

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 2002

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.

20020816 054

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to 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 2002	3. REPORT TYPE AND DATES COVERED Annual Summary (6 Apr 2001 - 5 Apr 2002)	
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) (abstract should contain no proprietary or confidential information) A new classification scheme was developed to classify mammographic masses as malignant and benign by using interval change information. The masses on both the current and the prior mammograms were automatically segmented using an active contour method. From each mass, texture morphological and spiculation features, were extracted. Additionally, difference features were obtained by subtracting the prior from the corresponding current features. The feature space consisted of the current, difference and prior features. Stepwise feature selection and linear discriminant analysis classification were used to select and merge the most useful features. A leave-one-case-out resampling scheme was applied to train and test the classifier using 140 temporal image pairs (85 malignant, 55 benign). An average of 10 features were selected from the 56 training subsets. The classifier achieved an average training A_z of 0.92 and a test A_z of 0.88. For comparison, a classifier was trained and tested using features extracted from the 120 current single images. This classifier achieved an average training A_z of 0.90 and a test A_z of 0.82. The information on the prior image significantly ($p=0.015$) improved the accuracy for classification of the masses. A pilot study was performed to evaluate the effects of computer classification on radiologists' estimates of the likelihood of malignancy of masses.				
14. Subject Terms (keywords previously assigned to proposal abstract or terms which apply to this award) Breast Cancer, Computer-aided diagnosis, Screening, Classification, Image Analysis			15. NUMBER OF PAGES 47	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

(3) Table of Contents

(1)	Front Cover.....	1
(2)	Standard Form (SF) 298, REPORT DOCUMENTATION PAGE	2
(3)	Table of Contents	3
(4)	Introduction	4
(5)	Body	5
(A)	Database collection and extraction of regions of interest	5
(B)	Further developments of methods for establishing corresponding locations in current and previous mammograms.....	5
(C)	Obtaining hand drawn mass boundaries from radiologists and evaluation of segmentation accuracy	6
(D)	Further develop methods for extracting morphological and texture features from masses segmented from ROIs extracted from current and prior mammograms.....	7
(E)	Analyze techniques for characterizing differences in these features	8
(F)	Evaluate the effectiveness of LDA classifiers and neural networks for classification	9
(G)	Evaluate the effectiveness of developed classifiers using receiver operating characteristic methodology	10
(H)	Identify the preferred features and classification methods.....	10
(I)	Compare the accuracy of computerized classification with the malignancy assessment of radiologists.....	11
(J)	Evaluate usefulness of temporal features for CAD by comparison of classification based on temporal features with classification based on features extracted from the current mammogram alone	12
(K)	Perform pilot ROC study for the design of a full-scale ROC experiment	13
(L)	Extension of the developed methods to the microcalcifications classification.....	14
(6)	Key research accomplishments in current year as a result of this grant	14
(7)	Reportable Outcomes	14
(8)	Conclusion.....	15
(9)	References	17
(10)	Appendix	19

(4) 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.

(5) Body

In the fourth year (4/6/01-4/5/02) of this grant, we have performed the following studies:

(A) Database collection and extraction of regions of interest (Task 1)

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 is identified by an Mammography Quality Standards Act (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 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 Breast Imaging Reporting and Data System (BI-RADS) lexicon.

(B) Further developments of methods for establishing corresponding locations in current and previous mammograms (Task 3)

We continued to improve our regional registration technique [3-6]. In the first step, we are still working on development of an automated method that will detect the nipple location in the breast image. The method is based on both the change of tangential direction and the change in the tissue density along the breast border.

In the third and final step we have designed a new method - an adaptive similarity measure (ASM), to improve automated identification of corresponding lesions on prior mammograms.

We are developing a new class of similarity measures (SM). It combines adaptive filtering to enhance the lesion and a SM as a figure-of-merit (FOM) measure. The filters are designed with a training set to maximize and minimize the FOM for the similar and dissimilar image pairs, respectively, by using a gradient optimization technique. The ASM is applied to the final stage of our multistage regional registration technique for mass identification on the prior mammogram. A search for the best match between the lesion template from the current mammogram and a structure on the prior mammogram is carried out within a search region, guided by the ASM.

This new approach was evaluated by using 179 temporal pairs of mammograms containing biopsy-proven masses.

We found that 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 lesion

centroids on the prior mammogram was 4.5 ± 6.7 mm. In comparison, the correct localization and the average distance using a conventional correlation SM were 84% and 4.9 ± 7.0 mm, respectively.

The preliminary results of this study are promising. The ASM improved the identification of the corresponding lesions on temporal pairs of mammograms. The average Euclidean distance between the computer estimate of the corresponding structure and the radiologist-identified location and their standard deviation were both reduced when compared to multistage registration using a conventional correlation SM.

We will present the preliminary results on this improved method at the International Workshop for Digital Mammography (IWDM), Bremen, Germany, June 22 - 25, 2002 [22].

In the previous years, when we increased the data set from 124 to 179 temporal pairs the detection accuracy was slightly reduced. For 124 temporal pairs 87% of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations [5], [6]. For 179 pairs it was reduced to 84 %. The average distance between the estimated and the true lesion centroids on the prior mammogram for 124 temporal pairs was 4.2 ± 5.7 mm. For 179 temporal pairs the average distance was 4.9 ± 7.0 mm. The main reason for the reduction of the detection accuracy was due to the more difficult 55 new temporal pairs. They included subtle masses surrounded by breast densities with brighter mammography appearance. Those structures were making the detection more difficult. A way to overcome this problem was to continue to improve the detection methods. We introduced density-weighted contrast enhancement (DWCE) technique [7] to improve the localization of the corresponding mass on the prior mammogram. The average distance between the estimated and the true centroid of the lesions on the prior mammogram was 4.8 ± 6.9 mm, which was slightly improved, however 84% of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations and it did not show improvement. By using the ASM method (described above) both the correct localization and the average distance were 86% and 4.5 ± 6.7 mm respectively, which is improved result compared to conventional and DWCE method.

We will continue our studies to improve the technique, expand it to different types of SM, and evaluate its accuracy on a larger data set.

(C) Obtaining hand drawn mass boundaries from radiologists and evaluation of segmentation accuracy (Task 9)

As we reported before we carried out an evaluation of the segmentation technique by comparison of the computer segmentations (K-means clustering [8] and active contour [9-11]) with hand segmentations using the expertise of the radiologist.

Obtaining hand drawn mass boundaries from radiologists

An MQSA-approved radiologist experienced in breast imaging outlined the mass boundaries of the masses on 239 regions of interest (ROI)s using an interactive image analysis tool on a UNIX workstation.

In the future year more MQSA-approved radiologists experienced in breast imaging will hand segment mass boundaries of the masses on the ROIs.

Formulate quantitative measures for assessing segmentation accuracy

For the purpose of our accuracy analysis, the radiologist's hand segmentations were used to compare with the computer segmentations. Three quantitative measures were used for evaluation of the accuracy of the computer segmentations: Hausdorff distance, average Hausdorff distance and the area overlap measure.

Hausdorff distance between two curves is defined as the maximum of the closest point distances (DCPs) between the two curves [12], [13]. The closest point distance (DCP) associates each point on both curves to a point on the other curve, and the Hausdorff distance finds the largest distance between the associated points. The average Hausdorff distance, on the other hand, finds the average of the DCPs between the two curves.

Area overlap is defined as follows:

$$A_{overlap} = \frac{A_1 \cap A_2}{A_1 \cup A_2} 100,$$

where A_1 is area inside the hand segmented mass outline and A_2 is area inside the computer segmented mass outline.

Evaluate quantitatively the accuracy of computer boundary segmentation using radiologists hand segmentation

Here we summarize again the segmentation results. For this study we used 239 ROI. All results presented in the following are average results for this 239 ROIs.

The results of the first stage of segmentation by the K-means clustering algorithm are the following: average area overlap of 40%, average Hausdorff distance of 5.58 mm, and average Hausdorff distance of 2.19 mm averaged over 239 ROIs.

The results of the active contour segmentation are the following: average Area overlap of 67%, average Hausdorff distance of 4.49 mm, and average Hausdorff distance of 1.27 mm averaged over 239 ROIs.

The active contour segmentation improved the segmentation accuracy. The above results confirm the visual satisfactory agreement between the active contour segmentation and radiologist's hand segmentation.

Evaluate qualitatively the computer segmentation by means of an observer study

Our experience showed that radiologists gave very subjective estimation with large variation within the data set when they visually evaluated the computer segmentation. We found that radiologist's hand segmentation and a quantitative comparison, described above, is a better way to evaluate the accuracy of the computer segmentation. This quantitation study would replace the visual qualitative evaluation proposed in our original research plan.

(D) Further develop methods for extracting morphological and texture features from masses segmented from ROIs extracted from current and prior mammograms (Task 10)

As we reported previously, we develop methods for extraction of texture, morphological and spiculation features from the segmented masses based on active contour segmentation [9-11]. Since the feature extraction is very important for the classification and we used the methods intensively during the past year, here we will summarize the methods.

The texture features used in this study were calculated from run-length statistics (RLS) matrices [14]. The RLS matrices were computed from the images obtained by the rubber band straightening transform (RBST)[15]. The RBST maps a band of pixels surrounding the mass onto the Cartesian plane (a rectangular region). In the transformed image, the mass border appears approximately as a horizontal edge, and spiculations appear approximately as vertical lines.

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 values of the filtered image. Five texture measures, namely, short run emphasis (SRE), long run emphasis (LRE), gray level nonuniformity (GLN), run length nonuniformity (RLN), and run percentage (RP) were extracted from the vertical and horizontal gradient images in two directions, $\theta = 0^\circ$, and $\theta = 90^\circ$. Therefore, a total of 20 RLS features were calculated for each ROI. The definition of the RLS feature measures can be found in the literature [14].

The morphological features were extracted from the automatically segmented mass shape. Five of morphological features were based on the normalized radial length (NRL), defined as the Euclidean distance from the object's centroid to each of its edge pixels, i.e., the radial length, and normalized relative to the maximum radial length for the object [16]. The following five NRL features were extracted: mean (NRLAVG), standard deviation (NRLSD), entropy (NRLENT), area ratio (NRLAREAR), zero crossing count (NRLZCC). In addition, the perimeter (PERIM), area (AREA), circularity (CIRC), rectangularity (SQR), contrast (CONT), perimeter-to-area ratio (CRR) and Fourier descriptor (FF) features were extracted. The detailed definition of the morphological features can be found in [17], [18].

A spiculation measure was defined for each pixel on the mass border by using the statistics of the image gradient direction relative to the normal direction to the mass border in a ring of pixels surrounding the mass [17], [19]. The spiculation measure for each border pixel was normalized to be between 0 and $\pi/2$, with a value of $\pi/4$ indicating a random orientation of image gradients, and larger values indicating a higher likelihood of spiculation. Three features were extracted from the spiculation measure. The first feature (AVG) was the average of the spiculation measure for all pixels on the mass boundary. The second feature (PERC_ABV) was the percentage of border pixels with a spiculation measure larger than $\pi/4$, and the third feature (AVE_ABV) was the average of the spiculation measure for those pixels with a spiculation measure larger than $\pi/4$.

A total of 35 features (20 RLS, 12 morphological and 3 spiculation) were therefore extracted from each ROI.

More detailed description of the above can be found in [23].

(E) Analyze techniques for characterizing differences in these features (Task 11)

Additionally, difference features were obtained by subtracting a prior feature from the corresponding current feature. Therefore, 35 difference features were derived from the 20 RLS, 12 morphological and 3 spiculation features.

We designed a new classification scheme allowing direct merge of current and prior information. The input feature space to the classifier includes the current, prior and difference features. This allows the classifier to choose the individual current and prior features or the difference features. Stepwise feature selection with simplex optimization is used to select the optimal feature subset. A linear discriminant classifier (LDA) is used to merge the selected features for classification of malignant and benign masses. A leave-one-case-out training and testing resampling scheme is used for feature selection and classification.

We have published the results on this method in Medical Physics Journal [23] as well as we presented the results at the RSNA meeting [20] and SPIE meeting [21].

(F) Evaluate the effectiveness of LDA classifiers and neural networks for classification (Task 12)

At this stage of our study we used 57 biopsy proven masses (33 malignant and 24 benign) in the 56 cases. The 241 mammograms contained different mammographic views (CC, MLO, and lateral views) and multiple examinations of the masses including the examination when the biopsy decision was made. By matching masses of the same view from two different examinations, a total of 140 temporal pairs were formed, of which 85 were malignant and 55 benign. A malignant temporal pair consisted of a biopsy proven malignant mass or a mass that was initially not recommended for biopsy and later found to be malignant by biopsy in a future year. A similar definition was used for the benign temporal pairs. Within a pair, the current mammogram was defined as the mammogram with the later date, and the prior mammogram was defined as the one with the earlier date. Therefore, in cases with three consecutive exams, more than one temporal pair could be formed and two of the mammograms could be called "current". Among the 140 temporal pairs, we had 120 unique current mammograms. Of the masses in the 120 current mammograms, 70 were malignant and 50 benign.

The current, prior, and difference features formed a multidimensional feature space for the classification task. Stepwise feature selection applied to linear discriminant analysis (LDA) was used to select the most useful features. The selected features were then used as the input predictor variables for the LDA classifier. The classifier was trained and tested by a leave-one-case-out resampling scheme. A case was considered to contain all ROIs from a given patient. In each resampling step, the temporal pairs from 55 cases were used for feature selection and formulation of the linear discriminant function, while the temporal pairs from the left-out case were used for testing the trained classifier. A total of 56 training and testing steps were obtained from the 56 cases. The classification results from the 56 test cases were accumulated to evaluate the classifier performance. Since the data set in this study was still small, we chose the feature selection parameters such that the dimensionality of the input feature vector for the LDA classifier was small in order to reduce the possibility of over-training.

To evaluate the improvement in the classifier performance designed by using the temporal change information, two additional classifiers with different input features were obtained. One of them was trained using the information extracted from the current single images of the temporal pairs. The other classifier was trained using the information extracted from the prior single images of the temporal pairs. Comparison of the three classifiers will reveal the effectiveness of interval change analysis for the classification of malignant and benign masses.

In this specific study we decided to use LDA classifier in order to have linear combination among the features. The neural network classifier (NNC) combines the input features in nonlinear way, which will make the analysis and comparisons more complicated. The use of NNC will involve uncertainties for the structure of the neural network (number of hidden layers, number of neurons in the hidden layers), number of iteration for training the NNC and how to have NNC not overtrained. One of initial aims of the recent study was to find the useful subset of features. For this purpose we extensively used the feature selection, which in case of NNC is very computationally intensive.

At this stage of our research we wanted to have simple classification method which allowed us to find useful features, and to design and compare different classifiers with different input

feature spaces (classifiers based on current and prior images) efficiently. This is the reason to concentrate mainly on the use of the LDA classification schemes.

(G) Evaluate the effectiveness of developed classifiers using receiver operating characteristic methodology (Task 13)

To evaluate the classifier performance, the training and test discriminant scores were analyzed using receiver operating characteristic (ROC) methodology [26]. The discriminant scores of the malignant and benign masses were used as decision variables in the LABROC1 program [27], which fits a binormal ROC curve based on maximum likelihood estimation. The classification accuracy was evaluated as the area under the ROC curve, A_z . The performances of the classifiers were also assessed by estimating the partial area index ($A_z^{(0.9)}$). 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 , $(1-TPF_0)$. 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 a cancer detection task.

The performances of the classifiers based on the temporal pairs, the current images, and the prior images are summarized in Table 1. The classifiers that achieved the highest test A_z values with a small average number of features were presented here. Table 2 is a summary of the features selected for each classifier.

For the 56 training subsets of temporal pairs used in this study, an average of 10 features were selected for the classification task. The most frequently selected features included 4 difference RLS features (3 SRE and 1 LRE), 4 RLS features (2 SRE, 1 RLN and 1 RP), 1 spiculation feature from the current image, and 1 spiculation feature from the prior image (Table 2). The LDA classifier achieved an average training A_z of 0.92 and a test A_z of 0.88. The test partial $A_z^{(0.9)}$ was 0.37.

Table 1. Classification results for the classifier based on the temporal change information, the classifier based on current single image information, and the classifier based on prior single image information.

Classification	Avg. No. of selected features	Training A_z	Test A_z	Test partial $A_z^{(0.9)}$
Temporal pairs	10	0.92	0.88 ± 0.03	0.37 ± 0.10
Current images	11	0.90	0.82 ± 0.04	0.32 ± 0.08
Prior images	4	0.78	0.76 ± 0.04	0.24 ± 0.08

(H) Identify the preferred features and classification methods (Task 14)

Texture and spiculation features were important for malignant and benign classification of mammographic masses for all three types of classifiers: the classifier based on temporal pair information, the classifier based on current image information, and the classifier based on prior image information [23]. One or more of the spiculation features were always selected in all training partitions for all three classifiers. The most frequently selected texture features were the

short run emphasis (SRE) features. They comprised more than 50 % of the texture features selected for the three classifiers (Table 2).

Temporal-information-based classifier showed improved performance compared to the classifiers based on current or prior image information alone. The input feature space to the temporal-information-based classifiers included the current, prior, and difference features. This allows the classifier to choose the individual features or the difference features. Using the stepwise feature selection procedure and the linear discriminant classifier, it was found that the texture and the spiculation features contained useful temporal information to perform malignant and benign mass classification. Texture features appeared to provide the best information by the difference features obtained from subtracting the prior from the corresponding current features (SRE and LRE difference features). On the other hand, the best use of the spiculation features appeared to be a direct combination of current and prior features in the input feature vector by the LDA since the individual features were chosen.

We found that better feature subsets could be selected by the stepwise feature selection in the subspaces than in the entire feature space. For example, for the temporal-information-based classifier, a better feature subset with a higher test A_z at 0.88 was found when the input feature space included only the texture and spiculation subspaces. The addition of the morphological feature subspace to the input feature space reduced the highest test A_z to 0.84. Similarly, in the case of the classifier based on prior image information, a better feature subset was obtained when the texture and spiculation feature subspaces were used in the input feature space for stepwise feature selection. Again the addition of the morphological feature subspace to the input feature space reduced the highest test A_z to 0.72. The classifier based on current image information was the only one, among the three, that obtained a better result, as shown in Table 1, when the morphological feature subspace was included in the input feature space.

One reason for the poor performance of the morphological features may be due to the fact that the masses were more subtle in the prior images. In fact, the experienced MQSA mammographer was not confident in seeing 25 of the "masses" on the prior images and could not provide a mass size estimation for them. Although the active contour model would stop the iteration based on the preset criteria and found an "outline" of the masses on the prior mammograms, generally these mass outlines were less reliable than those on the current masses in providing morphological characteristics of the masses. Texture features did not depend as strongly on the precise mass boundary as morphological features. Three out of the four features selected for classification of the malignant and benign masses on the prior images were RLS texture features. A spiculation feature was also found to be a good discriminator.

In this study, we employed a simple measure of temporal change by taking the difference between the feature from the current mass and the corresponding feature from the prior mass. We observed improvement in classification with this simple temporal information. It will be important to evaluate other similarity measures that can characterize small difference in image features of the object of interest. It can be expected that a more sensitive similarity measure will provide a better measurement of dissimilarity, or difference, between the current and prior masses and further improve the utilization of the temporal change information on mammograms.

(I) Compare the accuracy of computerized classification with the malignancy assessment of radiologists (Task 15).

We performed ROC analysis of the malignancy confidence ratings provided by the experienced MQSA radiologist for the current image data set (120 images) [23]. The radiologist estimated the likelihood of malignancy of the current masses on a 10-point confidence scale (1 –

definitely benign and 10 – definitely malignant) based on the 120 current mammograms alone without comparison with the prior.

The malignancy ratings resulted in an A_z value of 0.80 ± 0.04 . This indicates that the masses in the current mammograms cannot be easily distinguished as malignant or benign even by an experienced radiologist, consistent with the fact that all lesions had indeed undergone biopsy. The classifier based on the current image information has an A_z value of 0.82 ± 0.04 , similar to the accuracy of the radiologist for this data set.

Table 2. Selected features for classifiers based on temporal pairs, current images, and prior images. The letter “H” or “V” at the beginning of the texture feature labels indicates that the features were extracted from the horizontal or vertical gradient magnitude images, respectively. The number (0 or 90) at the end of the texture feature labels shows the direction at which the features were extracted.

Feature type	Group	Features	Temporal pairs			Current images	Prior images
			Curr	Pr	Diff	Curr	Pr
Texture	SRE	H_SRE_0			X		
		H_SRE_90	X		X		
		V_SRE_0	X		X	X	X
		V_SRE_90				X	
	LRE	V_LRE_0			X		X
		H_LRE_0				X	
	RLN	V_RLN_0	X			X	
	RP	H_RP_0	X				X
Spiculation		PERC_ABV	X			X	
		AVG		X			
		AVG_ABV					X
Morphological		CRR				X	
		NRLZCC				X	
		PERIM				X	
		NRLAVG				X	
		SQR				X	
		CONT				X	

(J) Evaluate usefulness of temporal features for CAD by comparison of classification based on temporal features with classification based on features extracted from the current mammogram alone (Task 15)

For classification of malignant and benign masses using the current single images (the 120 current images of the temporal pairs), the LDA classifier selected an average of 11 features for the 56 training subsets [23]. The most frequently selected features were 4 RLS features (2 SRE, 1 LRE and 1 RLN), 1 spiculation feature, and 6 morphological features (Table 2). The classifier achieved an average training A_z of 0.90, a test A_z of 0.82, and a test partial $A_z^{(0.9)}$ of 0.32.

For the classification of masses based on the prior single images alone, an average of 4 features were selected for the 56 training subsets. The most frequently selected features were 3 RLS features (1 SRE, 1 LRE, and 1 RP) and 1 spiculation feature. The LDA classifier achieved an average training A_z of 0.78, test A_z of 0.76, and test partial $A_z^{(0.9)}$ of 0.24.

The difference in the test A_z between the classifier based on the temporal pairs and that based on the current images alone is statistically significant ($p=0.015$). The difference in the test A_z between the classifier based on the temporal pairs and that based on the prior images alone is also statistically significant ($p=0.001$). The partial area index for the classifier based on the temporal pairs is also improved compared to the classifiers based on the current or the prior images alone, although the differences did not achieve statistical significance.

(K) Perform pilot ROC study for the design of a full-scale ROC experiment (Task 15 – final step of the project).

We have performed a pilot study as a first step to design an observer performance experiment with ROC methodology to evaluate the effects of computer classification on radiologists' estimates of the likelihood of malignancy of masses. A graphical user interface was developed on a PC to display side-by-side the temporal pairs of masses in a predesigned random order for each observer. The likelihood of malignancy and the BI-RADS assessment of the radiologist on each pair is automatically recorded when they select it on a slider.

253 temporal image pairs (136 malignant and 117 benign) from 95 patients containing masses on serial mammograms were chosen from patient files and digitized. Additional pairs containing normal structures were also included to simulate a more realistic clinical situation. The true mass locations were identified by an experienced radiologist on all mammograms. Regions of interest containing the corresponding masses were then extracted from the current and prior mammograms of each temporal pair and analyzed by the CAD program. All cases eventually underwent biopsy 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.

Two radiologists assessed the temporal pairs that were displayed on the display PC workstation. They provided estimates of the likelihood of malignancy and BI-RADS assessment without and then with CAD. The reading order of the temporal pairs was randomized for each observer. The classification accuracy was quantified by using the area under ROC curve, A_z .

For this data set, the computer classifier achieved a test A_z value of 0.86. The radiologists' A_z values for the likelihood of malignancy were 0.72 and 0.74 without CAD, and improved to 0.76 and 0.75, respectively, with CAD. The improvement was statistically significant ($p=0.0006$) for the first radiologist. For the BI-RADS assessments, the two radiologists obtained A_z values of 0.67 and 0.77 without CAD and improved to 0.73 and 0.79, respectively, with CAD. The improvements were also statistically significant ($p<0.001$).

This pilot study indicates that CAD using interval change analysis may be useful for assisting radiologists in classification of masses and thereby reducing unnecessary biopsies.

This pilot study will be the basis for our design of a full-scale ROC study. We have already recruited 6 radiologists to participate as observers. The results, described above, show that the study design will likely produce statistically significant results. The sample size is acceptable but we are continuing to enlarge the data set until the ROC study design is finalized. We expect that this ROC study can be completed within the no cost time extension year that we requested. This type of observer study is new and unique and the outcome will be important, providing a new

understanding of the potentials of computer aid to the radiologists in characterization of the temporal changes of mammographic masses.

(L) Extension of the developed methods to the microcalcifications classification.

We were encouraged by the above results and we were able to start the transfer of the developed methods for detection and classification of temporal masses to detection and classification of temporal microcalcification clusters. We carried out a preliminary study, which showed promising results and we applied for an BCRP-01, IDEA grant at the U.S. Army Medical Research and Materiel Command. The research grant was approved and we are very enthusiastic and encouraged that we will have the opportunity to extend the already developed methods and design new methods for detection, classification and analysis of temporal microcalcification clusters.

Some preliminary results were presented at RSNA 2001 [24], and SPIE 2002 [25].

(6) Key research accomplishments in current year as a result of this grant

- Database collection and extraction of regions of interest (Task 1).
- Further development of methods for establishing corresponding locations in current and previous mammograms (Task 3).
- Obtaining hand drawn mass boundaries from radiologists and evaluation of segmentation accuracy (Task 9).
- Further develop methods for extracting morphological and texture features from masses segmented from ROIs extracted from current and prior mammograms (Task 10).
- Analyze techniques for characterizing differences in these features (Task 11).
- Evaluate the effectiveness of LDA classifiers and neural networks for classification (Task 12).
- Evaluate the effectiveness of developed classifiers using receiver operating characteristic methodology (Task 13)
- Identify the preferred features and classification methods (Task 14)
- Compare the accuracy of computerized classification with the malignancy assessment of radiologists (Task 15).
- Evaluate usefulness of temporal features for CAD by comparison of classification based on temporal features with classification based on features extracted from the current mammogram alone (Task 15)
- Perform pilot ROC study for the design of a full-scale ROC experiment (Task 15– final step of the project).
- Extension of the developed methods to the microcalcifications classification.

(7) Reportable Outcomes

Publications in current year as a result of this grant

- [1] L. Hadjiiski, H.P. Chan, B. Sahiner, N. Petrick, M. Helvie, “Automated Registration of Breast Lesions in Temporal Pairs of Mammograms for Interval Change Analysis – Local

- Affine Transformation for Improved Localization”, *Medical Physics*, 28 (6), June 2001, pp. 1070-1079.
- [2] L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M. Helvie, M. Gurcan, “Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses”, *Medical Physics*, 28 (11), November 2001, pp. 2309-2317.
- [3] L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvie, M. Gurcan, “Analysis of temporal change of mammographic features for computer-aided characterization of malignant and benign masses ”, Oral Presentation at *SPIE International Symposium on Medical Imaging*, San Diego, California, February 19-22, 2001, *Proc. SPIE Medical Imaging, 2001, 4322*, pp.661-666.
- [4] L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvie, “An Adaptive Similarity Measure for Automated Identification of Breast Lesions in Temporal Pairs of Mammograms for Interval Change Analysis”, To be presented at the 6th International Workshop for Digital Mammography (IWDM), Bremen, Germany, June 22 - 25, 2002, To appear in *Proc. IWDM 2002*.
- [5] L. Hadjiiski, H.P. Chan, N. Petrick, B. Sahiner, M. Gurcan, M.A. Helvie, et al, “Computerized Regional Registration of Corresponding Microcalcification Clusters on Temporal Pairs of Mammograms for Interval Change Analysis”, Presented at the 87th *Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA)*, Chicago, Illinois, November 25 - 30, 2001. *Radiology* 2001; 221 (P): 425.
- [6] L. Hadjiiski, H.P. Chan, M. Gurcan, B. Sahiner, N. Petrick, M.A. Helvie, M. Roubidoux “Computer-Aided Characterization of Malignant and Benign Microcalcification Clusters Based on the Analysis of Temporal Change of Mammographic Features”, Presented at the *SPIE International Symposium on Medical Imaging*, San Diego, California, February 23-28, 2002. To appear in *Proc. SPIE Medical Imaging 2002*.

Copies of publications are enclosed with this report.

(8) Conclusion

During this year, we have continued the development of the regional registration technique. The adaptive similarity measure (ASM) improves the localization of the corresponding mass on the prior mammogram. 179 temporal pairs of mammograms containing biopsy-proven masses were used for evaluation of the detection accuracy. 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 4.5 ± 6.7 mm. In comparison, the correct localization and the average distance using a conventional correlation similarity measure were 84% and 4.9 ± 7.0 mm, respectively. The registration accuracy of the current method has been improved in comparison with that without ASM. 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 the coming year.

To evaluate the accuracy of computer segmentation of masses, 239 regions of interest containing the corresponding masses were identified by MQSA radiologist on the current and prior mammograms of the temporal pair. The masses were automatically segmented using a K-means clustering algorithm and active contour model. Additionally, hand drawn mass boundaries from radiologists were obtained and compared with the computer segmentations. The initial mass segmentation by the K-means clustering algorithm was satisfactory (average area overlap of 40%, average Hausdorff distance of 5.58 mm, and average Hausdorff distance of 2.19 mm averaged over 239 ROIs). The active contour model further improved the accuracy of mass segmentation (average Area overlap of 67%, average Hausdorff distance of 4.49 mm, and average Hausdorff distance of 1.27 mm averaged over 239 ROIs). The active contour model is therefore useful for precise mass segmentation.

For the task of feature extraction, we evaluated 35 features (20 texture, 12 morphological and 3 spiculation) extracted from each mass. Additional difference features were obtained by subtracting the features of the prior mass from those of the current mass. Therefore, 35 difference features were derived from the 20 texture, 12 morphological and 3 spiculation features. The feature space for each temporal pair consisted of the texture, spiculation and morphological features from both the prior and the current mammograms and the difference features. These features were evaluated for their effectiveness in classification of malignant and benign temporal masses as well as detection of temporal change.

We designed a new classification scheme allowing direct merge of current and prior information. The input feature space to the classifier included the current, prior and difference features. This allowed the classifier to choose the individual current and prior features or the difference features in order to obtain the best combination and merge of the features for high classification accuracy and optimal detection of interval change. It was found that the difference RLS texture features and spiculation features were useful for identification of malignancy in temporal pairs of mammograms. The information on the prior image was important for characterization of the masses; 5 out of the 10 selected features contained prior information. We found that the mass size descriptors were not discriminatory features for these difficult cases because many of the benign masses also grew over time. In comparison with the classification based on image information from the current images alone, the temporal change information significantly ($p=0.015$) improved the accuracy for classification of the masses in terms of the total area under the ROC curve (A_z). The partial area under the ROC curve for the classifier based on the temporal pairs ($A_z^{(0.9)} = 0.37$) is also improved compared to the classifier based only on the current images ($A_z^{(0.9)} = 0.32$), although the difference did not achieve statistical significance.

We performed a pilot study for the design of observer performance experiments with ROC methodology to evaluate the effects of computer classification on radiologists' estimates of the likelihood of malignancy of masses. Two radiologists read a data set of temporal pairs. For this data set, the computer classifier achieved a test A_z value of 0.86. The radiologists' A_z values for the likelihood of malignancy were 0.72 and 0.74 without CAD, and improved to 0.76 and 0.75, respectively, with CAD. The improvement was statistically significant ($p=0.0006$) for the first radiologist. For the BI-RADS assessments, the two radiologists obtained A_z values of 0.67 and 0.77 without CAD and improved to 0.73 and 0.79, respectively, with CAD. The improvements were also statistically significant ($p<0.001$). This pilot study will be the basis for our design of a full-scale ROC study to evaluate the effects of CAD of interval changes on the performance of radiologists.

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.

We were able to start the transfer of the developed methods for detection and classification of temporal masses to detection and classification of temporal microcalcification clusters. We carried out a preliminary study, which showed promising results and we applied for an IDEA grant at the U.S. Army Medical Research and Materiel Command. The research grant was approved and we will have the opportunity to extend the already developed methods and design new methods for detection, classification and analysis of temporal microcalcification clusters.

(9) 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] 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.
- [4] 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.
- [5] 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., *Proc. SPIE Medical Imaging, 2000*, 3979, pp.847-853.
- [6] L. Hadjiiski, H.P. Chan, B. Sahiner, N. Petrick, M. Helvie, "Automated Registration of Breast Lesions in Temporal Pairs of Mammograms for Interval Change Analysis – Local Affine Transformation for Improved Localization", *Medical Physics*, 28 (6), June 2001, pp. 1070-1079.
- [7] 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", Presented at *The World Congress on Medical Physics and Biomedical Engineering*, Chicago, July 23 - 28, 2000.
- [8] B. Sahiner, H. P. Chan, N. Petrick, D. Wei, 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 on mammograms," *Med. Phys.* 23, 1671-1684 (1996).
- [9] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Int. J. Comput. Vision* 1, 321-331 (1987).
- [10] L. D. Cohen, "On active contour models and balloons," *CVGIP: Image Understanding* 53, 211-218 (1991).

- [11] D. J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," *CVGIP: Image Understanding* 55, 14-26 (1992).
- [12] D.P. Huttenlocher, G.A. Klanderman and W.J. Ruclidge, "Comparing images using the Hausdorff distance", *IEEE Trans. Pattern Anal. Mach. Int.* 15, 850-863 (1993).
- [13] V. Chalana and Y.Kim, "A methodology for evaluation of image segmentation algorithms on medical images", *Proc. SPIE* 2710, 178-189 (1996).
- [14] M. M. Galloway, "Texture classification using gray level run lengths," *Comp. Graph. Img Proc.* 4, 172-179 (1975).
- [15] B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," *Med. Phys.* 25, 516-526 (1998).
- [16] J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computer-aided image analysis," *IEEE Trans. Med. Img.* 12, 664-669 (1993).
- [17] B. Sahiner, H.-P. Chan, N. Petrick, M. A. Helvie, and L. M. Hadjiiski, "Improvement of mammographic mass characterization using spiculation measures and morphological features," *Med. Phys.* (submitted), (2000).
- [18] N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," *Med. Phys.* 26, 1642-1654 (1999).
- [19] B. Sahiner, H. P. Chan, N. Petrick, L. M. Hadjiiski, M. A. Helvie, and S. Paquerault, "Active contour models for segmentation and characterization of mammographic masses," *The 5th International Workshop on Digital Mammography Proc. IWDM-2000*, (in press) (2000).
- [20] 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", Presented at the *86th Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA)*, Chicago, Illinois, November 26 - December 1, 2000. *Radiology* 2000; 217(P): 435.
- [21] L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvie, M. Gurcan, "Analysis of temporal change of mammographic features for computer-aided characterization of malignant and benign masses ", Oral Presentation at *SPIE International Symposium on Medical Imaging*, San Diego, California, February 19-22, 2001, *Proc. SPIE Medical Imaging, 2001*, 4322, pp.661-666.
- [22] L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvie, "An Adaptive Similarity Measure for Automated Identification of Breast Lesions in Temporal Pairs of Mammograms for Interval Change Analysis", To be presented at the *6th International Workshop for Digital Mammography (IWDM)*, Bremen, Germany, June 22 - 25, 2002, To appear in *Proc. IWDM 2002*.
- [23] L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M. Helvie, M. Gurcan, "Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses", *Medical Physics*, 28 (11), November 2001, pp. 2309-2317.
- [24] L. Hadjiiski, H.P. Chan, N. Petrick, B. Sahiner, M. Gurcan, M.A. Helvie, et al, "Computerized Regional Registration of Corresponding Microcalcification Clusters on Temporal Pairs of Mammograms for Interval Change Analysis", Presented at the *87th Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA)*, Chicago, Illinois, November 25 - 30, 2001. *Radiology* 2001; 221 (P): 425.

- [25] L. Hadjiiski, H.P. Chan, M. Gurcan, B. Sahiner, N. Petrick, M.A. Helvie, M. Roubidoux "Computer-Aided Characterization of Malignant and Benign Microcalcification Clusters Based on the Analysis of Temporal Change of Mammographic Features", Presented at the *SPIE International Symposium on Medical Imaging*, San Diego, California, February 23-28, 2002. To appear in *Proc. SPIE Medical Imaging 2002*.
- [26] C. E. Metz, "ROC methodology in radiologic imaging," *Invest Radiol* 21, 720-733 (1986).
- [27] C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Stat. Med.* 17, 1033-1053 (1998)

(10) Appendix

Copies of publications are enclosed with this report.

Automated registration of breast lesions in temporal pairs of mammograms for interval change analysis—local affine transformation for improved localization

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

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

(Received 15 August 2000; accepted for publication 9 April 2001)

Analysis of interval change is important for mammographic interpretation. The aim of this study is to evaluate the use of an automated registration technique for computer-aided interval change analysis in mammography. Previously we developed a regional registration technique for identifying masses on temporal pairs of mammograms. In the current study, we improved lesion registration by including a local alignment step. Initially, the lesion position on the prior mammogram was estimated based on the breast geometry. An initial fan-shaped search region was then defined on the prior mammogram. In the second stage, the location of the fan-shaped region on the prior mammogram was refined by warping, based on an affine transformation and simplex optimization in a local region. 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 and an average distance of 2.4 ± 2.1 mm between their centroids. The average distance between the estimated and the true centroid of the lesions on the prior mammogram over all 124 temporal pairs 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 in accuracy resulted from the improved geometry estimation and the local affine transformation. © 2001 American Association of Physicists in Medicine.

[DOI: 10.1118/1.1376134]

Key words: mammography, interval change, computer-aided diagnosis, breast cancer, affine transformation

I. INTRODUCTION

Mammography is currently the most effective method for early breast cancer detection.^{1,2} One of the important techniques used by radiologists in mammographic interpretation to detect developing malignancy is analysis of interval changes.^{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 identified 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 the most recent and the 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 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.*⁸⁻¹⁰ and Hadjiiski *et al.*¹¹ have developed a multistage technique that defines the transformation to locally map the position of the mass on a current mammogram to that of 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 transforma-

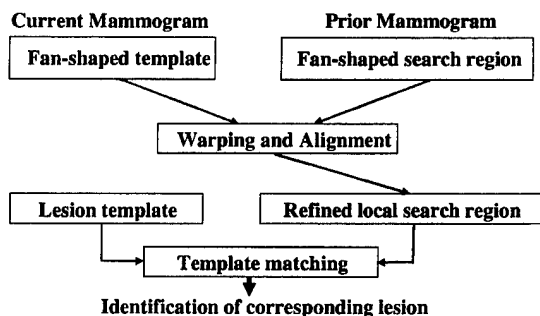


FIG. 1. Block diagram of the regional registration technique.

tion to map all points from the current mammogram onto a prior mammogram. The current mammogram is then subtracted from the prior mammogram.

The goal of our research is to develop a technique for computerized analysis of temporal differences between a mass 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. When fully developed, the technique will be applied to a mass on the current mammogram either identified by the radiologist or by an automated mass detection program, thus the interval change analysis can be an integrated part of an automated CAD system. In this study, we focused on the development of an automated registration technique that localizes the corresponding mass on the prior mammogram when the mass on the current mammogram is known. Therefore, we used radiologist-identified mass location on the current mammogram as a starting point and that on the prior mammogram as the ground truth for evaluation of the registration technique. A local registration technique was developed based on an affine transformation and simplex optimization and its usefulness in improving the localization of the mass on the prior mammogram was investigated.

II. REGISTRATION TECHNIQUE

A multistage regional registration technique was developed for identifying corresponding masses on temporal pairs of mammograms. The block diagram of the regional registration technique is shown in Fig. 1. In the first stage, an initial fan-shaped search region was defined on the prior mammogram based on the mass location on the current mammogram. In the second local alignment stage, the location of the search region on the prior mammograms was first refined by maximizing a correlation measure between a template of the fan-shaped region centered at the mass 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 and further improve the correlation. 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

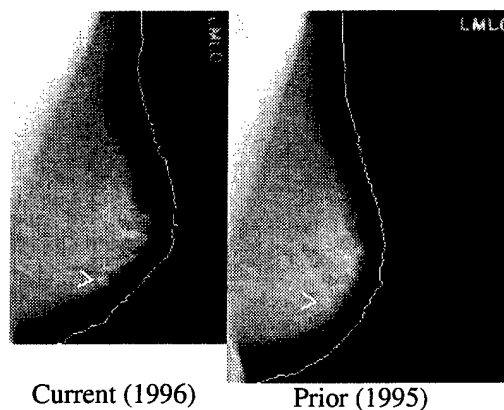


FIG. 2. An example of a pair of current and prior mediolateral oblique mammograms in our data set. The arrows point to the masses on the current and the prior mammograms. The white lines represent the breast boundary determined by the automated boundary detection procedure.

search region. A more detailed explanation for each of the stages will be presented in the following subsections.

A. Stage 1—Initial estimate of search region

We have modified our previous method to define a fan-shaped search region on the prior mammogram. Initially an automated procedure is used to detect the breast boundary on the mammograms (Fig. 2). The location of the mass on the current mammogram is determined in a polar coordinate system with the nipple as the origin. By using the radial distance R_{curr} between the nipple and mass centroid, $|NM|$, an arc is drawn which intersects the breast boundary at points **A** and **B** (Fig. 3). Three angles are estimated at the radial distance R_{curr} : The angle β between NM and NA , the angle φ between NM and NB , and the angle θ between NA and NB ($\theta = \beta + \varphi$). The location of the mass is determined by R_{curr} and the angle β or φ . The angle θ is the breast width at the radial distance R_{curr} . Using the radial distance R_{curr} to draw an arc centered at the nipple centroid on the prior mammogram, N' , the two intersect points A' and B' with the breast boundary on the prior mammogram are determined. The

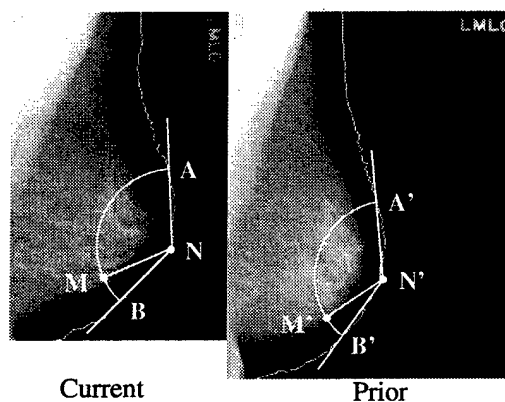


FIG. 3. Initial estimation of the mass location on the prior mammogram, based on the nipple-mass centroid distance and an angular distance from the breast periphery on the current mammogram.

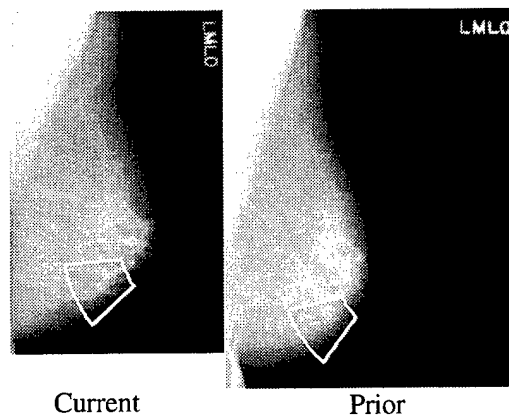


FIG. 4. Definition of an initial fan-shaped search region on the prior mammogram and a fan-shaped template on the current mammogram.

angle θ_p between the axes $|N'A'|$ and $|N'B'|$ is estimated. An angular scaling factor α can be calculated as the ratio of the prior and the current angles, $\alpha = \theta_p / \theta$.

In order to predict the angular location of the mass on the prior mammogram, the smaller angle between β and φ is selected as the angular coordinate of the mass on the current mammogram. The smaller angle is used because we found by experiment that it produces a smaller angular deviation error than using the larger angle. The angular deviation error is defined as the angle between the axis connecting the nipple and the true mass centroid and the axis connecting the nipple and the predicted mass centroid on the prior mammogram. The selected angle, multiplied by the angular scaling factor α , is used as the predicted angle from the corresponding axis on the prior mammogram. The radial distance R_{curr} is used to predict the radial position of the mass on the prior mammogram.

An initial fan-shaped search region is then defined on the prior mammogram centered at the predicted location of the mass centroid (Fig. 4). The size of the fan-shaped region is estimated previously¹⁰ to have the form $\epsilon = k_1 + k_2 / R_{curr}$ and $\delta = k_3$, where 2ϵ determines the angular width and 2δ determines the radial length of the fan-shaped region. The constants k_1, k_2 , and k_3 were chosen experimentally such that the estimated fan-shaped regions will essentially include all mass centroids on the prior mammograms. A fan-shaped template centered at the mass is also defined on the current mammogram. More details on defining the fan-shaped region can be found in Appendix A and in Ref. 10.

B. Stage 2—Refinement of search region by warping and alignment

The second stage combined two procedures. First, the location of the search region on the prior mammograms was refined by maximizing a correlation measure between the fan-shaped template extracted from the current mammogram and the breast structures on the prior mammogram. The template was shifted pixel by pixel within the initial fan-shaped search region and a correlation measure was calculated at each pixel location. The pixel location providing the maxi-

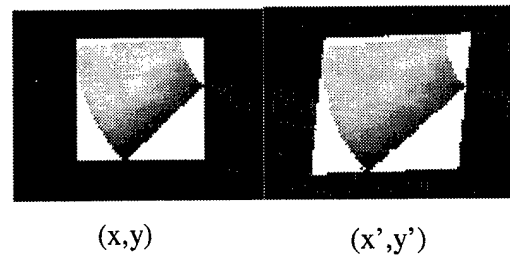


FIG. 5. The fan-shaped template (x,y) and the warped fan-shaped template (x',y') by the affine transformation.

mum correlation is used as the center of a refined search region. This is basically a template matching operation. Second, the affine transformation in combination with simplex optimization was iteratively used to warp the fan-shaped template and further maximize the correlation measure with the breast structures on the prior mammogram.

1. Affine transformation

An affine transformation¹³ is a linear transformation combining scaling, rotation, and translation. A two-dimensional affine transformation is defined as follows:

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

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 applying the affine transformation of Eq. (1) in combination with the simplex optimization (described below) to refine the fan-shaped search region is shown in Fig. 5. Since the affine transformation is linear, the transformed object is linearly resized and rotated. This can be observed from the edges of the bounding box of the fan-shaped region (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 lengths of the lines are linearly scaled.

2. Nonlinear simplex optimization

The nonlinear simplex optimization by Nelder and Mead^{14,15} is used to adjust the coefficients a, b, c, d, e , and f and to warp the fan-shaped template, thereby maximizing 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. The polygon is then "rolled" towards the minimum. The movement of the polygon (towards the minimum) is obtained by reflection in the direction opposite to the vertex with the maximal error. Figure 5 shows the result of application of the affine transformation whose coefficients were obtained by the nonlinear simplex optimization. A more detailed discussion on this optimization method can be found in Appendix B and Refs. 14 and 15.

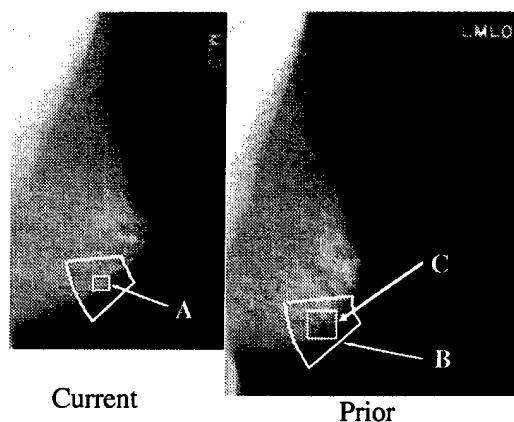


FIG. 6. A refined search region was defined on the prior mammogram. A search for the best match between the mass template from the current mammogram and a structure on the prior mammogram was carried out within the refined search region. (A—mass template on current mammogram, B—warped fan-shaped region from current mammogram, C—refined search region).

3. Stage 3—Mass template matching and localization of corresponding lesion

At this stage a new search region with a reduced size is defined on the prior mammogram (Fig. 6). The reduced size of the search region is determined experimentally by iterative adjustment of the size of the rectangular region targeting the improvement of the final result. A template containing the mass is extracted from the current mammogram. The mass location on the prior mammogram is then determined by maximizing the correlation between the template and a structure within the search region (Fig. 7).

III. 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. Different mammographic views of the same breast were also included. There were a total of 221 mammograms obtained from 54 cases. Temporal pairs were formed using the temporal se-

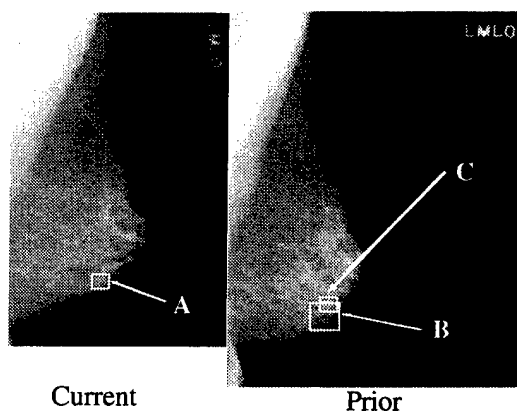
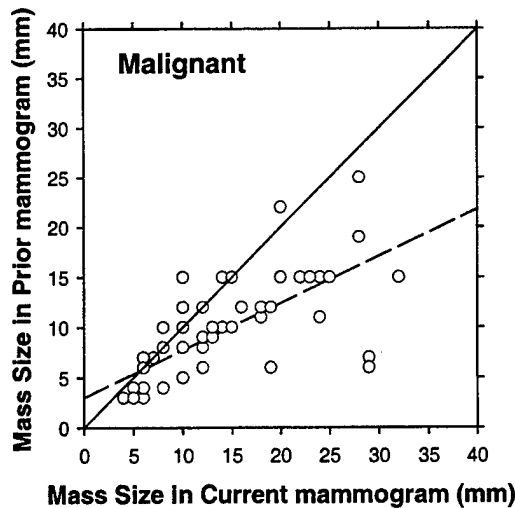


FIG. 7. Final identification of the corresponding mass on the prior mammogram. (A—Mass template on current mammogram, B—Refined search region, C—Identified mass location).

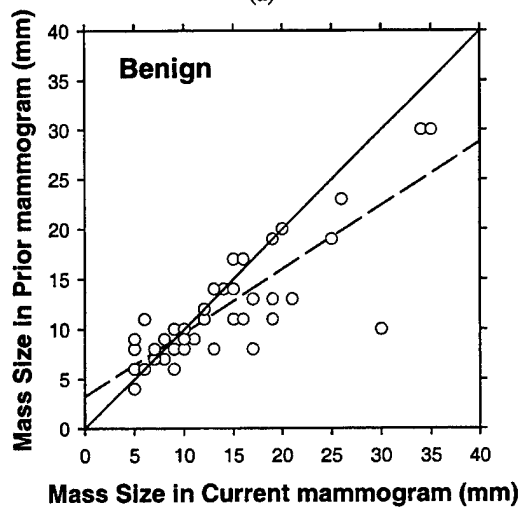
quence from the corresponding view. Some cases contained mammograms of multiple years and a combination of the mammograms from different prior years with the current-year mammogram formed multiple temporal pairs. Thirty five of the mammograms were digitized with a LUMISYS DIS-1000 laser scanner at a pixel resolution of $100 \mu\text{m} \times 100 \mu\text{m}$ 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 0.1–2.8 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 0–3.5. The remaining 186 mammograms were digitized with a LUMISCAN 85 laser scanner at a pixel size of $50 \mu\text{m} \times 50 \mu\text{m}$ and 4096 gray levels. The digitizer was calibrated so that the gray level values were linearly proportional to the OD within the range of 0–4 OD units, also with a slope of 0.001 OD/pixel value. Output from both digitizers was linearly converted so that large pixel value corresponded to a low-optical density. In order to process the mammograms digitized with these two different digitizers, the images were first averaged using a filter that has constant weights over the entire filter kernel and then were down-sampled. This filter will be referred to as a box filter. The images digitized with the LUMISCAN 85 digitizer were averaged with a 16×16 box filter and then were down-sampled by a factor of 16. The images digitized with the LUMISYS DIS-1000 digitizer were averaged with an 8×8 box filter and then were down-sampled by a factor of 8. Therefore, all resulting images had a pixel size of $800 \mu\text{m} \times 800 \mu\text{m}$.

The 54 cases contained 53 biopsy proven and one follow-up masses. The 221 mammograms contained different mammographic views and multiple years of the masses including the year when the biopsy was performed. Of the 124 temporal pairs of mammograms 73 were malignant and 51 benign. A malignant temporal pair consists of a biopsy proven malignant mass or a mass that was followed up and was found to be malignant when a biopsy was performed in a future year. Of the 124 temporal pairs of mammograms, 63 were CC-view pairs, 48 were MLO-view pairs, and 13 were lateral-view pairs. A Mammography Quality Standards Act (MQSA)-approved radiologist read the original mammogram to identify the mass and provide description of its characteristics. The radiologist defined a bounding box around the mass and marked the nipple location on every film.

The radiologist also measured the mass sizes, defined as the longest dimension of the mass, both on the current and prior mammograms. In Figs. 8(a) and 8(b) the mass sizes on the current mammograms were plotted against those on the prior mammograms for the malignant and the benign temporal pairs, respectively. Only 103 temporal pairs were plotted (54 malignant and 49 benign) due to the fact that the masses on the prior mammograms in the remaining 21 temporal pairs were too subtle for the radiologist to estimate their boundaries. On average the malignant masses appear to have a larger increase in size than the benign masses. The mean increase in size from prior to current for the malignant masses is 4.2 mm compared to 1.6 mm for the benign masses ($p=0.008$). The correlation coefficient is 0.71 for the malig-



(a)

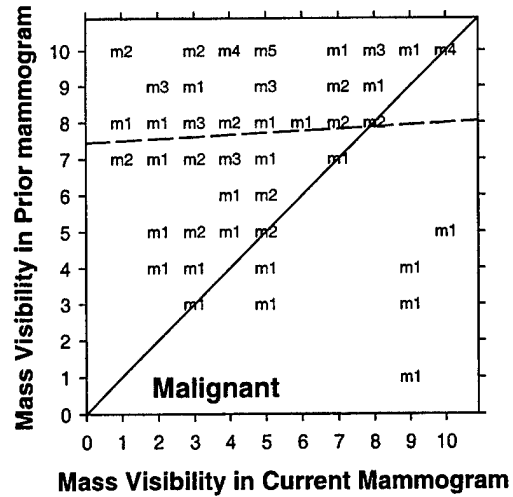


(b)

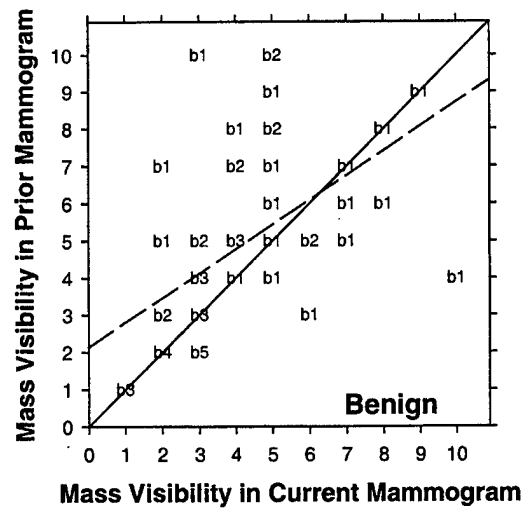
FIG. 8. Mass sizes measured by an MQSA-approved radiologist on the current mammograms plotted against those on the prior mammograms for (a) 54 malignant and (b) 49 benign temporal pairs. The diagonal line on the graph represents the case when the current and the prior mass sizes are identical. The dashed lines are the linear regression lines defined by $y = 0.469x + 3.012$ for (a) and by $y = 0.638x + 3.242$ for (b). The correlation coefficient for malignant masses is 0.71 and for benign masses is 0.83.

nant masses and 0.83 for the benign masses [Fig. 8(a) and 8(b)].

The radiologist also rated the visibility of the masses on the mammograms relative to those encountered in clinical practice on a 10-point scale, with one represents the most obvious and 10 the subtlest masses. The visibility of the masses on the current mammogram is plotted against those on the prior mammogram in Fig. 9 for the 73 malignant and 51 benign temporal pairs. Generally, the malignant masses were less visible on the prior mammograms while the visibility of the benign masses was found to be more similar. The mean difference in visibility between the prior and the current mammograms for the malignant masses is 2.8 compared to 0.7 mm for the benign masses ($p=0.0002$). The correlation coefficient is 0.06 for malignant masses and 0.54 for benign masses [Figs. 9(a) and 9(b)]. For most of the



(a)



(b)

FIG. 9. Visibility of the masses on the current mammogram plotted against those on the prior mammogram for (a) malignant and (b) benign temporal pairs. The visibility was rated on a 10-point discrete scale (1=most obvious, 10=subtlest). Because many of the data points overlap, we indicate the number of points with the same rating by a number next to the symbol (*m* or *b*). The diagonal line on the graph represents the case when the current and the prior mass sizes are identical. The dashed lines are the linear regression lines defined by $y = 0.055x + 7.44$ for (a) and by $y = 0.658x + 2.138$ for (b). The correlation coefficient for malignant masses is 0.06 and for benign masses is 0.54.

temporal pairs the time interval between the current and the prior mammogram was 12 months (Fig. 10).

IV. 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 the true lesion locations.

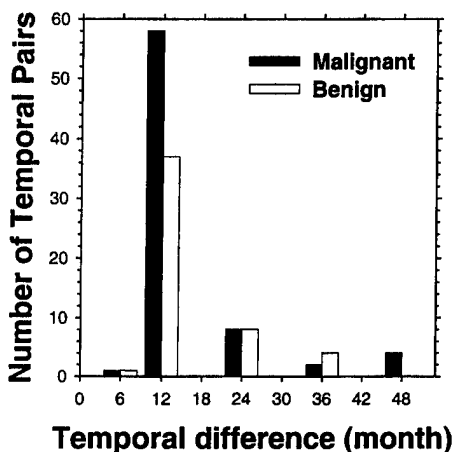


FIG. 10. Temporal interval between the current and the prior mammograms for the 124 temporal pairs in our data set.

V. REGISTRATION RESULTS

A. Stage 1—Initial estimate of search region

At this stage an initial estimation of the mass location on the prior mammogram was carried out based on the geometrical position of the mass on the current mammogram. Based on observation of the radial deviation errors and the angular deviation errors, the fan-shaped search region was estimated to be $\epsilon = 0.25 + 5/R_{\text{curr}}$ radians and $\delta = 20$ mm. This definition of the fan-shaped search region resulted in an average search area of 1462 mm^2 on the prior mammograms. For the 124 temporal image pairs used in this study, the Euclidean distance between the initial estimate of the centroid location of the corresponding structure on the prior mammogram and the center of the bounding box of the mass provided by the radiologist was estimated. For the 124 temporal image pairs, the average Euclidean distance error of the initial estimate was 8.4 ± 5.4 mm. The error distributions for both the malignant and the benign pairs are shown in Fig. 11. At this initial stage, 57% of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations and 27% resulted in an area overlap of at least 75% (Fig. 12).

B. Stage 2—Refinement of search region by warping and alignment

At the second stage, the location of the search region on the prior mammogram was first refined by maximizing a correlation measure between the fan-shaped 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. For the 124 temporal image pairs, the average Euclidean distance error after the second stage was 7.5 ± 5.4 mm. At this stage, 59% of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations, and 36% resulted in an area overlap of at least 75%. The average Euclidean distance error at this

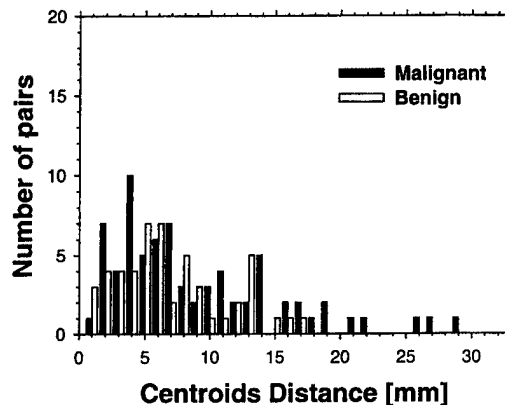


FIG. 11. Distribution of Euclidean distance error between the initial estimate of the mass centroid location on the prior mammogram and the center of the bounding box of the mass provided by the radiologist for the malignant and benign pairs after the first detection stage.

stage was reduced compared to that of the first stage, however, it did not achieve statistical significance ($p=0.07$).

After the simplex optimization, the search region was reduced to a constant size of $24 \text{ mm} \times 24 \text{ mm}$ ($=576 \text{ mm}^2$) centered at the refined fan-shaped region for every prior mammogram.

C. Stage 3—Mass template matching and localization of corresponding lesion

At this 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. This template matching resulted in 87% of the estimated lesion locations having an area overlap of at least 50% with the true lesion locations. The distributions of the Euclidean error for the malignant and the benign temporal pairs are shown in Fig. 13. The average distance between the estimated and the true centroids of the lesions on the prior mammogram for all 124 pairs was 4.2 ± 5.7 mm with a maximum of 31.6 mm. These results are summarized in Table I. For the 87% of the temporal pairs with 50% overlap, the

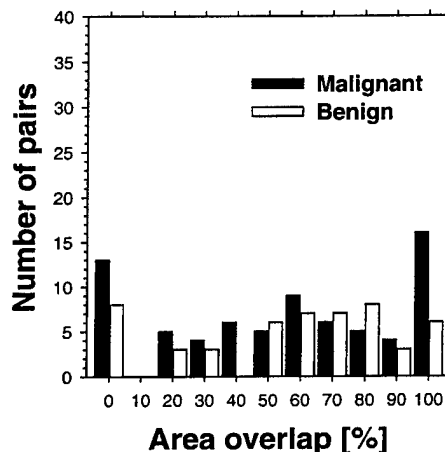


FIG. 12. Distribution of the area overlap between the estimated and the true lesion locations for 124 temporal pairs after the first detection stage.

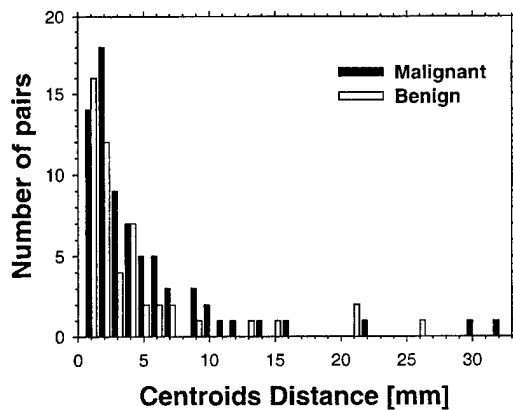


FIG. 13. Distribution of Euclidean distance error between the estimate of the mass centroid location on the prior mammogram and the center of the bounding box of the mass provided by the radiologist for the malignant and benign pairs after the final detection stage.

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. When a more stringent criterion of 75% overlap is imposed, 82% of the masses on the prior mammograms are considered to be localized (Fig. 14). For the 82% of the temporal pairs with 75% overlap, the average distance between the estimated and the true centroids of the lesions on the prior mammogram was 2.2 ± 1.9 mm with a maximum of 10.2 mm. The average Euclidean distance error at this stage was significantly reduced compared to the error of the first stage ($p=0.000001$) and the error of the second stage ($p=0.000001$).

D. Study of the importance of the stage 2 procedures

The effect of the two procedures at Stage 2 on the registration accuracy was studied. We removed them one at a time and evaluated the registration results. When the first correlation procedure was removed, the average Euclidean distance error increased to 5.6 ± 8.2 mm in the final stage. Only 81% of the estimated lesion locations resulted in an area overlap of at least 50% with the true lesion locations and 75% resulted in an area overlap of at least 75% with the true lesion locations. When the second warping procedure was removed, the average Euclidean distance error increased to 5.0 ± 6.3 mm in the final stage. Only 82% of the estimated

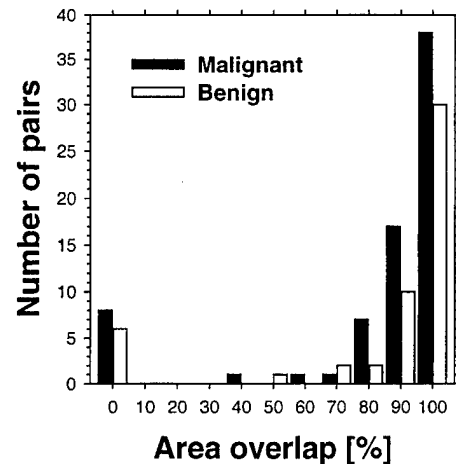


FIG. 14. Distribution of the area overlap between the estimated and the true lesion locations for 124 temporal pairs after the final detection stage.

lesion locations resulted in an area overlap of at least 50% with the true lesion locations and 76% resulted in an area overlap of at least 75% with the true lesion locations.

VI. DISCUSSION

The approach proposed here has simplified the first stage compared to our previous method.¹⁰ In the previous method, the distances between the nipple and the breast centroid on the current and prior mammograms were determined and used to estimate a radial scaling factor. The angular location of the mass was measured from the nipple–breast centroid axis. A global alignment procedure was used for determination of the breast centroids. With our new approach we eliminated the scaling for the radial distance between the nipple and the mass location of the prior mammogram. The breast periphery was used as a reference for the estimation of the angular position of the mass. Therefore, there was no need to determine the breast centroids on the current and the prior mammograms and the global alignment procedure could be eliminated. This is possible because the local alignment step provides better compensation for the displacement of the corresponding masses on the current and the prior mammogram caused by different compression and positioning of the breast.

It was found that the estimation of the angular position from the breast periphery allowed more precise localization of the mass position on the prior mammogram compared to our previous method where the angular position of the mass was estimated based on the nipple–breast centroid axis.¹⁰ There is a large variability in the estimation of the breast centroid location because the extend of the breast imaged on the mammogram at the chest wall and at the axillary tail in the MLO view depends on the breast positioning and compression. This causes an uncertainty in defining the region to calculate the breast centroid. In the previous study using 74 temporal pairs, the estimated Euclidean distance error at the first stage was 9.8 ± 6.0 mm. The fan-shaped search region was defined as $\epsilon = 0.35 + 5/r$, resulting in an average area of 1865 mm^2 for the fan-shaped search region. In the current

TABLE I. The Euclidean distance between the true and the estimated centroids of the mass on the prior mammogram for the three detection stages.

		Overall	50% overlap	75% overlap
Stage 1	Mean distance	8.4 mm	5.6 mm	4.5 mm
	Standard. Deviation.	5.4 mm	2.8 mm	2.6 mm
	Max. distance	29.0 mm	16.2 mm	13.8 mm
Stage 2	Mean distance	7.5 mm	4.9 mm	3.9 mm
	Standard. Deviation.	5.4 mm	3.0 mm	2.6 mm
	Max. distance	32.0 mm	16.9 mm	11.6 mm
Stage 3	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

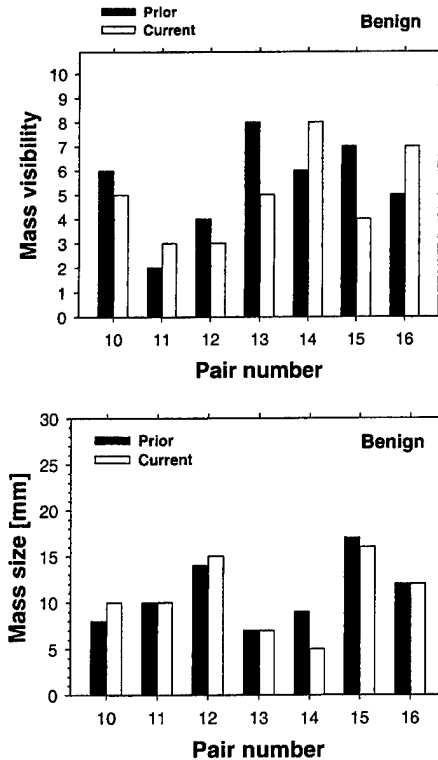
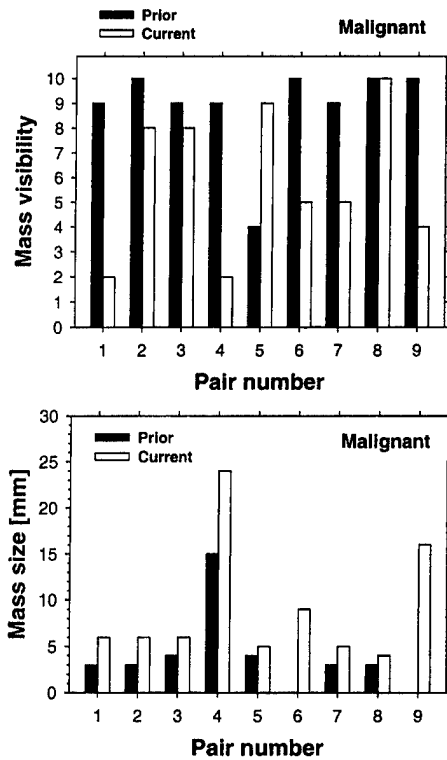


FIG. 15. The visibility and the mass size of nine malignant temporal pairs having area overlap less than 50%. The radiologist was unable to define the prior mass sizes of pairs 6 and 9 due to the subtlety of these masses.

FIG. 16. The visibility and the mass size of seven benign temporal pairs having area overlap less than 50%.

study, the estimated Euclidean distance error at the first stage was reduced to 8.4 ± 5.4 mm even though the data set was increased to 124 temporal pairs of mammograms. This allows the fan-shaped region to be reduced to $\epsilon = 0.25 + 5/r$, resulting in an average fan-shaped search area of 1462 mm^2 on the prior images. The reduction of the search area improves the chance of correctly localizing the mass on the prior mammogram.

The second stage combined two procedures: First the localization of the search region on the prior mammograms was refined by maximizing a correlation measure between the fan-shaped 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 and locally align the template with the breast structures. Both procedures improved the detection process. When one of these procedures was removed the registration results deteriorated, as discussed in the Results section.

With these improvements, the accuracy of the current regional registration technique is improved over the previous method.¹⁰ The current technique produced an average Euclidean distance error of 4.2 ± 5.7 mm, compared to 5.4 ± 7.5 mm when the previous technique was applied to the current data set. This difference is statistically significant ($p=0.03$). 82% of the estimated lesion locations resulted in an area overlap of at least 75% with the true lesion locations compared with 72% when applying the previous technique to the current data set. It is interesting to note that, of the 21

“masses” on the prior mammograms that the experienced radiologist could not confidently define the mass and measure its size, our registration technique localize 19 of them with an area overlap greater than 50%.

The average distance between the estimated and the true centroid of the lesions on the prior mammogram for the subset of temporal pairs having 50% overlap is about half of that of the entire data set (Table I). The maximum distance for this subset is about 1/3 of that for the entire data set.

With the current regional registration technique, 16 temporal pairs (13% of 124 temporal pairs) have an area overlap less than 50%. Twelve of the 16 computer estimated locations do not overlap at all with the radiologist's identified locations, and the other four pairs have an overlap between 1% and 49%. Seven of them are benign and nine are malignant. A major cause of the misregistration was that the mass was small and subtle and a breast structure within the search region had a higher correlation with the mass template from the current mammogram. Figures 15 and 16 show the visibility ratings and sizes of these misregistered masses. Eight of the nine misregistered malignant masses have visibility ratings of 9 or 10 and sizes below 5 mm. The misregistered benign masses are somewhat more obvious and larger in sizes than the malignant ones. Since many of the masses on the prior mammograms were not interpreted as a mass without reference to the current mammograms, the automatic registration with template matching would be difficult with these masses if the search region contains normal, but dense breast structures. We are currently investigating the application of local mass detection in the search region to focus

template matching to a few suspicious areas. Morphological and texture features will be extracted from the potential mass areas to provide additional matching information in the feature space.

The interval change analysis, when fully developed, will be one of the functions provided in an integrated CAD system. The mass on the current mammogram can be detected by an automated mass detection algorithm or identified by a radiologist. The CAD system will then analyze whether the mass is an existing or a newly developed lesion and will estimate its likelihood of malignancy. We are developing methods for characterization of malignant and benign masses based on analysis of interval changes in the mass features.¹⁶ Investigation of criteria to determine whether a mass exists on the prior mammogram is underway. If the mass is a newly developed lesion on the current mammogram, it will then undergo a single-exam analysis by the CAD system.

VII. CONCLUSION

We are developing an automated registration technique for analysis of interval change of a mass from a previous mammographic exam to the current one. In this study we found that a local affine transformation in combination with nonlinear simplex optimization can improve the localization and reduce the size of the search region. With the improved method, 87% of the estimated lesion locations in 124 randomly selected temporal pairs resulted in an area overlap of at least 50% with the true lesion locations. When the threshold for correct localization was set to 75% area overlap, 82% of the temporal pairs still exceeded this threshold. The average distance between the estimated and the true centroids of the lesions on the prior mammogram over all pairs was 4.2 ± 5.7 mm. The registration accuracy of the current method has been improved in comparison with that of our previous method¹⁰ even though 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. Additional studies are currently underway to develop a feature matching method to further improve lesion localization.

ACKNOWLEDGMENTS

This work is supported by a Career Development Award from the U.S. Army Medical Research and Material Command (DAMD 17-98-1-8211) (L.H.), a USPHS Grant CA 48129, a USAMRMC grant (DAMD 17-96-1-6254), and a USAMRMC Career Development Award DAMD 17-96-1-6012 (B.S.). The content of this publication does not necessarily reflect the position of the government and no official endorsement of any equipment or product of any companies mentioned in the publication should be inferred.

APPENDIX A: DEFINITION OF THE FAN-SHAPED REGION ON THE PRIOR MAMMOGRAM

Refer to Figs. 3 and 4, the fan-shaped region on the prior mammogram is drawn based on the nipple centroid on the prior mammogram, N' , as the center of the coordinate sys-

tem. The two bounding arcs are drawn using the radial distances $R_{\text{curr}} + \delta$ and $R_{\text{curr}} - \delta$, both centered at N' . The two sides of the fan-shaped region are bounded by two radial lines that form angles ϵ and $-\epsilon$ with the line $|N'M'|$. Thus the initial fan-shaped search region is centered as the predicted location of the mass centroid M' on the prior mammogram (Fig.4).

The constants k_1 , k_2 , and k_3 were chosen experimentally based on analysis of the angular deviation errors and the corresponding radial deviation errors for the 124 temporal pairs. The radial deviation error is defined as the difference between the predicted and the true distance of the mass from the nipple on the prior mammogram. The constants k_1 , k_2 are obtained in such a way that ϵ is the smallest upper bound that can enclose all angular deviation errors for all radial distances (R_{cur}) and all temporal pairs. The selection of the parametric form of ϵ was discussed in detail in Ref. 10. It reduced ϵ at larger R_{cur} . The constant k_3 was chosen to be equal to the maximum radial deviation error.

APPENDIX B: SIMPLEX OPTIMIZATION

An optimization problem can be defined as an error function that has to be minimized by iterative selection of the values of the function parameters n . We can define $n+1$ dimensional space, where n dimensions (degree of freedom) correspond to the error function parameters, and one dimension is the error function itself. When the optimization function is calculated for all possible values of the n parameters, and error surface in $(n+1)$ -dimensional space will be obtained. Usually the error functions for the real world applications are complex and nonlinear and the corresponding error surfaces contain local minima.

The nonlinear simplex optimization by Nelder and Mead^{14,15} defines a hyper-polygon with $n+1$ vertexes in a $(n+1)$ dimensional space. For each vertex the error function is calculated. The polygon is then "rolled" towards the minimum. The movement of the polygon (towards the minimum) is obtained by reflection in the direction opposite to the vertex (K) with the maximal error. To achieve this the center of masses (L) of the hyper-polygon vertexes is calculated. A line KL connects the center of the masses with the vertex with the maximal error. The new vertex (K') is obtained by central projection of the vertex K on the line KL with center L and $|K'L| = t|KL|$. The coefficient t determines how far the new vertex will be projected and what the corresponding size of the hyper-polygon will be. The larger the hyper-polygon is, the easier it will avoid ("roll over") the local minima on the error surface. However, it will be difficult to get close to the global minimum if its size is too large. On the other hand, although a small hyper-polygon will allow it to get to a close proximity to the global minimum, it is more likely to be trapped in a local minimum. The magnitude of the coefficient t is controlled adaptively by the Nelder and Mead algorithm. In case a large reduction in the error is detected for the new vertex, the magnitude of t is increased. In case the error is found to be increased for the new vertex, the magnitude of t is decreased.

In this paper, the nonlinear simplex optimization by Nelder and Mead was used to adjust the coefficients a , b , c , d , e , and f and to warp the fan-shaped template, thereby maximizing the correlation (C) between the template and a breast structure on the prior mammogram. Therefore, the dimensionality of the space was 7: Six parameters to be adjusted and the error function to be minimized was defined as $1 - C$.

^{a)} Author to whom correspondence should be addressed. Telephone: (734) 647-8552. Fax: (734) 647-8557. Electronic mail: lhadjisk@umich.edu

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**, 993-1005 (1987).

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

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

⁵ 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).

⁶ D. Brzakovic, N. Vujovic, M. Neskovic, P. Brzakovic, and K. Fogerty, "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. Image Process.* **6**, 1388-1399 (1997).

⁸ 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).

⁹ 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).

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

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

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

¹³ L. Quan and T. Kanade, "Affine structure from line correspondence with uncalibrated affine cameras," *IEEE Trans. Pat. Anal. Machine Intel.* **19(8)**, 834-845 (1997).

¹⁴ S. S. Rao, *Optimization: Theory and Applications* (Wiley, New York, 1979).

¹⁵ *Numerical Methods for Non-Linear Optimization*, edited by F. A. Lootsma (Academic, New York, 1972).

¹⁶ L. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. Gurcan, "Computer-aided classification of malignant and benign breast masses by analysis of interval change of features in temporal pairs of mammograms," *Radiology* **217(P)**, 435 (2000).

Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses

Lubomir Hadjiiski,^{a)} Berkman Sahiner, Heang-Ping Chan, Nicholas Petrick, Mark A. Helvie, and Metin Gurcan

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

(Received 22 May 2001; accepted for publication 27 August 2001)

A new classification scheme was developed to classify mammographic masses as malignant and benign by using interval change information. The masses on both the current and the prior mammograms were automatically segmented using an active contour method. From each mass, 20 run length statistics (RLS) texture features, 3 speculation features, and 12 morphological features were extracted. Additionally, 20 difference RLS features were obtained by subtracting the prior RLS features from the corresponding current RLS features. The feature space consisted of the current RLS features, the difference RLS features, the current and prior speculation features, and the current and prior mass sizes. Stepwise feature selection and linear discriminant analysis classification were used to select and merge the most useful features. A leave-one-case-out resampling scheme was used to train and test the classifier using 140 temporal image pairs (85 malignant, 55 benign) obtained from 57 biopsy-proven masses (33 malignant, 24 benign) in 56 patients. An average of 10 features were selected from the 56 training subsets: 4 difference RLS features, 4 RLS features, and 1 speculation feature from the current image, and 1 speculation feature from the prior, were most often chosen. The classifier achieved an average training A_z of 0.92 and a test A_z of 0.88. For comparison, a classifier was trained and tested using features extracted from the 120 current single images. This classifier achieved an average training A_z of 0.90 and a test A_z of 0.82. The information on the prior image significantly ($p=0.015$) improved the accuracy for classification of the masses. © 2001 American Association of Physicists in Medicine. [DOI: 10.1118/1.1412242]

Key words: computer-aided diagnosis, interval change, classification, feature analysis, mammography, malignancy

I. 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 in mammographic interpretation to detect developing malignancy.^{3,4} A variety of computer-aided diagnosis (CAD) techniques have been developed to detect abnormalities and to distinguish malignant and benign lesions on mammograms. We are studying the use of CAD techniques to assist radiologists in interval change analysis.

Commonly used lesion classification methods for CAD employ information from a single image. These methods have been shown to perform well in lesion classification problems.⁵⁻¹² However, when mammograms from multiple examinations are available, it can be expected that even higher accuracy may be achieved if the computer can utilize the interval change information for classification. New computer vision methods will have to be designed to extract features characterizing temporal changes and to improve the differentiation between benign and malignant masses.

A number of researchers have developed algorithms to register the mass on current and prior mammograms. Sallam *et al.*¹³ have proposed a warping technique for mammogram registration based on manually identified control points. A mapping function was calculated for matching each point on the current mammogram to a point on the prior mammo-

gram. Brzakovic *et al.*¹⁴ have investigated a three-step method for comparison of the most recent and the 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.*¹⁵ 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. Furthermore, these studies aimed at registration without using the results for interval change analysis.

Gopal *et al.*^{16,17} and Hadjiiski *et al.*¹⁸⁻²⁰ have developed a multistage technique that defines a transformation to locally map the position of the mass on a current mammogram to a search region on the prior mammogram. A local search for the exact mass location is then performed on the prior mammogram. Good *et al.*²¹ have developed a technique that defines a transformation to map all points from the current mammogram onto a prior mammogram. The current mammogram is then subtracted from the prior mammogram.

Few studies have been performed so far in the area of automated classification of breast masses based on the interval change information. Gopal *et al.*²² and Hadjiiski *et al.*^{23,24} have carried out a preliminary study of the classification scheme that combines prior and current information automatically extracted from masses on prior and current mammograms, respectively. The classifier using the combined prior and current information performed better than the classifier using current information alone. To our knowledge, no other studies that describe automated classification of malignant and benign breast lesions based on temporal changes of mammographic features have been reported.

The goal of our research is to develop a CAD method for automated analysis of interval changes to be used as an aid to radiologists for detection and classification of malignant and benign lesions on mammograms. In this study, we conducted a preliminary investigation to demonstrate the feasibility of analyzing temporal differences in the texture and morphological features between a mass on the most recent mammogram and a prior mammogram of the same view for the classification task. Additionally, we compared this method with two classification methods, one of which is based on information extracted from the current mammograms alone, the other one is based on information extracted from the prior mammograms alone.

II. MATERIALS AND METHODS

The new classification technique is based on the design of features that characterize the temporal change in the lesion of interest between two mammographic examinations. The mass to be analyzed can either be identified manually by a radiologist or automatically by a computerized detection program. In this study, the mass on each mammogram was identified by an MQSA certified radiologist. The masses on both the current and the prior mammograms were automatically segmented using an active contour method that has been discussed in detail elsewhere.^{25,26} Examples of the segmentation are shown in Figs. 2 and 3 for a malignant and a benign mass, respectively. Features that characterized mammographic masses including texture features, morphological features, and spiculation features were extracted from each mass. Three of the morphological features are related to the mass size. Additionally, difference features were obtained by subtracting a feature of the prior mass from the corresponding feature of the current mass. The current, prior, and difference features formed a multidimensional feature space for the classification task. Stepwise feature selection applied to linear discriminant analysis (LDA) was used to select the most useful features. The selected features were then used as the input predictor variables for the LDA classifier (Fig. 1). The classifier was trained and tested by a leave-one-case-out resampling scheme. A case was considered to contain all regions of interest from a given patient. In each resampling step, the temporal pairs from 55 cases were used for feature selection and formulation of the linear discriminant function, while the temporal pairs from the left-out case were used for testing the trained classifier. A total of 56 training and testing

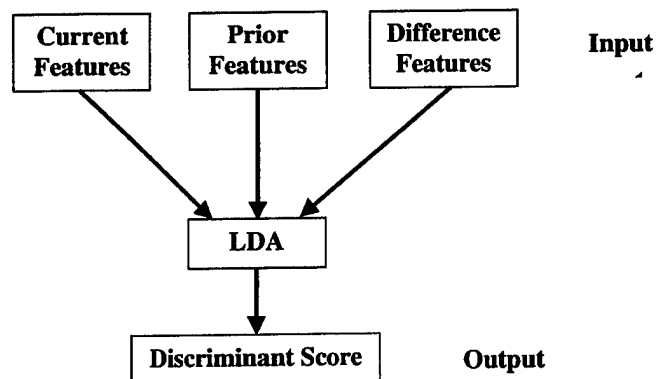


FIG. 1. Block diagram of the classification method.

steps were obtained from the 56 cases. The classification results from the 56 test cases were accumulated to evaluate the classifier performance. Since the data set in this study was still small, we chose the feature selection parameters such that the dimensionality of the input feature vector for the LDA classifier was small in order to reduce the possibility of over-training. The feature selection procedure is discussed in Sec. II C.

To evaluate the improvement in the classifier performance designed by using the temporal change information, two additional classifiers were obtained. One of them was trained using the information extracted from the current single images of the temporal pairs. We will refer to these images as current images. The other classifier was trained using the information extracted from the prior single images of the temporal pairs and we will refer to these images as prior images. Comparison of the three classifiers will reveal the effectiveness of interval change analysis for the classification of malignant and benign masses.

A. Data set

A set of 140 temporal pairs of mammograms containing biopsy-proven masses on the current mammograms was used to examine the performance of this approach. The data set consisted of 241 mammograms from 56 patients. The mammograms were digitized with a LUMISCAN 85 laser scanner at a pixel resolution of $50\ \mu\text{m} \times 50\ \mu\text{m}$ 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 0–4 OD units, with a slope of 0.001 OD/pixel value. The digitizer output was linearly converted so that a large pixel value corresponded to a low optical density. The image matrix size was reduced by averaging every 2×2 adjacent pixels and downsampled by a factor of 2, resulting in images with a pixel size of $100\ \mu\text{m} \times 100\ \mu\text{m}$ for further analysis.

There were 57 biopsy-proven masses (33 malignant and 24 benign) in the 56 cases. The 241 mammograms contained different mammographic views (CC, MLO, and lateral views) and multiple examinations of the masses including the examination when the biopsy decision was made. By matching masses of the same view from two different examinations, a total of 140 temporal pairs were formed, of which

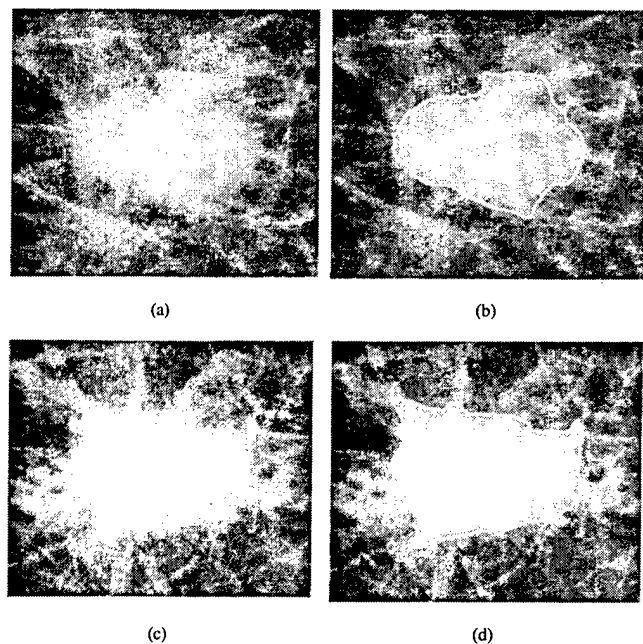


FIG. 2. A malignant mass: (a) the mass in a prior year mammogram (1997), (b) mass outline obtained by active contour segmentation. (c) the mass in a current year mammogram (1998), (d) mass outline obtained by active contour segmentation.

85 were malignant and 55 benign. A malignant temporal pair consisted of a biopsy-proven malignant mass or a mass that was initially not recommended for biopsy and later found to be malignant by biopsy in a future year. A similar definition was used for the benign temporal pairs. Within a pair, the current mammogram was defined as the mammogram with the later date, and the prior mammogram was defined as the one with the earlier date. Therefore, in cases with three consecutive exams, more than one temporal pair could be formed and two of the mammograms could be called "current." Among the 140 temporal pairs, we had 120 unique current mammograms. Of the masses in the 120 current mammograms, 70 were malignant and 50 benign.

Since all cases in this data set had undergone biopsy, the benign masses in this set could not be distinguished easily from the malignant ones based on current mammographic criteria. Changes occurred for the benign masses that prompted the radiologists to recommend biopsy. Examples of such cases are shown in Figs. 2 and 3. The malignant mass in Fig. 2 did not increase in size but changed its density. The benign mass (Fig. 3), on the other hand, appeared to have spicules. For the malignant masses in this data set, the average mass size, estimated by the radiologist as the longest dimension of the mass on the mammogram, was 8.2 mm on the prior mammograms and 12.7 mm on the current mammograms. The corresponding sizes were 10.6 and 12.2 mm, respectively, for the benign masses. As discussed in Sec. IV, 25 of the masses on the prior mammograms were too subtle for the radiologist to estimate their sizes. The average sizes given previously were obtained after excluding all temporal pairs that involved these masses.

The radiologist also rated the visibility of the masses on

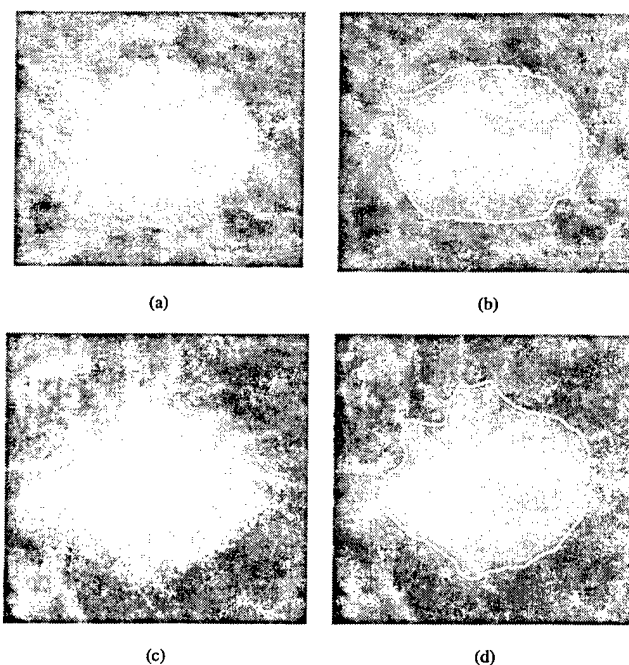


FIG. 3. A benign mass: (a) the mass on a prior year mammogram (1995), (b) mass outline obtained by active contour segmentation, (c) the mass on a current year mammogram (1996), (d) mass outline obtained by active contour segmentation.

the mammograms relative to those encountered in clinical practice on a 10-point scale, with 1 representing the most obvious and 10 representing the most subtle masses. The visibility of the masses on the current mammogram is plotted against those on the prior mammogram in Fig. 4 for the malignant and benign temporal pairs. Generally the malignant masses were less visible on the prior than on the current mammograms while the visibility of the benign masses was found to be more similar on the current and prior mammograms. The mean difference in the visibility rating between the prior and the current mammograms for the malignant masses is 2.8 compared to 1.2 for the benign masses ($p = 0.0007$ with an unpaired *t*-test between the malignant and benign masses). The correlation coefficient is 0.02 for malignant masses [Fig. 4(a)] and 0.37 for benign masses [Fig. 4(b)]. In addition, the radiologist also estimated the likelihood of malignancy of the current masses on a 10-point confidence scale (1—definitely benign and 10—definitely malignant) based on the 120 current mammograms alone without comparison with the prior (Fig. 5). The temporal pairs had a time interval of 6–36 months (Fig. 6). More than 70% of the pairs had a time interval of 12 months.

B. Feature extraction

A rectangular region of interest (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 masses on both the current and the prior mammograms were automatically segmented

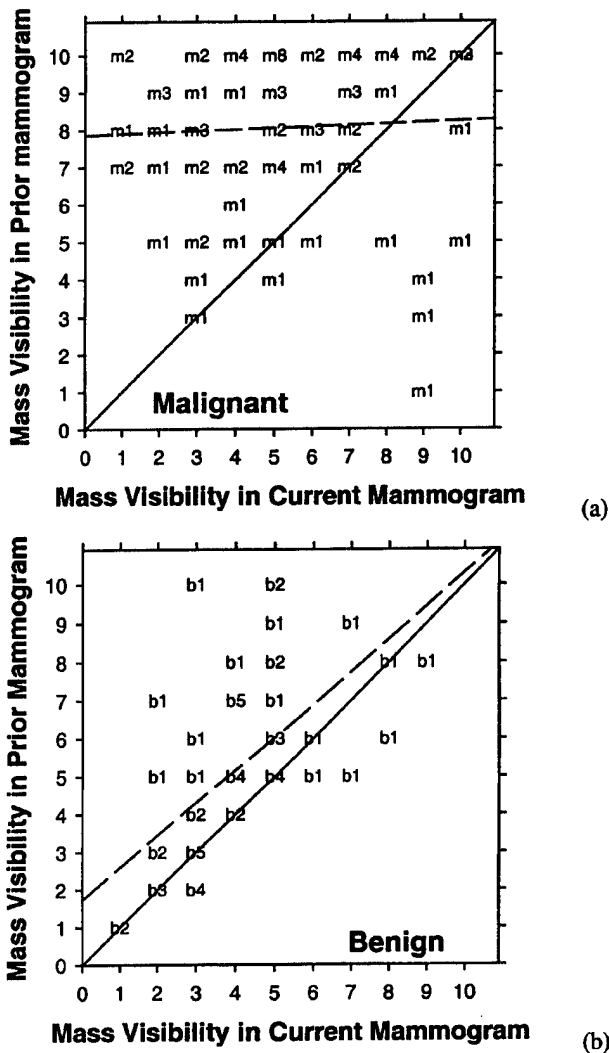


FIG. 4. Visibility of the masses on the current mammogram plotted against those on the prior mammogram for (a) malignant and (b) benign temporal pairs. The visibility was rated on a 10-point discrete scale (1 = most obvious, 10 = most subtle). Because many of the data points overlap, we indicate the number of points with the same rating by a number next to the symbol (m or b). The diagonal line on the graph represents the cases when the current and the prior mass sizes are identical. The dashed lines are the linear regression lines for the data defined by $y=0.038x+7.86$ for (a) and by $y=0.857x+1.742$ for (b). The correlation coefficient for malignant masses is 0.02 and for benign masses is 0.37.

within the ROI using a two-dimensional active contour method that was initialized by K-mean clustering.^{25,26}

The texture features used in this study were calculated from run-length statistics (RLS) matrices.²⁷ The RLS matrices were computed from the images obtained by the rubber band straightening transform (RBST).⁶ The RBST maps a band of pixels surrounding the mass onto the Cartesian plane (a rectangular region). In the transformed image, the mass border appears approximately as a horizontal edge, and spiculations appear approximately as vertical lines. A complete description of the RBST can be found in the literature.⁶

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

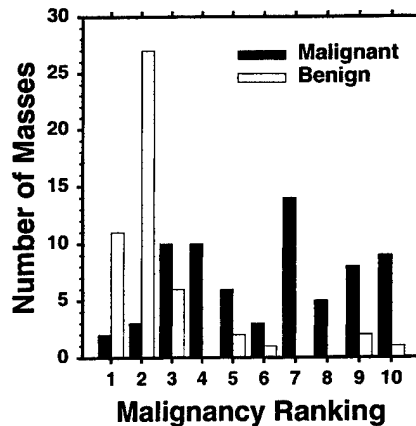


FIG. 5. The distribution of the malignancy ranking of the masses in the 120 current mammograms. The rating was performed by an experienced MQSA radiologist (1: definitely benign, 10: definitely malignant).

oriented Sobel filters and computing the absolute gradient values of the filtered image.⁶ Five texture measures, namely, short run emphasis (SRE), long run emphasis (LRE), gray level nonuniformity (GLN), run length nonuniformity (RLN), and run percentage (RP) were extracted from the vertical and horizontal gradient images in two directions, $\theta = 0^\circ$, and $\theta = 90^\circ$. Therefore, a total of 20 RLS features were calculated for each ROI. The definition of the RLS feature measures can be found in the Appendix and in the literature.²⁷

Morphological features were extracted from the automatically segmented mass shape. Five of the morphological features were based on the normalized radial length (NRL), defined as the Euclidean distance from the object's centroid to each of its edge pixels, i.e., the radial length, and normalized relative to the maximum radial length for the object.¹¹ The following five NRL features were extracted: mean (NRLAVG), standard deviation (NRLSD), entropy (NRL-ENT), area ratio (NRLAREAR), zero crossing count (NRLZCC). In addition, the perimeter (PERIM), area (AREA), circularity (CIRC), rectangularity (SQR), contrast (CONT), perimeter-to-area ratio (CRR), and Fourier descriptor (FF)

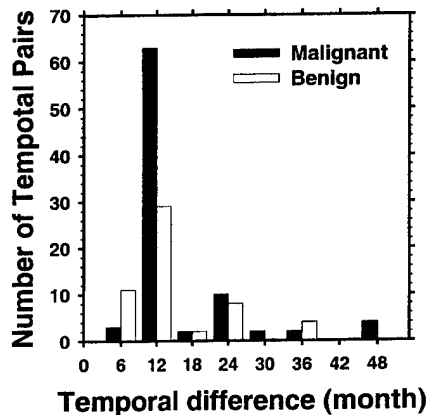


FIG. 6. Temporal interval between the current and the prior mammograms for the 140 temporal pairs in our data set.

features were extracted. The definitions of the morphological features can be found in the literature.^{26,28} Three of the morphological features (perimeter, area, and perimeter-to-area ratio) are related to the mass size and thus are feature descriptors of the mass size.

A spiculation measure was defined for each pixel on the mass border by using the statistics of the image gradient direction relative to the normal direction to the mass border. The statistics was determined in a 90° sector centered about the normal at the border pixel and outside of the mass border.^{25,26} The spiculation measure for each border pixel was normalized to be between 0 and $\pi/2$, with a value of $\pi/4$ indicating a random orientation of image gradients, and larger values indicating a higher likelihood of spiculation. Three features were extracted from the spiculation measure. The first feature (AVG) was the average of the spiculation measure for all pixels on the mass boundary. The second feature (PERC_ABV) was the percentage of border pixels with a spiculation measure larger than $\pi/4$, and the third feature (AVE_ABV) was the average of the spiculation measure for those pixels with a spiculation measure larger than $\pi/4$.

A total of 35 features (20 RLS, 12 morphological, and 3 spiculation) were therefore extracted from each ROI. Additionally, difference features were obtained by subtracting a prior feature from the corresponding current feature. Therefore, 35 difference features were derived from the 20 RLS, 12 morphological, and 3 spiculation features.

C. Feature selection

In order to reduce the number of the features and to obtain the best feature subset to design an effective classifier, feature selection with stepwise linear discriminant analysis²⁹ 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 squares³⁰) was used as a selection criterion. The optimization procedure used a threshold F_{in} for feature entry, a threshold F_{out} for feature removal, and a tolerance threshold T for measuring feature correlation with the other features. In 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} and its correlation value with the rest of the features in the pool is below T . In a feature removal step, the features that have already been entered in the selected feature pool are removed one at a time 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} , F_{out} and T are not known *a priori*, we examined a range of F_{in} , F_{out} , and T values using an automated simplex optimization method.^{31,32} The appropriate thresholds were chosen in such

TABLE I. Classification results for the classifier based on the temporal change information, the classifier based on current single image information, and the classifier based on prior single image information.

Classification	Avg. No. of selected features	Training A_z	Test A_z	Test partial $A_z^{(0.9)}$
Temporal pairs	10	0.92	0.88 ± 0.03	0.37 ± 0.10
Current images	11	0.90	0.82 ± 0.04	0.32 ± 0.08
Prior images	4	0.78	0.76 ± 0.04	0.24 ± 0.08

a way that a minimum number of features were selected to achieve a high accuracy of classification by LDA. More details about the stepwise linear discriminant analysis and its application to CAD can be found elsewhere.^{5,6}

The feature selection in this study was performed by applying the stepwise feature selection to the entire feature space (combination of texture, spiculation, and morphological features altogether) as well as subspaces obtained by different combinations of the three feature subspaces: texture, spiculation, and morphological features. The stepwise feature selection uses a sequential forward inclusion and backward elimination approach. The procedure does not exhaustively evaluate all possible combinations of individual features. It is therefore not optimal, especially when the feature space is large and the training sample is small. By limiting the input to the feature subspaces, the dimensionality was reduced compared to the entire feature space. We found that better feature subsets could be selected by the stepwise feature selection in the subspaces than in the entire feature space.

D. Evaluation methods

To evaluate the classifier performance, the training and test discriminant scores were analyzed using receiver operating characteristic (ROC) methodology.³³ The discriminant scores of the malignant and benign masses were used as decision variables in the LABROC1 program,³⁴ which fits a binormal ROC curve based on maximum likelihood estimation. The classification accuracy was evaluated as the area under the ROC curve, A_z . The performances of the classifiers were also assessed by estimating the partial area index ($A_z^{(0.9)}$). 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 , $(1 - TPF_0)$. 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 a cancer detection task.

III. RESULTS

The performances of the classifiers based on the temporal pairs, the current images, and the prior images are summarized in Table I. The classifiers that achieved the highest test A_z values with a small average number of features were presented here. Table II is a summary of the features selected for each classifier. For the 56 training subsets of temporal pairs used in this study, an average of 10 features were selected for

TABLE II. Selected features for classifiers based on temporal pairs, current images, and prior images. The letter "H" or "V" at the beginning of the texture feature labels indicates that the features were extracted from the horizontal or vertical gradient magnitude images, respectively. The number (0 or 90) at the end of the texture feature labels shows the direction at which the features were extracted.

Feature type	Group	Features	Temporal pairs			Current images	Prior images
			Curr	Pr	Diff	Curr	Pr
Texture	SRE	H_SRE_0			×		
		H_SRE_90	×		×		
		V_SRE_0	×		×	×	×
	LRE	V_SRE_90				×	
		V_LRE_0			×		×
	RLN	H_LRE_0				×	
		V_RLN_0	×			×	
RP	H_RP_0	×				×	
Spiculation		PERC_ABV	×			×	
		AVG		×			
		AVG_ABV					×
Morphological		CRR				×	
		NRLZCC				×	
		PERIM				×	
		NRLAVG				×	
		SQR				×	
		CONT				×	

the classification task. The most frequently selected features included 4 difference RLS features (3 SRE and 1 LRE), 4 RLS features (2 SRE, 1 RLN and 1 RP), 1 spiculation feature from the current image, and 1 spiculation feature from the prior image (Table II). The LDA classifier achieved an average training A_z of 0.92 and a test A_z of 0.88. The test partial $A_z^{(0.9)}$ was 0.37.

For classification of malignant and benign masses using the current single images (the current images of the temporal pairs), the LDA classifier selected an average of 11 features for the 56 training subsets. The most frequently selected features were 4 RLS features (2 SRE, 1 LRE and 1 RLN), 1 spiculation feature, and 6 morphological features (Table II). The classifier achieved an average training A_z of 0.90, a test A_z of 0.82, and a test partial $A_z^{(0.9)}$ of 0.32.

For the classification of masses based on the prior single images alone, an average of 4 features were selected for the 56 training subsets. The most frequently selected features were 3 RLS features (1 SRE, 1 LRE, and 1 RP) and 1 spiculation feature. The LDA classifier achieved an average training A_z of 0.78, test A_z of 0.76, and test partial $A_z^{(0.9)}$ of 0.24.

The test ROC curves for the three classifiers are compared in Fig. 7. The difference in the test A_z between the classifier based on the temporal pairs and that based on the current images alone is statistically significant ($p=0.015$). The difference in the test A_z between the classifier based on the temporal pairs and that based on the prior images alone is also statistically significant ($p=0.001$). The partial area index for the classifier based on the temporal pairs is also improved compared to the classifiers based on the current or the prior images alone, although the differences did not achieve statistical significance.

IV. DISCUSSION

Texture and spiculation features were important for malignant and benign classification of mammographic masses for all three types of classifiers: the classifier based on temporal pair information, the classifier based on current image information, and the classifier based on prior image information. One or more of the spiculation features were always selected in all training partitions for all three classifiers. The most frequently selected texture features were the short run emphasis (SRE) features. They comprised more than 50% of the texture features selected for the three classifiers (Table II).

The temporal-information-based classifier showed improved performance compared to the classifiers based on cur-

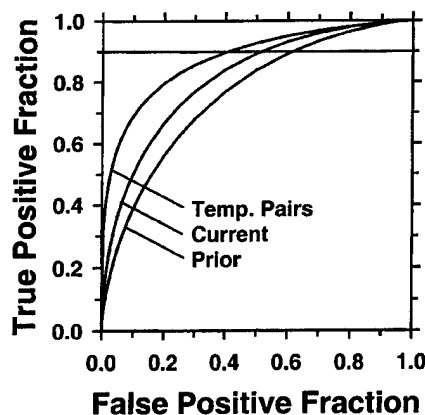


FIG. 7. The test ROC curves for the classifiers based on temporal pair information, current image information, and prior image information.

rent or prior image information alone. The input feature space to the temporal-information-based classifiers included the current, prior, and difference features. This allows the classifier to choose the individual features or the difference features. Using the stepwise feature selection procedure and the linear discriminant classifier, it was found that the texture and the spiculation features contained useful temporal information to perform malignant and benign mass classification. Texture features appeared to provide the best information by the difference features obtained from subtracting the prior from the corresponding current features (SRE and LRE difference features). On the other hand, the best use of the spiculation features appeared to be a direct combination of current and prior features in the input feature vector by the LDA since the individual features were chosen.

We found that better feature subsets could be selected by the stepwise feature selection in the subspaces than in the entire feature space. For example, for the temporal-information-based classifier, a better feature subset with a higher test A_z at 0.88 was found when the input feature space included only the texture and spiculation subspaces. The addition of the morphological feature subspace to the input feature space reduced the highest test A_z to 0.84. Similarly, in the case of the classifier based on prior image information, a better feature subset was obtained when the texture and spiculation feature subspaces were used in the input feature space for stepwise feature selection. Again the addition of the morphological feature subspace to the input feature space reduced the highest test A_z to 0.72. The classifier based on current image information was the only one, among the three, that obtained a better result, as shown in Table I, when the morphological feature subspace was included in the input feature space.

One reason for the poor performance of the morphological features may be due to the fact that the masses were more subtle in the prior images. In fact, the experienced MQSA mammographer was not confident in seeing 25 of the "masses" on the prior images and could not provide a mass size estimation for them. Although the active contour model would stop the iteration based on the preset criteria and found an "outline" of the masses on the prior mammograms, generally these mass outlines were less reliable than those on the current masses in providing morphological characteristics of the masses. Texture features did not depend as strongly on the precise mass boundary as morphological features. Three out of the four features selected for classification of the malignant and benign masses on the prior images were RLS texture features. A spiculation feature was also found to be a good discriminator.

We also performed ROC analysis of the malignancy confidence ratings provided by the experienced MQSA radiologist for the current image data set (120 images). The distribution of the malignancy ratings is shown in Fig. 5, which resulted in an A_z value of 0.80 ± 0.04 . This indicates that the masses in the current mammograms cannot be easily distinguished as malignant or benign even by an experienced radiologist, consistent with the fact that all lesions had indeed undergone biopsy. The classifier based on the current image

information has an A_z value of 0.82 ± 0.04 , similar to the accuracy of the radiologist for this data set.

In this study, the locations of the masses were identified manually on both the current and the prior mammograms by a radiologist. This simulated the situation when a radiologist finds a mass either in a diagnostic or a screening setting and call upon the CAD algorithm to seek a second opinion on the likelihood of malignancy of the mass based on the interval change information. We are developing an automated regional registration technique that can automatically locate the mass on the prior mammogram based on its location on the current mammogram. The location of the mass on the current mammogram can be identified by a radiologist or by an automated mass detection algorithm. In the latter case, the process of mass detection, current and prior mass registration, and classification can be fully automated. The analysis of interval change can be incorporated as one of the functions provided by a CAD system for interpretation of mammograms.

In this study, we employed a simple measure of temporal change by taking the difference between the feature from the current mass and the corresponding feature from the prior mass. We observed improvement in classification with this simple temporal information. It will be important to evaluate other similarity measures that can characterize small difference in image features of the object of interest. It can be expected that a more sensitive similarity measure will provide a better measurement of dissimilarity, or difference, between the current and prior masses and further improve the utilization of the temporal change information on mammograms.

V. CONCLUSION

We performed a preliminary study to evaluate the effectiveness of interval change analysis for classification of malignant and benign masses on mammograms. It was found that the difference RLS texture features and spiculation features were useful for identification of malignancy in temporal pairs of mammograms. The information on the prior image was important for characterization of the masses; 5 out of the 10 selected features contained prior information. We found that the mass size descriptors were not discriminatory features for these difficult cases because many of the benign masses also grew over time. In comparison with the classification based on image information from the current images alone, the temporal change information significantly ($p = 0.015$) improved the accuracy for classification of the masses in terms of the total area under the ROC curve (A_z). The partial area under the ROC curve for the classifier based on the temporal pairs ($A_z^{(0.9)} = 0.37$) is also improved compared to the classifier based only on the current images ($A_z^{(0.9)} = 0.32$), although the difference did not achieve statistical significance. Further studies are under way to improve this temporal change classification technique and to evaluate its performance on a larger data set.

ACKNOWLEDGMENTS

This work is supported by a Career Development Award from the USAMRMC (No. DAMD 17-98-1-8211) (L.H.), USPHS Grant No. CA 48129, and a USAMRMC grant (No. DAMD 17-96-1-6254). The content of this publication does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned in the publication should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program.

APPENDIX: RUN LENGTH STATISTICS TEXTURE FEATURES

A gray level run length is a set of consecutive collinear pixels all having the same gray level value. The length of the run is the number of pixels in the run. For a given image it is possible to compute a gray level run length matrix for runs in any given direction. In this study, two directions are used: $\theta=0^\circ$, and $\theta=90^\circ$. Let $p(i, j)$ be the number of times there is a run of length j that has a gray level i . Let N_g be the number of gray levels and N_r be the number of runs. The short run emphasis is defined as

$$\text{SRE} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i, j)}{j^2}}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j)}$$

This feature divides the frequency of each run length by the length of the run squared. This tends to emphasize short runs. The denominator is the total number of runs in the image and serves as a normalizing factor. The long run emphasis is defined as

$$\text{LRE} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} j^2 p(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j)}$$

This feature multiplies the frequency of each run length by the length of the run squared. This tends to emphasize long runs.

The gray level nonuniformity is defined as

$$\text{GLN} = \frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_r} p(i, j))^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j)}$$

This feature squares the number of run lengths for each gray level. This measures the gray level nonuniformity of the image. If the runs are equally distributed over all gray levels, the feature takes on its lowest values. A larger run length contributes more to the feature value.

Run length nonuniformity is defined as

$$\text{RLN} = \frac{\sum_{j=1}^{N_r} (\sum_{i=1}^{N_g} p(i, j))^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j)}$$

This feature measures the nonuniformity of the run lengths. If the runs are equally distributed over all lengths, the feature will have a low value. A larger run contour contributes more to the feature value.

Run percentage is defined as

$$\text{RP} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j)}{P}$$

This feature is a ratio of the total number of runs to the total number of possible runs (P) if all runs have a length of one.

The above-given definitions are based on Galloway²⁷ and more details can be found in this reference.

^{a)}Electronic mail: lhadjisk@umich.edu

¹H. 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).

²L. Tabar and P. B. Dean, "The control of breast cancer through mammography screening," *Radiol. Clin. North Am.* **25**, 993-1005 (1987).

³L. W. Bassett, B. Shayestehfar, and I. Hirbawi, "Obtaining previous mammograms for comparison: Usefulness and costs," *AJR, Am. J. Roentgenol.* **163**, 1083-1086 (1994).

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

⁵H. 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 discriminant analysis in texture feature space," *Phys. Med. Biol.* **40**, 857-876 (1995).

⁶B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," *Med. Phys.* **25**, 516-526 (1998).

⁷H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, K. L. Leung, D. D. Adler, and M. M. Goodsitt, "Computerized classification of malignant and benign microcalcifications on mammograms: Texture analysis using an artificial neural network," *Phys. Med. Biol.* **42**, 549-567 (1997).

⁸L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, and M. A. Helvie, "Classification of malignant and benign masses based on hybrid ART2LDA approach," *IEEE Trans. Med. Imaging* **18**, 1178-1187 (1999).

⁹Y. 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* **187**, 81-87 (1993).

¹⁰V. Goldberg, A. Manduca, D. L. Evert, J. J. Gisvold, and J. F. Greenleaf, "Improvement in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence," *Med. Phys.* **19**, 1475-1481 (1992).

¹¹J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computer-aided image analysis," *IEEE Trans. Med. Imaging* **12**, 664-669 (1993).

¹²Z. M. 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.* **5**, 155-168 (1998).

¹³M. 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).

¹⁴D. Brzakovic, N. Vujovic, M. Neskovic, P. Brzakovic, and K. Fogarty, "Mammogram analysis by comparison with previous screenings," in *Digital Mammography*, edited by A. G. Gale, S. M. Astley, D. R. Dance, and A. Y. Cairns (Elsevier, Amsterdam, 1994).

¹⁵N. Vujovic and D. Brzakovic, "Establishing the correspondence between control points in pairs of mammographic images," *IEEE Trans. Med. Imaging* **6**, 1388-1399 (1997).

¹⁶S. S. Gopal, H.-P. Chan, N. Petrick, T. E. Wilson, B. Sahiner, M. A. Helvie, and M. Goodsitt, "A regional registration technique for automated analysis of interval changes of breast lesions," *Proc. SPIE* **3338**, 118-131 (1998).

¹⁷S. S. Gopal, H.-P. Chan, T. E. Wilson, M. A. Helvie, N. Petrick, and B.

- Sahiner, "A regional registration technique for automated interval change analysis of breast lesions on mammograms," *Med. Phys.* **26**, 2669–2679 (1999).
- ¹⁸L. M. Hadjiiski, H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, and S. Sanjay-Gopal, "Automated identification of breast lesions in temporal pairs of mammograms for interval change analysis," *Radiology* **213(P)**, 229–230 (1999).
- ¹⁹L. M. Hadjiiski, H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, S. Paquerault, and C. Zhou, "Interval change analysis in temporal pairs of mammograms using a local affine transformation," *Proc. SPIE* **3979**, 847–853 (2000).
- ²⁰L. M. Hadjiiski, H. P. Chan, B. Sahiner, N. Petrick, and M. A. Helvie, "Automated registration of breast lesions in temporal pairs of mammograms for interval change analysis—Local affine transformation for improved localization," *Med. Phys.* **28**, 1070–1079 (2001).
- ²¹W. F. Good, B. Zheng, Y. H. Chang, Z. H. Wang, and G. S. Maitz, "Generalized procrustean image deformation for subtraction of mammograms," *Proc. SPIE* **3661**, 1562–1573 (1999).
- ²²S. S. Gopal, H.-P. Chan, B. Sahiner, N. Petrick, T. E. Wilson, and M. A. Helvie, "Evaluation of interval change in mammographic features for computerized classification of malignant and benign masses," *Radiology* **205(P)**, 216 (1997).
- ²³L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. Gurcan, "Computer-aided classification of malignant and benign breast masses by analysis of interval change of features in temporal pairs of mammograms," *Radiology* **217(P)**, 435 (2000).
- ²⁴L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. Gurcan, "Analysis of temporal change of mammographic features for computer-aided characterization of malignant and benign masses," *Proc. SPIE* **4322**, 661–666 (2001).
- ²⁵B. Sahiner, H. P. Chan, N. Petrick, L. M. Hadjiiski, M. A. Helvie, and S. Paquerault, "Active contour models for segmentation and characterization of mammographic masses," *The Fifth International Workshop on Digital Mammography Proceedings, 2000, IWDM-2000*, pp. 357–362.
- ²⁶B. Sahiner, H.-P. Chan, N. Petrick, M. A. Helvie, and L. M. Hadjiiski, "Improvement of mammographic mass characterization using spiculation measures and morphological features," *Med. Phys.* **28**, 1455–1465 (2001).
- ²⁷M. M. Galloway, "Texture classification using gray level run lengths," *Comput. Graph. Image Process.* **4**, 172–179 (1975).
- ²⁸N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," *Med. Phys.* **26**, 1642–1654 (1999).
- ²⁹M. J. Norusis, *SPSS for Windows Release 6 Professional Statistics* (SPSS, Chicago, IL, 1993).
- ³⁰M. M. Tatsuoka, *Multivariate Analysis, Techniques for Educational and Psychological Research*, 2nd ed. (Macmillan, New York, 1988).
- ³¹S. S. Rao, *Optimization: Theory and Applications* (Wiley, New York, 1979).
- ³²F. A. Lootsma, *Numerical Methods for Non-linear Optimization* (Academic, New York, 1972).
- ³³C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.* **21**, 720–733 (1986).
- ³⁴C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Stat. Med.* **17**, 1033–1053 (1998).

Analysis of temporal change of mammographic features for computer-aided characterization of malignant and benign masses

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

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

ABSTRACT

A new classification scheme was developed to classify mammographic masses as malignant and benign by using interval change information. The masses on both the current and the prior mammograms were automatically segmented using an active contour method. From each mass, 20 run length statistics (RLS) texture features, 3 spiculation features, and mass size were extracted. Additionally, 20 difference RLS features were obtained by subtracting the prior RLS features from the corresponding current RLS features. The feature space consisted of the current RLS features, the difference RLS features, the current and prior spiculation features, and the current and prior mass sizes. Stepwise feature selection and linear discriminant analysis classification (LDA) were used to select and merge the most useful features. A leave-one-case-out resampling scheme was applied to train and test the classifier using 140 temporal image pairs (85 malignant, 55 benign) obtained from 57 biopsy-proven masses (33 malignant, 24 benign) in 56 patients. An average of 10 features were selected from the 56 training subsets: 4 difference RLS features, 4 RLS features and 1 spiculation feature from the current image, and 1 spiculation feature from the prior, were most often chosen. The classifier achieved an average training A_z of 0.92 and a test A_z of 0.88. For comparison, a classifier was trained and tested using features extracted from the 120 current single images. This classifier achieved an average training A_z of 0.90 and a test A_z of 0.82. The information on the prior image significantly ($p=0.01$) improved the accuracy for classification of the masses.

Keywords: Computer-Aided Diagnosis, Interval Changes, Classification, Feature analysis, 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 in mammographic interpretation to detect developing malignancy^{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.

Commonly used classification methods for CAD use information from a single image. These methods have been shown to perform well in lesion classification problems⁶⁻¹³. However, when multiple-year mammograms of a mass are available, it is not trivial to design computer vision methods to use the temporal information for computer-aided classification and to improve the differentiation between benign and malignant masses.

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 can be used to assist radiologists in evaluating interval changes and thus distinguishing between malignant and benign masses for CAD. We have previously presented⁵ preliminary results that demonstrated the feasibility of classifying malignant and benign masses based on interval change analysis. In this study, we continue the development of this approach. Additionally, we compared this method with a classification method based on information extracted from the current mammogram alone.

2. CLASSIFICATION TECHNIQUE

A new classification scheme was developed to classify mammographic masses as malignant and benign by using interval change information. The technique is based on the generation of features that we expect will represent adequately the temporal information and will discriminate between malignant and benign masses.

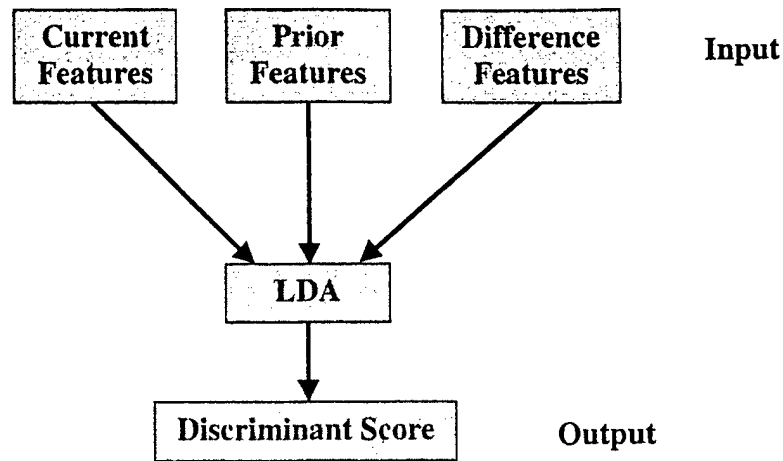


Figure 1. Block-diagram of the classification method.

The mass to be analyzed can either be identified manually by a radiologist or automatically by a computerized detection program. In this study, the masses were identified by an MQSA radiologist on each mammogram. The masses on both the current and the prior mammograms were automatically segmented using an active contour method. An example of the segmentation is shown in Figure 2 and Figure 3 for a malignant and a benign mass, respectively. Features such as texture features, spiculation features and mass size were extracted from each mass. Additionally, difference features were obtained by subtracting a prior feature from the corresponding current feature. The feature space consisted of current, prior, and difference features. Stepwise feature selection applied to linear discriminant analysis (LDA) were used to select the most useful features. The selected features were then used as the input predictor variables of the LDA classifier (Figure 1). A leave-one-case-out resampling scheme was employed to train and test the classifier. The LDA classifier was used in order to keep the discrimination function simple, thereby reducing the possibility of over-training.

To evaluate the improvement in the classifier performance designed by using the temporal change information, an additional classifier was trained using the information extracted from the current single images of the temporal pairs. We will refer to these images as current images. Comparison of the two classifiers will reveal the effectiveness of interval change analysis on classification of malignant and benign masses.

3. DATA SET

A set of 140 temporal pairs of mammograms containing biopsy-proven masses on the current mammograms was used to examine the performance of this approach. The data set consisted of a total of 241 mammograms from 56 patients. The mammograms were digitized with a LUMISCAN 85 laser scanner at a pixel resolution of $50 \mu\text{m} \times 50 \mu\text{m}$ 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 to 4 OD units, with a slope of 0.001 OD/pixel value. Outside this range, the slope of the calibration curve decreased gradually. The digitizer output was linearly converted so that a large pixel value corresponded to a low optical density. The images were averaged and down-sampled by a factor of 2 resulting in images with a pixel size of $100 \mu\text{m} \times 100 \mu\text{m}$ for further analysis.

The 56 cases contained 57 biopsy proven masses (33 malignant and 24 benign). The 241 mammograms contained different mammographic views and multiple years of the masses including the year when the biopsy was performed. By matching masses of the same view from two different exams, a total of 140 temporal pairs were formed, of which 85 were malignant and 55 benign. A malignant temporal pair consisted of a biopsy proven malignant mass or a mass that was followed up and found to be malignant by biopsy in a future year. Similar definitions were used for the benign temporal pairs. Within the 140 temporal pairs, a total of 120 mammograms were current mammograms. Of the 120 current mammograms, 70 were malignant and 50 benign.

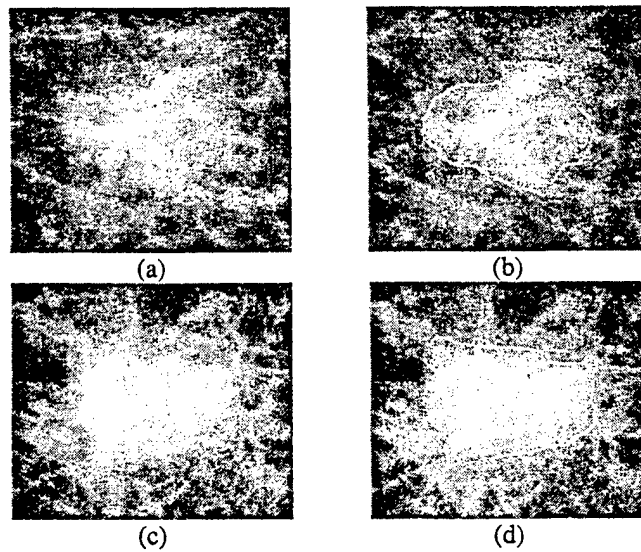


Figure 2. A malignant mass: (a) the mass in a prior year mammogram (1997), (b) mass outline obtained by active contour segmentation, (c) the mass in a current year mammogram (1998), (d) mass outline obtained by active contour segmentation.

Since all cases in this data set had undergone biopsy, the benign masses in this set could not be distinguished easily from the malignant ones based on current mammographic criteria. Examples of such cases are shown in Figure 2 and Figure 3. The malignant mass in Figure 2 did not increase in size but changed its density. The benign mass (Figure 3), on the other hand, appeared to have spicules. For the malignant masses in this data set, the average mass size was 8.2 mm on the prior mammograms and 12.7 mm on the current mammograms. The corresponding sizes were 10.6 mm and 12.2 mm, respectively, for the benign masses. The temporal pairs had a time interval of 6 to 36 months. More than 70% of the pairs had a time interval of 12 months.

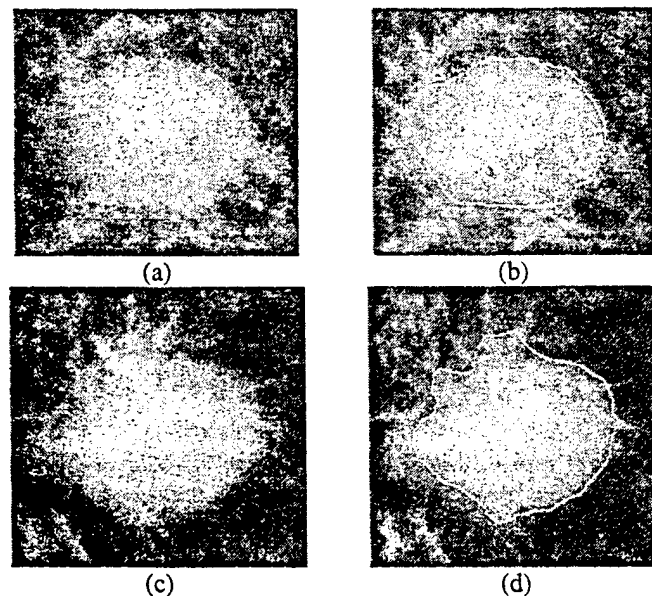


Figure 3. A benign mass: (a) the mass on a prior year mammogram (1995), (b) mass outline obtained by active contour segmentation, (c) the mass on a current year mammogram (1996), (d) mass outline obtained by active contour segmentation.

4. FEATURE EXTRACTION

A rectangular region of interest (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 masses on both the current and the prior mammograms were automatically segmented using a 2D active contour method, initialized by adaptive thresholding^{14,15}.

The texture features used in this study were calculated from run-length statistics (RLS) matrices¹⁸. The RLS matrices were computed from the images obtained by the rubber band straightening transform (RBST)⁷. The RBST maps a band of pixels surrounding the mass onto the Cartesian plane (a rectangular region). In the transformed image, the mass border appears approximately as a horizontal edge, and spiculations appear approximately as vertical lines. A complete description of the RBST can be found in the literature⁷.

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. 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, $\theta = 0^\circ$, and $\theta = 90^\circ$. Therefore, a total of 20 RLS features were calculated for each ROI. The definition of the RLS feature measures can be found in the literature¹⁸.

The morphological features were extracted from the automatically segmented mass shape, and included features such as the area, circularity, rectangularity, compactness, and the axis ratio¹⁵. Spiculation features were extracted by using the statistics of the image gradient direction relative to the normal direction to the mass border in a ring of pixels surrounding the mass^{14,15}.

A total of 35 features (20 RLS, 12 morphological and 3 spiculation) were therefore extracted from each ROI. Additionally, difference features were obtained by subtracting a prior feature from the corresponding current feature. Therefore 20 RLS, 12 morphological and 3 spiculation difference features were obtained.

5. FEATURE SELECTION

In order to reduce the number of the features and to obtain the best feature subset to design an effective classifier, feature selection with stepwise linear discriminant analysis^{19,20} was applied. At each step of the stepwise selection procedure one feature is entered or removed from the feature pool based on analysis of its effect on the selection criterion. The stepwise selection procedure is controlled by a simplex optimization method^{16, 17} in such a way that a minimum number of features were selected to achieve a high accuracy of classification by LDA. More details about the stepwise linear discriminant analysis and its application to CAD can be found elsewhere^{6, 7}.

6. EVALUATION METHODS

To evaluate the classifier performance, the training and test discriminant scores were analyzed using receiver operating characteristic (ROC) methodology²¹. The discriminant scores of the malignant and benign masses were used as decision variables in the LABROC1 program²², which fits a binormal ROC curve based on maximum likelihood estimation. The classification accuracy was evaluated as the area under the ROC curve, A_z . The performances of the classifiers were also assessed by estimation of the partial area index ($A_z^{(0.9)}$). 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 , $(1-TPF_0)$. 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 a cancer detection task.

7. CLASSIFICATION RESULTS

For the data set used in this study, an average of 10 features were selected from the 56 training subsets. The most frequently selected features included 4 difference RLS features, 4 RLS features and 1 spiculation feature from the current image, and 1 spiculation feature from the prior. The LDA classifier achieved an average training A_z of 0.92 and a test A_z of

0.88. The LDA classifier using features extracted from the current single images (the current images of the temporal pairs) achieved an average training A_z of 0.90 and a test A_z of 0.82. An average of 11 features were selected from the 56 training subsets. The most frequently selected features were 4 RLS features, 1 spiculation feature from the current image and 6 morphological features. The difference in the test A_z between the two classifiers is statistically significant ($p=0.01$). The classifier based on temporal pairs achieved a test partial $A_z^{(0.9)}$ of 0.37 and the classifier based on current images achieved a test $A_z^{(0.9)}$ of 0.32. These results are summarized in Table 1.

Table 1. Classification results for the classifier based on the temporal change information and the classifier based on current single image information.

Classification	Avg. no. of selected features	Training A_z	Test A_z	Test partial $A_z^{(0.9)}$
Temporal pairs	10	0.92	0.88 ± 0.028	0.37 ± 0.1
Current images	11	0.90	0.82 ± 0.038	0.32 ± 0.08

8. CONCLUSION

The difference RLS texture features and spiculation features were useful for identification of malignancy in temporal pairs of mammograms. The information on the prior image was important for characterization of the masses; 5 out of the 10 selected features contained prior information. We found that the size of the mass was not a discriminatory feature for these difficult cases because many of the benign masses also grew over time. The temporal change information significantly ($p=0.01$) improved the accuracy for classification of the masses in terms of the total area under the ROC curve (A_z). The partial area under the ROC curve is also improved for the classifier based on current and prior images ($A_z^{(0.9)} = 0.37$) compared to the classifier based only on the current images ($A_z^{(0.9)} = 0.32$), although the difference did not achieve statistical significance. Further studies are underway to improve this temporal change classification technique and to evaluate its performance on a larger data set.

ACKNOWLEDGMENTS

This work is supported by a Career Development Award from the USAMRMC (DAMD 17-98-1-8211) (L.H.), a USPHS Grant CA 48129, and a USAMRMC grant (DAMD 17-96-1-6254).

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. 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", *Radiology* 2000 **217(P)**: 435.
6. 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.* **40**, pp. 857-876, 1995.

7. B. Sahiner, H. P. Chan, N. Petick, M. A. Helvie, and M. M. Goodsitt, Computerized Characterization of Masses on Mammograms: The Rubber Band Straightening Transform and Texture Analysis, *Med. Phys.* **25** (4), pp. 516-526, April 1998.
8. 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 Microcalcifications on mammograms: Texture analysis using an Artificial Neural Network, *Phys. Med. Biol.* **42**, pp. 549-567, 1997.
9. L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M. 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.
10. 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* **187**, pp. 81-87, 1993.
11. 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.*, **19**, pp. 1475-1481, 1992.
12. J. Kilday, F. Palmieri, and M. D. Fox, Classifying Mammographic Lesions Using Computerized Image Analysis, *IEEE Transaction on Medical Imaging*, Vol. **12**, No. 4, pp. 664-669, Dec. 1993.
13. 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.*, **5**, pp. 155-168, 1998.
14. B. Sahiner, H. P. Chan, N. Petrick, L. M. Hadjiiski, M. A. Helvie and S. Paquerault, "Active contour models for segmentation and characterization of mammographic masses," *Madison, WI: Medical Physics Publishing, The 5th International Workshop on Digital Mammography*, Toronto, Canada, Proc. IWDM-2000, (in press), 2000.
15. B. Sahiner, H.-P. Chan, N. Petrick, M. A. Helvie and L. M. Hadjiiski, "Improvement of mammographic mass characterization using spiculation measures and morphological features," *Medical Physics* (submitted), 2000.
16. S.S. Rao, *Optimization: Theory and Applications*", Wiley Eastern Limited, 1979.
17. F.A. Lootsma, (ed). *Numerical methods for non-linear optimization*, Academic Press, New York, 1972.
18. M. M. Galloway, Texture Analysis Using Gray Level Run Length, *Comput. Graph. Image Process.* **4**, pp. 172-179, 1975.
19. M. J. Norusis, *SPSS Professional Statistics 6.1* (SPSS Inc., Chicago, 1993).
20. M. M. Tatsuoka, *Multivariate Analysis, Techniques for Educational and Psychological Research* (Macmillan, New York, 1988).
21. C. E. Metz, ROC methodology in radiographic imaging, *Invest. Radiol.*, **21**, pp. 720-733, 1986.
22. 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 Annual Meeting of the American Statistical Association, Anaheim, CA*, 1990.

An Adaptive Similarity Measure for Automated Identification of Breast Lesions in Temporal Pairs of Mammograms for Interval Change Analysis

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

PURPOSE: An adaptive similarity measure (ASM) is designed to improve automated identification of corresponding lesions on prior mammograms. This technique is the basis for interval change analysis of breast lesions in CAD applications.

MATERIALS AND METHODS: A new class of similarity measures (SM) is proposed. It combines adaptive filtering to enhance the lesion and a SM as a figure-of-merit (FOM) measure. The filters are designed with a training set to maximize and minimize the FOM for the similar and dissimilar image pairs, respectively, by using a gradient optimization technique.

The ASM was applied to the final stage of our multistage regional registration technique for mass identification on the prior mammogram. 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 a search region, guided by the ASM. This new approach was evaluated by using 179 temporal pairs of mammograms containing biopsy-proven masses.

RESULTS: 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 lesion centroids on the prior mammogram was 4.5 ± 6.7 mm. In comparison, the correct localization and the average distance using a conventional correlation SM were 84% and 4.9 ± 7.0 mm, respectively.

CONCLUSION: The ASM improved the identification of the corresponding lesions on temporal pairs of mammograms. Further studies are underway to improve the technique, expand it to different types of SM, and evaluate its accuracy on a larger data set.

Computerized Regional Registration of Corresponding Microcalcification Clusters on Temporal Pairs of Mammograms for Interval Change Analysis

L.M. Hadjiiski, PhD, Ann Arbor, MI (lhadjisk@umich.edu) • H. Chan, PhD • N.A. Petrick, PhD • B. Sahiner, PhD • M.N. Gurcan, PhD • M.A. Helvie, MD • et al

PURPOSE: To develop a regional registration technique for identifying corresponding microcalcification clusters on current and prior mammograms of the same view. The technique will be useful for computerized analysis of interval changes of microcalcification clusters in computer aided diagnosis (CAD).

METHOD AND MATERIALS: A multi-stage regional registration technique is being developed. In the first stage, an initial fan-shape search region was estimated on the prior mammogram based on the cluster location on the current mammogram. In the second stage, detection of cluster candidates within the search region was performed with an automated cluster search program. The cluster (TP) on the current image was paired with every detected cluster (TP or FP) in the search region. In the final stage, a correspondence classifier was designed to reduce the false pairs (TP-FP) within the search region. Texture and morphological features were extracted from the clusters on the current and the prior mammograms. Similarity measures were derived from the extracted features of the TP or FP clusters for each temporal pair. Stepwise feature selection with simplex optimization was used to select the optimal feature subset. A linear discriminant classifier was used to merge the selected features for classification of the TP-TP and TP-FP cluster pairs. In this preliminary study, a data set of 51 temporal pairs of mammograms from 19 patients containing biopsy-proven microcalcification clusters was used. The true cluster locations were identified by an MQSA radiologist. A leave-one-case-out training and testing resampling scheme was used for feature selection and classification.

RESULTS: Using a search region with an average area of 1350 mm² allowed all clusters of interest to be localized in the search region. The average distance between the estimated and the true centroid of the microcalcification clusters on the prior mammogram was 7.9 ± 4.1 mm after the first stage. The cluster search program detected 90% (46/51) of the true clusters with an average of 0.69 FP cluster within the search region on the prior mammograms. The correspondence classifier reduced the FP rate to an average of 0.41 FP cluster at the cost of misclassifying 1 true pair.

CONCLUSIONS: Our preliminary study demonstrated that the regional registration technique is a promising approach for identifying corresponding microcalcification clusters on temporal pairs of mammograms. Further studies are underway to improve the technique and to evaluate its accuracy on a larger data set.

4684-79, Session 14

Computer-Aided Characterization of Malignant and Benign Microcalcification Clusters Based on the Analysis of Temporal Change of Mammographic Features

Lubomir Hadjiiski, Heang-Ping Chan, Metin Gurcan, Berkman Sahiner, Nicholas Petrick, Mark A. Helvie, Marilyn Roubidoux (Department of Radiology, The University of Michigan, Ann Arbor, MI 48109-0904)

We have previously demonstrated that interval change analysis can improve differentiation of malignant and benign masses. In this study, a new classification scheme using interval change information was developed to classify mammographic microcalcification clusters as malignant and benign. From each cluster, 20 run length statistic texture features (RLSF) and 21 morphological features (MF) were extracted. Twenty difference RLSF were obtained by subtracting a prior RLSF from the corresponding current RLSF. The feature space consisted of the current RLSF, the difference RLSF, and the current and prior MF. A leave-one-case-out resampling was used to train and test the classifier using 65 temporal image pairs (19 malignant, 46 benign) containing biopsy-proven microcalcification clusters. Stepwise feature selection and a linear discriminant classifier, designed with the training subsets alone, were used to select and merge the most useful features. An average of 10 features were selected from the training subsets, of which 2 difference RLSF and 6 MF were consistently selected from most of the training subsets. The classifier achieved an average training A_z of 0.97 and a test A_z of 0.85. For comparison, a classifier based on the current single image features achieved an average training A_z of 0.93 and test A_z of 0.79. These results indicate that the use of temporal information improved the accuracy of microcalcification characterization.