registration. They first determined several control points independently on the current and prior mammograms based on the intersection points of prominent anatomical structures in the breast. A correspondence between these control points was established based on a search in a local neighborhood around the control point of interest.

The previous techniques depend on the identification of control points. However, because the breast is mainly composed of soft tissue that can change over time, there are no obvious landmarks on mammograms. The crossing line structures are often fibrous tissue from different depths of the breast which overlap in a projection image. These crossing points are not invariant landmarks on the different mammograms. Because of the elasticity of the breast tissue, there is large variability in the positioning and compression used in mammographic examination. As a result, the relative positions of the breast tissues projected onto a mammogram vary from one examination to the other. Techniques that depend on identification of control points may not be generally applicable to registration of breast images.

Gopal et al.[8,9,10] and Hadjiiski et al.[11] have developed a multistage technique that defines a transformation to locally map the position of the mass on a current mammogram to that on the prior mammogram. A local search for the mass is then performed on the prior mammogram. Good et al.[12] also have developed a technique that defines a transformation to map all points from the current mammogram onto a prior mammogram. Then the current mammogram is subtracted from the prior mammogram.

The goal of our research is to develop a technique for computerized analysis of temporal differences between a lesion on the most recent mammogram and a prior mammogram of the same view. The computer algorithm will assist radiologists in quantifying interval changes and thus distinguishing between benign and malignant masses for CAD. In this study, we developed a local registration technique based on affine transformation and simplex optimization and evaluated its usefulness in improving the localization of the mass on the prior mammogram.

2. REGISTRATION TECHNIQUE

A multistage regional registration technique was developed for identifying corresponding masses on temporal pairs of mammograms. It combined both a global and a local alignment procedure. The block diagram of the regional registration technique is shown in Fig. 1. In the first stage, the breast images from the current and prior mammograms were globally aligned by maximizing a mutual information measure. An initial fan-shape search region was then defined on the prior mammogram based on the mass location on the current mammogram.
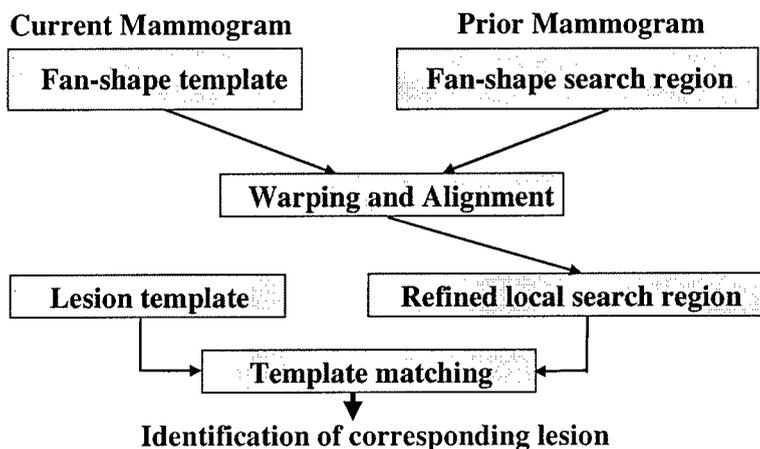
Figure 1. Block-diagram of the regional registration technique.

In the local alignment stage, the location of the search region on the prior mammograms was first refined by maximizing a correlation measure between a template extracted from the current mammogram and the breast structures on the prior mammogram. The affine transformation in combination with simplex optimization was then employed to warp this local region. In the final stage a search for the best match between the lesion template from the current mammogram and a structure on the prior mammogram was carried out within the refined search region. A more detailed explanation for each of the stages will be presented in the following subsections.

2.1. Stage1: Global alignment

Initially an automatic procedure is used to detect the breast boundary for all of the mammograms (Fig. 2). In the first stage, the breast images from the current and prior mammograms were globally aligned by maximizing a mutual information measure. The nipple locations of the two breast images were used as the pivot points in a common reference frame for the alignment.

Current (1996)   Prior (1995)
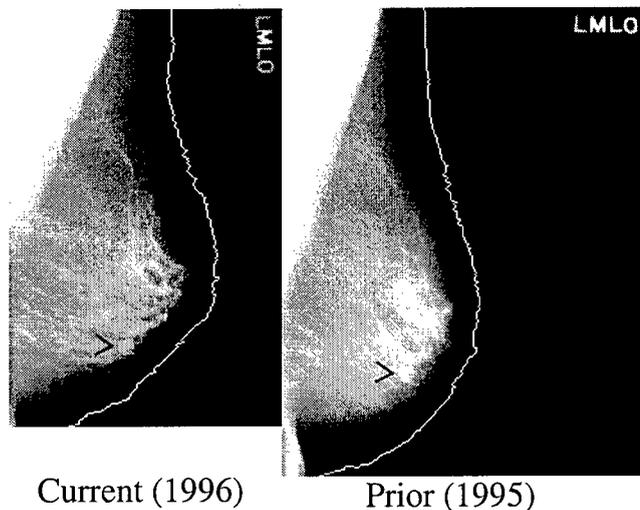
Figure 2. Current and prior mammogram. The arrows point to the
masses on the current and the prior mammograms.

2.2. Stage 1: Definition of fan-shape regions

The location of the mass on the current mammogram is determined in a polar coordinate system with the nipple as the origin. The location is represented as the radial distance from the nipple and the angle between the nipple-mass centroid axis and the breast periphery. Angular and radial scaling are performed and the position of the mass on the prior mammogram is predicted in a polar coordinate system defined in a similar manner (Fig. 3). An initial fan-shape search region is then defined on the prior mammogram centered at the predicted location of the mass centroid (Fig. 4).

Current   Prior   Current   Prior
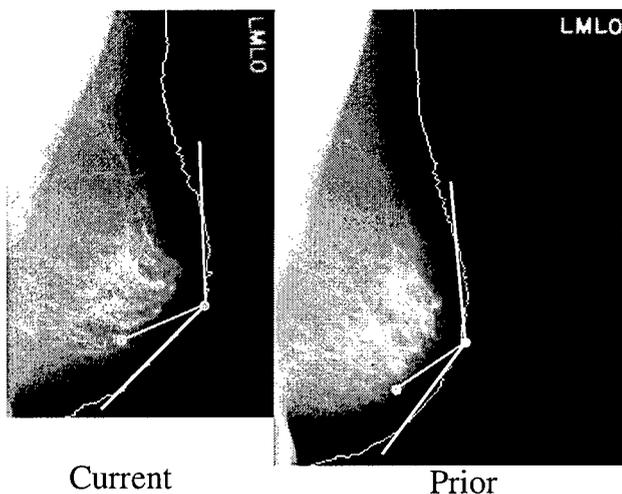
Figure 3. Initial estimation of the mass centroid position on the prior mammogram based on the nipple-mass distance and the angle between the nipple-mass axis and breast periphery on the current mammogram.
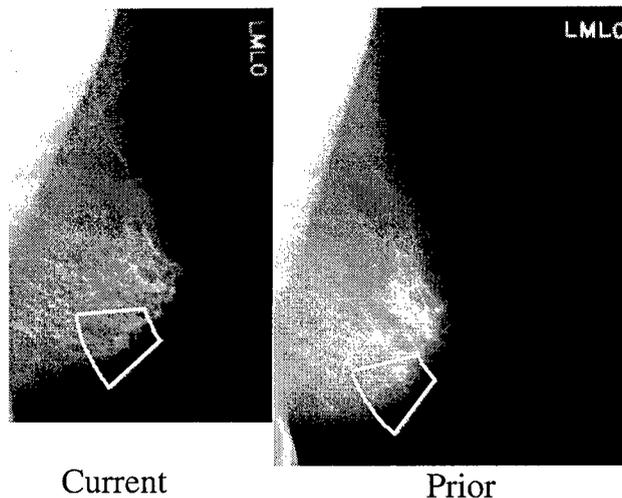
Figure 4. Definition of an initial fan-shape search region on the prior mammogram and a fan-shape template on the current mammogram.

The size of the fan-shape region is predefined using a training set so that it will include the mass centroid on the prior mammograms. A fan-shape template centered at the mass is also defined on the current mammogram.

2.3. Stage 2: Warping and alignment

In the second stage the fan-shape search region is refined. By allowing warping of the fan-shape template from the current mammogram, it may provide better compensation for local geometric distortions due to differences in positioning of the breast and in breast compression and may improve the localization of the mass on the prior mammogram. The warping procedure is based on the affine transformation combined with simplex optimization. In the following the warping procedure is explained in greater detail.

2.3.1. Affine transformation

An affine transformation[13] is a linear transformation combining rotation and translation. A two dimensional affine transformation is defined as follows:

$$
\begin{aligned}
x' &= ax + by + c \\
y' &= dx + ey + f
\end{aligned}
\tag{1}
$$

where (x, y) are the original coordinates, (x', y') are the transformed coordinates, and a, b, d, e, c, f are the transformation coefficients. The coefficients a, b, d, e determine a scaling and a rotation, and the coefficients c and f determine a translation. The result of the application of the affine transformation of Eq. (1) in combination with simplex optimization (described below) is shown in Fig. 5. Since the affine transformation is linear, the transformed objects are linearly resized and rotated. This can be observed from the edges of the fan-shape region bounding box (the white box in Fig. 5). After the transformation the edges are still straight lines, however, the corner angles are different from 90 degrees and the length of the lines is also linearly scaled.

2.3.2. Nonlinear Simplex Optimization

The Nelder and Mead[14,15] nonlinear simplex optimization is used in order to adjust the coefficients a, b, c, d, e and f and to warp the fan-shaped template in order to maximize the correlation between the template and a breast structure on the prior mammogram. This optimization defines a hyper polygon. For each vertex an error function is calculated. Then the polygon is "rolled" towards the minimum. The movement of the polygon (toward the minimum) is obtained by the reflection in the direction opposite to the vertex with maximal error. Fig. 5 shows the result of application of the affine transformation whose coefficients were obtained by the nonlinear simplex optimization.

Fan-shape
template
(x,y)

Warped
fan-shape
template
(x',y')
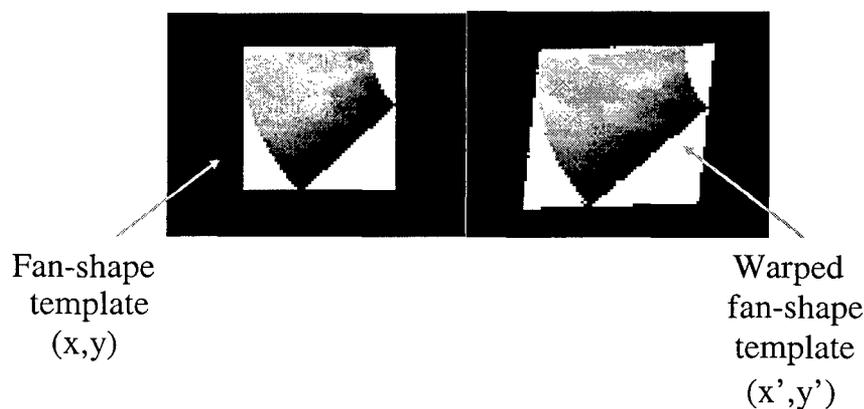
Figure 5. Fan-shaped template and warped fan-shaped template by the affine transformation.

2.4. Mass template alignment and identification of corresponding lesion

In this stage a new search region with a reduced size is defined on the prior mammogram (Fig. 6). A template containing the mass is extracted from the current mammogram. Then, the mass location on the prior mammogram is determined by maximizing the correlation between the template and a structure within the search region (Fig. 7).
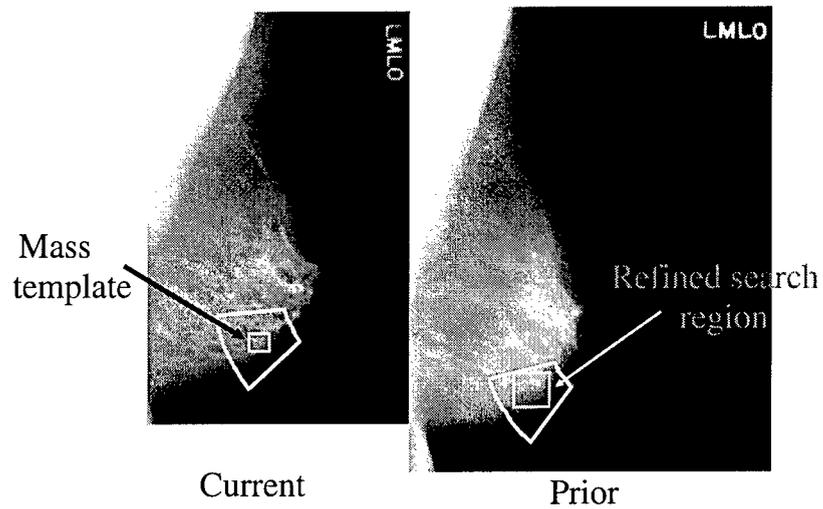
Figure 6. A refined search region is defined on the prior mammogram. A search for the best match between the mass template from the current mammogram and a structure on the prior mammogram was carried out within the refined search region.
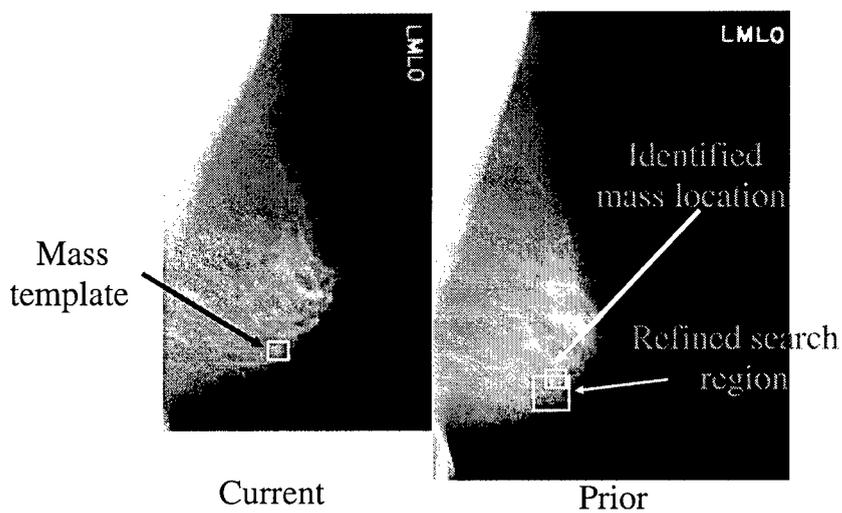
Figure 7. Final identification of the corresponding mass on the prior mammogram.

3. DATA SET

A set of 124 temporal pairs of mammograms containing biopsy-proven masses on the current mammograms was used to examine the performance of this approach. A total of 221 mammograms from 43 cases were digitized. Thirty five of the mammograms were digitized with a LUMISYS DIS-1000 laser scanner at a pixel resolution of 100 μm × 100 μm and 4096 gray levels. The digitizer was calibrated so that gray level values were linearly proportional to the optical density (OD) within the range of 0.1 to 2.8 OD units, with a slope of 0.001 OD/pixel value. Outside this range, the slope of the calibration curve decreased gradually. The OD range of the digitizer was 0 to 3.5. The remaining 186 mammograms were digitized with a LUMISCAN 85 laser scanner at a pixel resolution of 50 μm × 50 μm and 4096 gray levels. The digitizer was calibrated so that gray level values were linearly proportional to the OD within the range of 0 to 4 OD units, with a slope of 0.001 OD/pixel value. Output from both digitizers was linearly converted so that a large pixel value corresponded to a low optical density. In order to process the mammograms digitized with these two different digitizers, the images digitized with LUMISCAN 85 digitizer were averaged with a 16×16 box filter, resulting in 800 μm images. The images digitized with LUMISYS DIS-1000 digitizer were averaged with a 8×8 box filter, resulting in 800 μm images. The true lesion locations were identified by an experienced radiologist on all mammograms. The 221 mammograms contained 219 biopsy-proven and 2 follow-up masses. From all 124 temporal pairs of mammograms, 63 were CC-view pairs, 48 were MLO-view pairs, and 13 were lateral-view pairs.

4. EVALUATION METHODS

The accuracy of the multistage regional registration was analyzed in terms of two measures. The first measure is the overlap area between the estimated and the true lesions on the prior mammogram. The fractions of registered temporal pairs that could provide an accuracy of over 50% area overlap and over 75% area overlap were examined. The second measure is the average Euclidean distance between the centroids of the estimated and true lesion locations.

5. REGISTRATION RESULTS

In this study 86% of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations. The average distance between the estimated and the true centroids of the lesions on the prior mammogram was 4.4 ± 5.9 mm with a maximum of 30.6 mm. These results are presented in Table 1 and Table 2. For the 86% of the temporal pairs with 50% overlap, the average distance between the estimated and the true centroids of the lesions on the prior mammogram was 2.4 ± 2.1 mm with a maximum of 10.2 mm.

Table 1. The area overlap between the true and the estimated masses on the prior mammogram.

Pairs	50% overlap	75% overlap
Number	107	99
%	86%	80%

Table 2. The distance between the true and the estimated centroids of the mass on the prior mammogram.

	Overall	50% overlap	75% overlap
Mean distance	4.4 mm	2.4 mm	2.2 mm
Standard. Deviation.	5.9 mm	2.1 mm	2.0 mm
Max. distance	30.6 mm	10.2 mm	10.2 mm

6. CONCLUSION

The regional registration technique can localize the corresponding mass on the prior mammogram. The warping procedure based on an affine transformation in the local alignment stage reduces the size of the search region. Eighty-six percent of the

estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations. The average distance between the estimated and the true centroids of the lesions on the prior mammogram was 4.4 ± 5.9 mm. When the threshold was set to 75% area overlap, 80% of the temporal pairs could still exceed the threshold. The registration accuracy of the current method has been improved in comparison with that of our previous method,[10] although the data set was increased from 74 pairs to 124 pairs. This improvement is obtained mainly from the second stage affine transformation and simplex optimization. Further study is underway to develop a feature matching method to improve lesion localization within the search region.

REFERENCES

1. H. C. Zuckerman, "The role of mammography in the diagnosis of breast cancer," in *Breast Cancer, Diagnosis and Treatment*, edited by I. M. Ariel and J. B. Cleary (McGraw-Hill, New York, 1987), pp. 152-172.

2. L. Tabar and P. B. Dean, "The control of breast cancer through mammographic screening: What is the evidence," *Radiol. Clin. N. Amer.* **25**, pp. 993-1005, 1987.

3. L. W. Bassett, B. Shayestehfar and I. Hirbawi, "Obtaining previous mammograms for comparison: usefulness and costs," *Amer. J. Roentgenology* **163**, pp. 1083-1086, 1994.

4. E. A. Sickles, "Periodic mammographic follow-up of probably benign lesions: results in 3183 consecutive cases," *Radiology* **179**, pp. 463-468, 1991.

5. M. Sallam and K. Bowyer, "Detecting abnormal densities in mammograms by comparison with previous screenings" in *Digital Mammography '96*, edited K. Doi, M. L. Giger, R. M. Nishikawa and R. A. Schmidt (Elsevier, Amsterdam, 1996).

6. D. Brzakovic, N. Vujovic, M. Neskovic, P. Brzakovic and K. Fogarty, "Mammogram analysis by comparison with previous screenings" in *Digital Mammography '96*, edited by K. Doi, M. L. Giger, R. M. Nishikawa and R. A. Schmidt (Elsevier, Amsterdam, 1996).

7. N. Vujovic and D. Brzakovic, "Establishing the correspondence between control points in pairs of mammographic images," *IEEE Trans. Imag. Proc.* **6**, pp. 1388-1399, 1997.

8. S. Sanjay-Gopal, H. P. Chan, B. Sahiner, N. Petrick, T. Wilson, M. Helvie, "Evaluation of interval change in mammographic features for computerized classification of malignant and benign masses," *Radiology* **205(P)**, pp. 216, 1997.

9. S. Sanjay-Gopal, H. P. Chan, N. Petrick, T. Wilson, B. Sahiner, M. Helvie, M. Goodsitt, "A regional registration technique for automated analysis of interval changes of breast lesions," *Proc. SPIE* **3338**, pp. 118-131, 1998.

10. S. Sanjay-Gopal, H.P. Chan, T.E. Wilson, M.A. Helvie, N. Petrick, B. Sahiner, "A regional registration technique for automated interval change analysis of breast lesions on mammograms", *Medical Physics,* **26**: pp. 2669-2679, 1999.

11. L. Hadjiiski, H.P. Chan, B. Sahiner, N. Petrick, M.A. Helvie, S.S. Gopal, "Automated identification of breast lesions in temporal pairs of mammograms for interval change analysis", *Radiology,* **213(P)**: pp. 229-230, 1999.

12. W. Good, B. Zheng, Y.H. Chang, X. Wang, G. Maitz, "Generalized procrustean image deformation for substraction of mammograms", *Proc. SPIE* **3661**, pp. 1562-1573, 1999.

13. L. Quan, T. Kanade, "Affine structure from line correspondance with uncalibrated affine cameras", *IEEE Trans. Pat. Anal. Machine Intel.*, Vol. 19, No. **8**, Aug. 1997.

14. S.S. Rao, *Optimization: Theory and Applications*", Wiley Eastern Limited, 1979.

15. F.A. Lootsma, (ed). *Numerical methods for non-linear optimization*, Academic Press, New York, 1972.

ＣHICAGO World Congress on Medical Physics and Biomedical Engineering
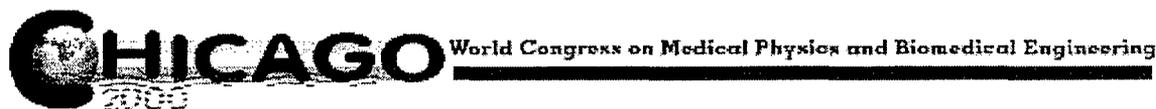
Abstract Submission - Status

Online help is available

Abstract #5945

The registration associated with this abstract has not been paid. This abstract will not be held in the program after May 15, 2000.
You may make changes to this registration or pay on-line by clicking here.

Converted documents have been viewed and approved. To change the document status, you must reset the documents and re-submit.

Click on the blue link(s) below to view converted PDFs

Abstract text
5945-96185   5945-96185.pdf

Status: Document received.   Reset Abstract Doc. This will delete your abstract.

Short Paper
5945-96300

Status: Not received yet.

Abstract Information
Click on headings below to edit information

Title: Regional registration of masses on current and prior mammograms using DWCE segmentation

Key Words: Computer-Aided Diagnosis
DWCE Segmentation
Interval Changes
Mammography

Track: 01 Diagnostic Physics, Medical Imaging, and Image Processing

Requested Presenting Mode: Oral

By-Line (Authors / Institutions): L Hadjiiski*, N Petrick, HP Chan, B Sahiner, M Helvie, C Zhou, M Gurcan, S Paquerault, The University of Michigan, Ann Arbor, Michigan, U.S.A, Department of Radiology, University of Michigan, Ann Arbor, MI 48109, USA

Authors: note: if you edit authors, you must edit by-line as well

Lubomir M Hadjiiski - Presenting Author

Nicholas Petrick - Author

Heang-Ping Chan - Author

Berkman Sahiner - Author

Mark A Helvie - Author

Chuan Zhou - Author

Metin N Gurcan - Author

Sophie Paquerault - Author

Corresponding Author Information:

Lubomir M Hadjiiski
Radiology, University of Michigan,
CGC B2103, 1500 E. Med. Center Dr
Ann Arbor, MI 48109-0904 United States

Phone: 7346478552

Fax: 7346478557

Email: lhadjisk@umich.edu

Get Acrobat

Click here to log off the system.

AbstractID: 5945 Title: Regional registration of masses on current and prior mammograms using DWCE segmentation

We are developing a computerized method to analyze interval change of mammographic lesions. In this study, we investigated the use of the density-weighted contrast enhancement (DWCE) technique to improve the localization of the corresponding mass on the prior mammogram.

A regional registration technique was developed for identifying masses on temporal pairs of mammograms. The breast images from the current and prior mammogram were first aligned globally based on the mutual information. An initial fan-shape search region was defined on the prior mammogram using a polar coordinate system with the origin at the nipples. The location of the fan-shaped region was refined by affine transformation and simplex optimization. The DWCE technique was then used to segment dense structures within the search region. A search for the best match between the lesion template from the current mammogram and a structure on the prior mammogram was performed within the DWCE segmented densities. The technique was evaluated on 124 temporal pairs of mammograms containing biopsy-proven masses. The true corresponding mass location on the prior mammogram was identified by an experienced radiologist. It was found that 85% of the estimated mass locations resulted in an area overlap of at least 50% with the true mass locations. The average distance between the estimated and the true centroid of the lesions on the prior mammogram was 2.6mm (std 2.1mm) for this subset and was 4.7mm (std 6.2mm) over the entire data set. The DWCE segmentation improved the accuracy of matching by directing the search to the dense structures and thus reducing the chance of mismatch.

ABSTRACT SUBMISSION FOR RSNA 2000
86th Scientific Assembly and Annual Meeting

The corresponding author (to receive confirmation of receipt and acceptance/rejection notice) is:
Lubomir M Hadjiiski PhD

Radiology
The University of Michigan
CGC B2 103
1500 E. Med Cntr Dr
Ann Arbor, MI 48109-0904   USA
Tel: 734-647-8552
Fax: 734-647-8557
lhadjisk@umich.edu

Presentation Type: Scientific Presentation
Category: Physics: Image Processing, CAD, etc
Keyword 1: Breast radiography, technology
Keyword 2: Computers, diagnostic aid
Keyword 3: Images, analysis

Work supported by grant from RSNA: No
Work presented previously: No
Paper submitted for publication: No
Principal Investigator: Other
Presentation Preference: Scientific Paper

Authors
1   Lubomir M Hadjiiski PhD   **Phone:** 734-647-8552 **Fax** 734-647-8557
        SSN: ▓▓▓▓E-Mail: lhadjisk@umich.edu **Member Number:**   **Conflict:** No
2   Heang-Ping Chan PhD   **Phone:** 734-936-4357 **Fax** 734-936-7948
        SSN: E-Mail: chanhp@umich.edu **Member Number: Conflict:** No
3   Berkman Sahiner PhD   **Phone:** 734-647-7429 **Fax** 734-647-8557
        SSN: E-Mail: berki@umich.edu **Member Number: Conflict:** No
4   Nicholas Petrick PhD   **Phone:** 734-647-7428 **Fax** 734-647-8557
        SSN: E-Mail: petrick@umich.edu **Member Number: Conflict:** No
5   Mark A Helvie MD   **Phone:** 734-936-4352 **Fax** 734-936-9723
        SSN: E-Mail: mahelvie@umich.edu **Member Number: Conflict:** No
6   Metin N Gurcan PhD   **Phone:** 734-647-8554 **Fax** 734-647-8557
        SSN: E-Mail: gurcan@umich.edu **Member Number: Conflict:** No

More than 6 co-authors: No
Presenter #: 1

Title:

Computer-Aided Classification of Malignant and Benign Breast Masses by Analysis of Interval Change of Features in Temporal Pairs of Mammograms

Abstract:

PURPOSE: To develop a computer-aided diagnosis method to assist radiologists in analysis of interval changes of corresponding masses on a temporal pair of mammograms.

METHOD/MATERIALS: An automated method was developed to extract and analyze features extracted from corresponding masses on a temporal pair of mammograms. Regions of interest containing the corresponding masses were identified on the current and prior mammograms of the temporal pair. The masses were automatically segmented using an active contour model. 20 Run Length Statistics (RLS) features, 3 spiculation features, and mass size were extracted from each mass. An additional 20 difference RLS features were obtained by subtracting the RLS features of the prior mass from those of the current mass for each temporal pair. The feature space for each temporal pair consisted of the RLS and spiculation features from both the prior and the current mammograms and the difference RLS features. Stepwise feature selection with simplex optimization was used to select the optimal feature subset. A linear discriminant classifier (LDA) was used to merge the selected features for classification of malignant and benign masses.

In this study, 93 temporal image pairs from 38 patients containing biopsy-proven masses on the current mammograms were chosen from patient files. The true mass locations were identified by an experienced radiologist on all mammograms. All cases were selected from a biopsy database so that interval change was observed for most of the masses even if they were found to be benign after biopsy. This was therefore a difficult data set for interval change analysis. A leave-one-case-out training and testing resampling scheme was used for feature selection and classification. The classification accuracy was analyzed by receiver operating characteristic (ROC) methodology.

RESULTS: An average of 9 features was selected from the 37 training subsets. The selected features included 3 difference RLS features, 3 RLS and 1 spiculation features from the current image, and 2 spiculation features from the prior. The classifier achieved a training Az of 0.92 and a test Az of 0.82.

CONCLUSIONS: The size of the mass is not a useful feature for difficult cases because many benign masses grow over time. The difference RLS and prior spiculation features are useful for identification of malignancy in temporal pairs of mammograms. Further studies are underway to improve the technique and to evaluate the performance on a larger data set.