

AD_____

AWARD NUMBER: W81XWH-06-1-0726

TITLE: Correlative Feature Analysis for Multimodality Breast CAD

PRINCIPAL INVESTIGATOR: Yading Yuan

CONTRACTING ORGANIZATION: University of Chicago
Chicago, IL 60637

REPORT DATE: September 2009

TYPE OF REPORT: Annual Summary

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

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

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

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE 1 September 2009		2. REPORT TYPE Annual Summary		3. DATES COVERED 1 Sep 2006 – 31 Aug 2009	
4. TITLE AND SUBTITLE Correlative Feature Analysis for Multimodality Breast CAD				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-06-1-0726	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Yading Yuan E-Mail: yading@uchicago.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Chicago Chicago, IL 60637				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The purpose of the study is to develop correlative feature analysis methods for integrating image information from multi-modality breast images, taking advantage of the information from different views and/or different modalities, and thus improving the sensitivity and specificity of breast cancer diagnosis. Identifying the corresponding image pair of a lesion is an essential step for this purpose. During the past three years, we have built a multi-modality database which includes FFDM , breast US and DCE-MR images. We also developed computerized correlative feature analysis methods including automatic lesion segmentation, feature extraction and selection, feature correlation analysis and image pair classification in differentiating corresponding and non corresponding lesions across different mammographic views and/or different imaging modalities. The results show that the proposed correlative feature analysis is effective and robust for the discrimination between corresponding and non-corresponding lesion pairs.					
15. SUBJECT TERMS Multi-modality database, lesion segmentation, feature extraction, feature selection, correlative feature analysis, correspondence analysis					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 66	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

Table of Contents

	<u>Page</u>
Introduction.....	4
Body.....	5
Key Research Accomplishments.....	17
Reportable Outcomes.....	19
Conclusion.....	22
References.....	23
Appendices.....	25

INTRODUCTION

In breast imaging, it has been well recognized that merging information from different images will greatly benefit the diagnosis of breast cancer [1-4], as well as contribute to the assessment of tumor response and image-guided therapy. However, interpreting images from different modalities is not trivial as different images of the same lesion may exhibit different physical lesion characteristics, and the image acquisitions are performed under different breast positioning protocols. Also, the breast is a non-rigid object, and thus conventional image registration methods are not appropriate. So the essential problem of merging information from different modalities is to address the non-trivial task of identifying corresponding images of lesions as seen with different views and/or different imaging modality techniques. The purpose of this research is to develop correlative feature analysis methods for integrating image information from multi-modality breast images, taking advantage of the information from different views and/or different modalities, and thus improving the sensitivity and specificity of breast cancer diagnosis. A novel aspect of the proposed research is the development of correlative feature analysis (CFA) into the decision-making process. Our hypothesis is that the proposed correlative feature analysis can benefit computerized corresponding image analysis, leading to improved discrimination between corresponding and non-corresponding lesion pairs. This report summarizes the progress of this Predoctoral Traineeship Award project made by the recipient during the past three years.

BODY

Training Accomplishments

At the time of this report, the recipient of the Predoctoral Traineeship Award has taken 22 out of the 22 required courses towards the Ph.D. degree in medical physics. The courses include physics of medical imaging, physics of radiation therapy, mathematics for medical physicists, image processing, statistics, machine learning, numerical computation, computer vision, anatomy of the body, radiation biology, and teaching assistant training.

Research Accomplishments

1. Database collection

We have collected a multi-modality breast image database from the University of Chicago Medical Center (UCMC) under protocols approved by the Institute Board Review (IRB). This database includes full-field digital mammographic (FFDM) images, breast ultrasound (US) images and breast dynamic contrast-enhanced magnetic resonance (DCE-MR) images. The FFDM database consists of 229 malignant and 164 benign lesions. All the images were obtained from GE Senographe 2000D systems with a spatial resolution of $100\mu\text{m}\times 100\mu\text{m}$. The US database consists of 205 malignant solid lesions, 113 simple cysts and 139 benign solid lesions. The US images were obtained with a Philips HDI 5000 US unit and a 12-5MHz linear array probe. The pixel size varied from $53\mu\text{m}$ to $212\mu\text{m}$, with the average value of $114\mu\text{m}$. The MR database consists of 359 malignant and 133 benign lesions. In MR imaging, patients were scanned in the prone position with a standard double breast coil on a 1.5 T whole-body GE MRI system. Images were obtained by the use of a T1-weighted 3D spoiled gradient echo sequences. For each case, one pre-contrast and five post-contrast series were taken. Each series contained 60 coronal slices with an image of 128×256 pixels and a range of planar spatial resolution from $1.25\times 1.25\text{mm}^2$ to $1.6\times 1.6\text{mm}^2$. Slice thickness ranged from 3 to 4 mm depending on breast size.

All the lesions in the multi-modality database were identified by expert breast radiologists based on visual criterion and either biopsy or aspiration proven reports. An expert radiologist (Dr. C. S.) also helped us identify the correspondence of lesions appeared in different modality images.

2. Investigation of lesion segmentation on FFDM images

2.1 Development of a dual-stage segmentation algorithm

Mass lesion segmentation on mammograms is a challenging task since mass lesions are usually embedded and hidden in varying densities of parenchymal tissue structures. We have developed a dual-stage method for automatic delineation of lesion boundaries on FFDM images [5]. This method utilizes a geometric active contour model that minimizes an energy function based on the homogeneities inside and outside of the evolving contour.

Prior to the application of the active contour model, a radial gradient index (RGI) based segmentation method is applied to yield an initial contour closer to the lesion boundary location in a computationally efficient manner. Based on the initial segmentation, an automatic background estimation method is applied to identify the effective circumstance of lesion, and a dynamic stopping criterion is implemented to terminate the contour evolution when it reaches the lesion boundary. A full description of the method is in reference [5] which is attached as Appendix A.

2.2 Evaluation of the proposed segmentation algorithm

By using the FFDM database described above, we compared, both qualitatively and quantitatively, the proposed algorithm with a conventional region-growing method [6] and a RGI-based algorithm [7]. Figure 1 shows an example of lesion segmentation using region growing, RGI-based segmentation alone and the proposed dual-stage segmentation method. The result of the dual-stage segmentation method visually demonstrates a closer agreement with the radiologist's outline of the lesion.

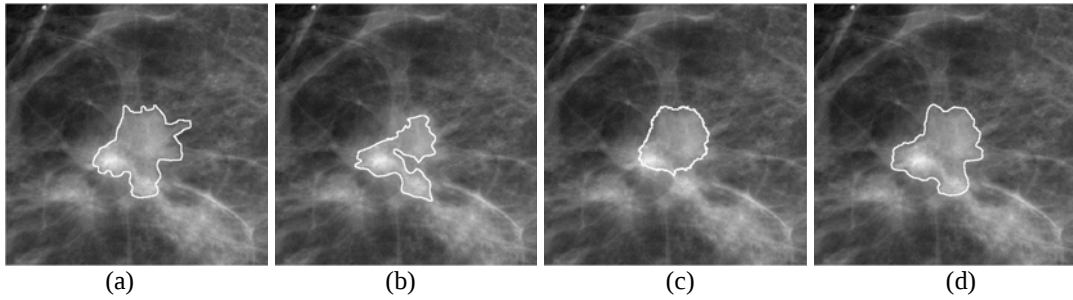


Figure 1: Segmentation results for a malignant lesion. (a) radiologist's outline, (b) region growing, (c) RGI-based segmentation and (d) the dual-stage segmentation method.

We quantitatively evaluated the performance difference among the three automatic segmentation methods by use of the area overlap ratio between computer segmentation and manual segmentation by an expert radiologist. Figure 2 shows the fraction of lesions correctly segmented at various overlap threshold levels. At the overlap threshold of 0.4, for benign lesions, 87% of the images are correctly segmented with the proposed method, while 72% and 81% of the images are correctly segmented by the region-growing and RGI-based methods, respectively. For malignant lesions, 84% of the images are correctly segmented with the proposed method, while 66% and 67% of the images are correctly segmented by region-growing and RGI-based methods, respectively.

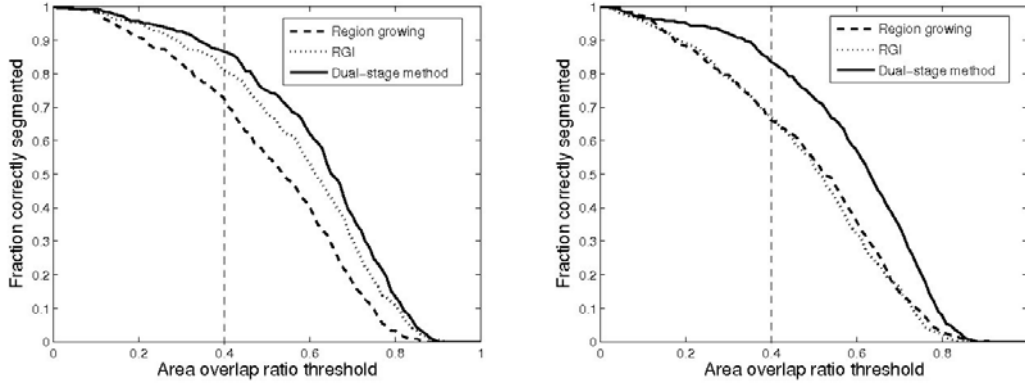


Figure 2: Performance of three different segmentation methods in terms of area overlap ratio on a clinical FFDM database. Left: Evaluated on 327 benign images; Right: Evaluated on 412 malignant images. The results show that the dual-stage segmentation method is statistically superior to both region-growing and RGI-based method.

2.3 Investigation of lesion segmentation to FFDM computer-aided diagnosis (CAD)

We continued to evaluate the performance of the proposed segmentation algorithm for the task of assessing the likelihood of malignancy of lesion. As characteristic features are extracted from the segmented lesion itself, or the neighborhood of the segmented lesion, the more accurately segmented lesions, the more meaningful and stable features are expected. In our study [8], 15 features [9] were automatically extracted to quantify the characteristics of spiculation, margin, contrast, shape and texture. An effective subset of features were automatically selected by a stepwise method and merged with a Bayesian Artificial Neural Network (BANN) [10] to yield a discriminant score, which estimates the probability of malignancy (PM) for a given lesion. The performance of individual features and the selected feature subset was evaluated using receiver operating characteristic (ROC) analysis [11], with the area under the ROC curve (AUC) as a figure of merit.

We compared the classification performance of the proposed method with that of a conventional region-growing method, which was used for our previously developed CAD system for mammographic images. The FFDM dataset used in this study included 146 malignant and 134 benign cases. The results showed that the performances of most of the spiculation features were improved with the dual-stage segmentation method. However, we failed to observe improvement for other features. In leave-one-out evaluation by lesion, the effective feature subset by the dual-stage segmentation, including two spiculation features and one gradient texture, yielded an AUC of 0.78, while the prior-selected feature subset from the region growing segmentation, including three spiculation features, one margin sharpness and one average gray level of lesion, yielded an AUC of 0.72. The difference is statistically significant ($p=0.04$). This work has been reported at the AAPM annual meeting, 2008. A more detailed summary is in reference [8], which is also attached as Appendix B.

3. Investigation of correlative feature analysis across (CFA) mammographic views

We firstly investigated on a dual-stage correlative feature analysis method to address the task of classifying corresponding images of lesions as seen in different mammographic views [12]. The proposed CFA method is different from conventional image registration methods in the following two aspects. (1) The task of image registration is to align two images known to represent the same object, while CFA is to assess the probability that the given two images represent the same object. (2) The key point of image registration is to determine a geometrical transformation that minimizes some cost functions defined by intensities, contours and mutual information, in which various geometrical landmarks, such as control points and inherent image landmarks are identified and matched. The CFA technique is feature based, which is motivated by the studies on fusion of two-view information for computer-aided detection, as well as our prior research on the task of automated classification of breast lesions, i.e., in the determination of benign and malignant breast lesions based on computer-extracted features. A full description of the method is in reference [12], which is also attached as Appendix C.

3.1 Design of corresponding and non-corresponding datasets

Based on the correspondence of lesions identified by the radiologist, we constructed 123 corresponding pairs and 82 non-corresponding pairs [12]. Each pair consists of a cranio-caudal (CC) view and a medio-lateral (ML) view. Considering the most realistic scenario of lesion mismatch in clinical practice, the non-corresponding pairs were constructed from cases of the same patients but different physical lesions. In Figure 3, (a) shows an example with multiple lesions seen on mammograms in CC and ML views, and (b) lists the corresponding and non-corresponding pairs constructed from (a).

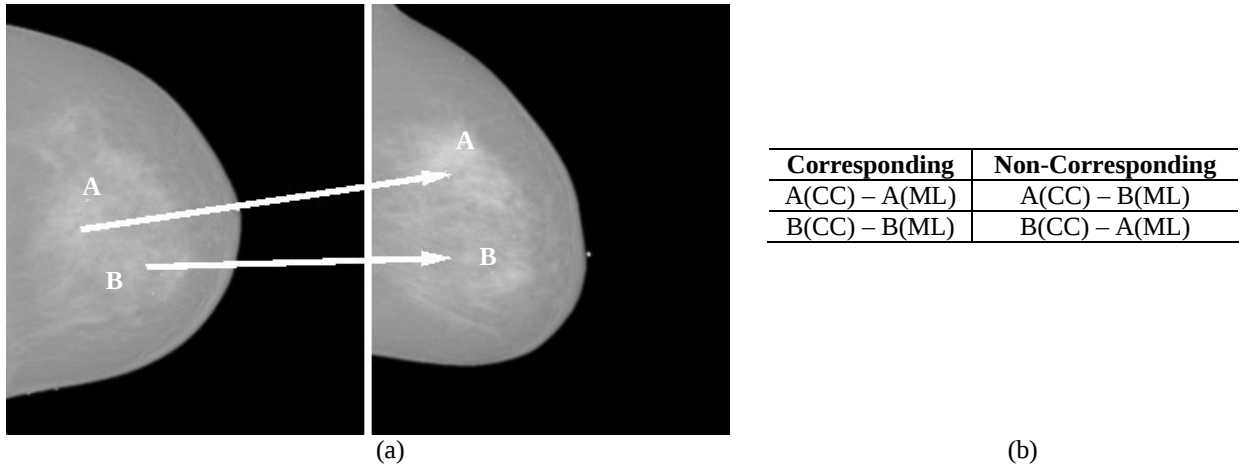


Figure 3: (a) An example of two lesions in the same breast as seen in CC view (left) and ML view (right). The arrow indicates the correspondence of the same physical lesion in different views. (b) The corresponding and non-corresponding pairs constructed from (a).

3.2 Investigation of feature correlation criterion

We evaluated the correlation performance of individual computerized features extracted from the FFDM images of a lesion obtained in CC and ML views. In order to evaluate the robustness of the correlation performance to lesion segmentation, besides the radiologist's outlines, three automatic segmentation methods were employed to extract the mass lesion from the surrounding tissues, which includes a conventional region-growing method, a RGI-based method and the newly-developed dual-stage segmentation method. 15 computer-extracted features, which have been described in Section 2.3, were calculated from each lesion in both views. For each feature, Pearson correlation coefficient between the two views and the p-value of the derived correlation coefficient were obtained. Our results show that the features characterizing shape, contrast and texture performed better among the 15 individual features despite of segmentation methods and pathology. This is because the features representing large-scale information are less sensitive to the change of position than those representing small-scale information, which results in the higher correlation between large-scale features from different views than that of small-scale features. This work provides a guide for discriminating corresponding and non-corresponding lesion pairs within the CAD framework. It is also helpful for guiding the development of new features to improve the accuracy of image matching in disease diagnosis and prognosis. A more detailed summary can be found in reference [13] which is also attached as Appendix D.

Mutual information (MI) is another measure of the dependence between two variables. It is well understood that mutual information measures the general dependence, while the correlation coefficient measures the linear dependence. So we also investigated the mutual information among the features and assessed its effect on the choice of discriminating features as compared with the use of linear correlation coefficient between features. For each feature described above, mutual information between the two views was obtained using a density estimation method (e.g., Parzen windows) [14]. However, the dependence rank of features determined by mutual information highly agreed with that determined by linear correlation coefficient, yielding a correlation coefficient of 0.87. This result indicated that linear correlation coefficient is a good metric to represent the dependence between features from different views. Moreover, since linear correlation coefficient is bounded to $[-1, 1]$, we will use linear correlation coefficient as the metric to choose the discriminating features in correlative feature analysis across mammographic views.

3.3 Development of new features

Since features characterizing large-scale information usually have better correlation performance, we developed two sets of "large-scale" features. Firstly, we extracted a set of texture features based on a gray-level co-occurrence matrix (GLCM). For each region, four GLCMs were constructed along four different directions of 0° , 45° , 90° and 135° . Assuming that there is no directional texture features in mammograms, a non-directional GLCM was obtained by summing all the directional GLCMs. Texture features were then computed from each non-directional GLCM. To avoid sparse GLCMs for smaller lesions, the gray level range of image was scaled down to 6 bits, resulting in GLCM of size

64×64. Among the texture features, correlation feature performed best with a correlation coefficient of 0.67 (p-value < 10^{-3}) [12].

In clinic practice, radiologists commonly use the distance from nipple to the center of a lesion to correlate the lesion in different views. It is generally believed that this distance keeps fairly constant. Thus, we developed a distance feature to measure the Euclidean distance between the nipple location and the mass center of lesion. We also developed an automatic nipple localization scheme to tracking nipple markers on each FFDM images. With computer-identified nipples, the distance features in CC views are highly correlated with those in ML views, yielding a correlation coefficient of 0.88 (p-value < 10^{-3}) [12].

3.4 Investigation of scheme to combine multiple classifiers in correlative feature analysis

For each pairwise set of features extracted from lesions in CC and ML views, a BANN classifier was employed to merge each feature pair into a correspondence score, which is the estimate of the probability that the two lesion images depict the same physical lesion. As an example, Figure 4(a) shows the distribution of distance feature generated from lesions taken in different views for corresponding and non-corresponding datasets. The histogram in Figure 4(b) demonstrates, for the corresponding and non-corresponding datasets, the distribution of these correspondence scores output from the BANN.

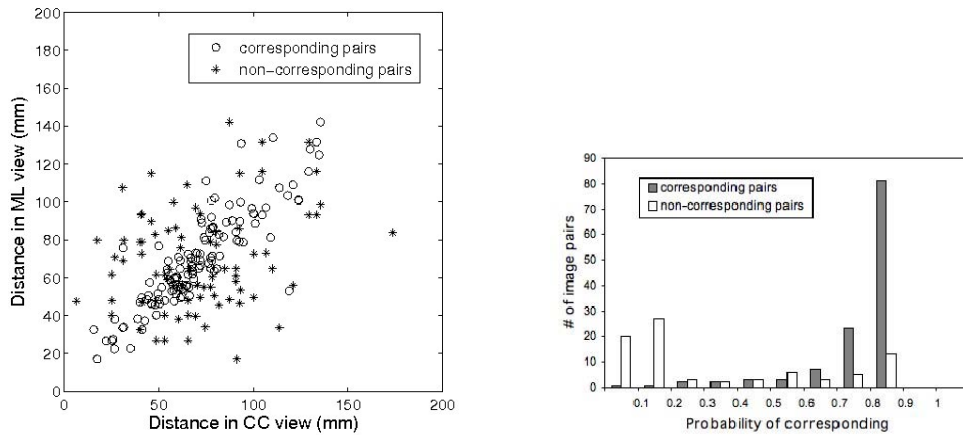


Figure 4: (a) The scatter plot of distance feature generated from lesions seen on CC and ML views. (b) The distribution of the output correspondence scores of the distance feature from a BANN classifier.

Linear stepwise feature selection [15] with Wilks lambda criterion was employed on all feature-based correspondence scores to select a subset of scores for the final task of distinguishing corresponding and non-corresponding pairs. Another BANN was then trained with the selected correspondence scores to yield an overall estimate of probability of correspondence.

With leave-one-out (by physical lesion) validation, the distance feature performed best among all the individual features, yielding an AUC of 0.81 [12]. A subset of features, which included distance, ROI-based correlation and gradient texture, was selected and

merged with BANN to yield an AUC of 0.87. The improvement by using multiple features was statistically significant compared to that of single feature ($p = 0.01$).

3.5 Investigation of robustness of the proposed correlative feature analysis

To evaluate the robustness of the proposed correlative feature analysis over different mammographic view pairs, we extended the application of CFA to CC versus medio-lateral oblique (MLO) views [16]. We constructed 104 corresponding image pairs and 95 non-corresponding pairs. The distance was also the best individual feature with AUC of 0.78 ± 0.03 . The selected feature subset, including distance, ROI-based energy and ROI-based homogeneity, yielded an AUC of 0.88 ± 0.02 . This improvement was statistically significant ($p < 0.001$). Although most of the selected mathematical descriptors were different from those based on CC versus ML views, they did represent the same physical characteristic, i.e. texture. From the entire FFDM database, we constructed a dataset of lesions with CC, MLO and ML views, and obtained 83 corresponding pairs and 66 non-corresponding pairs. The leave-one-out (by physical lesion) was performed on the paired lesions only. The proposed CFA method yielded an AUC of 0.87 ± 0.02 for CC versus ML views and 0.90 ± 0.02 for CC versus MLO views. The difference was not statistically significant ($p=0.49$). The results showed that CFA was robust across two different view pairs (CC versus ML and CC versus MLO). This work has been reported in the 9th International Workshop on Digital Mammography (IWDM) [16], please refer to Appendix E for more details.

4. Investigation of correlative feature analysis across imaging modalities

The computerized method developed in the above work can be extended for assessing the probability of correspondence across different imaging modalities. Here our investigation will be based on FFDM images versus DCE-MR images. The key difference between this scheme and the prior mammographic one is that, in DCE-MR images, all computerized features will be extracted in three-dimensional (3D) space. Kinetic, morphologic, and texture features are widely used in the computerized analysis of breast DCE-MR images [17, 18]. However, in the proposed correlative feature analysis study, only morphologic and texture features will be used, because FFDM images do not carry functional information, such as enhancement-variance dynamics.

4.1 Design of corresponding and non-corresponding datasets

Depending on which mammographic view was used, we constructed three data sub-categories from the multi-modality database: MR images versus CC views, MR images versus ML views, and MR images versus MLO views. Note that a mammographic image does not correlate with a single MR image, but a set of MR images depicting a physical lesion in 3D space. Table 1 lists the detailed information regarding the corresponding and non-corresponding datasets in each sub-category.

Table 1: The number of lesion pairs in corresponding and non-corresponding datasets. The non-corresponding pairs were constructed from cases of the same breasts but different physical lesions.

	Corresponding dataset	Non-corresponding dataset
MR vs. CC view	243	98
MR vs. ML view	219	88
MR vs. MLO view	122	77

4.2 Investigation of feature correlation criterion

In the correlative feature analysis across mammographic views, we used Pearson correlation coefficient as the criterion to evaluate the correlative performance of computerized features from different views. Pearson correlation coefficient is a widely used criterion that measures the linear correlation between two feature vectors. However, it is more natural to extract a general correlation between two feature vectors when we compare images from different modalities. Thus, we introduced Spearman rank order correlation coefficient [19] in this study.

As compared to Pearson correlation coefficient, Spearman rank order correlation coefficient is a non-parametric measure of correlation, that is, it assesses how well an arbitrary monotonic function could describe the relationship between two variables, without make any other assumptions about the particular nature of the relationship between the variables. Thus, Spearman rank order correlation coefficient could measure the non-linear relationship between feature vectors. Although mutual information (MI) can also depict the non-linear relationship between variables, its calculation, however, is not standardized and thus hard to infer statistically. Moreover, the MI value is not bounded. Therefore, we used Spearman rank order correlation coefficient as the metric to choose the discriminating features in correlative feature analysis across different imaging modalities.

We used lesion size feature to demonstrate the effectiveness of Spearman rank order correlation. The lesion size feature in mammograms (F_{MG}) is conventionally calculated as the equivalent diameter that measures the diameter of a circle yielding the same area as the segmented lesion, while the lesion size feature in MR images (F_{MR}) is conventionally calculated as the volume of the segmented lesion. The equivalent diameter of lesion in MR images can be obtained as: $F'_{MR} = \sqrt[3]{6 \times F_{MR}}$. In Figure 5, (a) shows the scatter plot between F_{MG} and F_{MR} , while (b) shows the scatter plot between F_{MG} and F'_{MR} . Table 2 lists the associated Pearson correlation coefficient (γ) and Spearman rank order correlation coefficient (ρ). The results show that Spearman rank order correlation coefficient could effectively extract the non-linear relationship between features obtained from MG images and MR images.

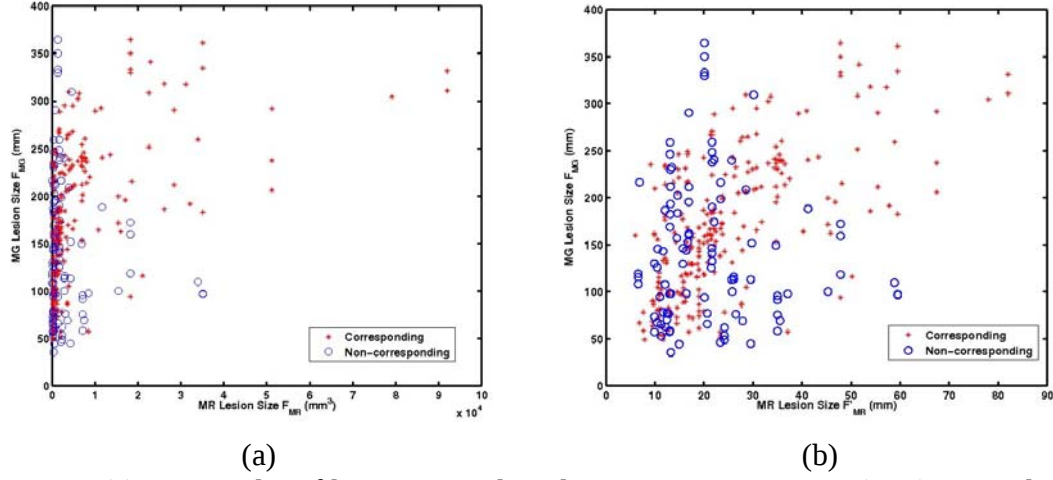


Figure 5: (a) Scatter plot of lesion equivalent diameter in MG images (F_{MG}) versus lesion volume in MR images (F_{MR}); (b) Scatter plot of lesion equivalent diameter in MG images (F_{MG}) versus lesion equivalent diameter in MR images (F'_{MR}).

Table 2: Pearson correlation coefficient (γ) versus Spearman rank order correlation coefficient (ρ).

	F_{MG} versus F_{MR}		F_{MG} versus F'_{MR}	
	γ	ρ	γ	ρ
Corresponding dataset	0.49	0.65	0.65	0.65
Non-corresponding dataset	-0.11	0.03	-0.06	0.03

4.3 Investigation of geometric features

Investigation of geometric features is motivated by the observation that radiologists ordinarily use location information to identify lesions in different images. Also, in the previous study of correlative feature analysis across mammographic views, we have shown that the distance feature performs well for the task of finding the corresponding lesions in different mammographic views.

Figure 6 illustrates the geometric configurations of CC view, MLO view and MR images. Note that we did not include ML view in the figure because its geometric property is close to that of MLO view. We developed an automatic method to locate the mass center of a lesion $L(L_x, L_y, L_z)$ in a Cartesian coordinate system. The origin of the coordinate system is set at the nipple, and its z direction is from nipple to the chest wall. Then, we extracted geometric feature pairs depending on which mammographic view was used, as listed in Table 3.

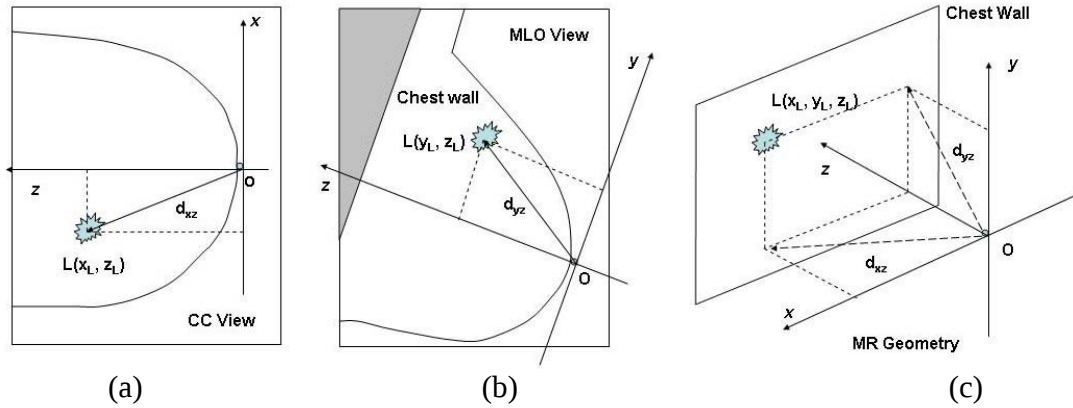


Figure 6: Illustration of geometric configuration of CC view (a), MLO view (b) and MR images (c).

Table 3: Geometric feature pairs between MR images and mammographic views. d_{xz} is calculated as the projection of the Euclidean distance between nipple and the center of lesion on the x-z plane, and d_{yz} is calculated as the projection of the Euclidean distance between nipple and the center of lesion on the y-z plane.

MR images	CC view	MLO view	ML view
x	x	-	-
y	-	y	y
z	z	z	z
d_{xz}	d_{xz}	-	-
d_{yz}	-	d_{yz}	d_{yz}

For each pair-wise set of geometric features extracted from lesions in MR images and mammographic views, a BANN classifier was employed to merge each feature pair into a correspondence score, which is the estimate of the probability that the two lesion images depict the same physical lesion. Table 4 lists the AUC values of correspondence scores obtained from geometric feature pairs, which were evaluated with leave-one-out (by lesion) validation.

Table 4: Performance of correspondence scores obtained from geometric feature pairs.

Geometric features	MR vs. CC (AUC $\pm$ se)	MR vs. MLO (AUC $\pm$ se)	MR vs. ML (AUC $\pm$ se)
x	0.59 ± 0.04	-	-
y	-	0.72 ± 0.04	0.69 ± 0.03
z	0.70 ± 0.03	0.74 ± 0.04	0.66 ± 0.04
d_{xz}	0.65 ± 0.03	-	-
d_{yz}	-	0.75 ± 0.04	0.73 ± 0.03

4.4 Evaluation of the proposed CFA method

Linear stepwise feature selection with Wilks lambda criterion was employed on all feature-based correspondence scores to select a subset of scores for the final task of distinguishing corresponding and non-corresponding pairs. Another BANN was then trained with the selected correspondence scores to yield an overall estimate of probability of correspondence.

For MR images versus CC view, a subset of features, which included three morphological features, three texture features and two geometric features, was selected and merged with BANN to yield an AUC of 0.78, with leave-one-out (by physical lesion) validation. The improvement by using multiple features was statistically significant compared to that of single feature ($p = 0.02$).

For MR images versus MLO view, a subset of features, which included lesion size, two texture features and two geometric features, was selected and merged with BANN to yield an AUC of 0.86, with leave-one-out (by physical lesion) validation. The improvement by using multiple features was statistically significant compared to that of single feature ($p = <0.001$).

For MR images versus ML view, a subset of features, which included two morphological features and three geometric features, was selected and merged with BANN to yield an AUC of 0.84, with leave-one-out (by physical lesion) validation. The improvement by using multiple features was statistically significant compared to that of single feature ($p = <0.001$).

5. Multi-modality breast cancer classification with mammography and DCE-MRI

In order to evaluate the roles of corresponding lesions and their features in breast CAD, we investigated the multimodality breast cancer classification with mammography [9] and DCE-MR images [17]. In this pilot study [20], we used a FFDM database including 321 lesions (167 malignant and 154 benign) and a DCE-MRI database including 181 lesions (97 malignant and 84 benign). From these two databases, we constructed a multimodality dataset of 51 lesions (29 malignant and 22 benign). Mammograms and DCE-MR images are available for these lesions.

For each lesion on each modality, computer automatically segmented the mass lesions and extracted a set of features. Linear stepwise feature selection was firstly performed on single modalities, yielding one feature subset for each modality. Then, these selected features served as the input to another feature selection procedure when extracting useful information from both modalities. The selected features were merged by a Linear Discriminant Analysis (LDA) into a discriminant score. ROC analysis was used to evaluate the performance of each selected feature subset in the task of distinguishing between malignant and benign lesions.

With leave-one-out (by physical lesion) evaluation on the multimodality dataset, the mammography-only features yielded an AUC of 0.62 and the DCE-MRI-only features

yielded an AUC of 0.90. The combination of these two modalities, which included a spiculation feature from mammography and a kinetic feature from DCE-MRI, yielded an AUC of 0.94. The improvement of combining multi-modality information was statistically significant as compared to the use of mammography only ($p=10^{-4}$). However, we failed to show statistically significant improvement with the multi-modality features as compared to DCE-MRI, mostly due to the limited multi-modality dataset ($p=0.22$).

In the previous studies [9][17], spiculation and kinetic features have been justified as the best features when distinguishing malignant and benign lesions for mammography and DCE-MRI, respectively. Our feature selection method correctly captured these two features when combining information from different modalities. The results showed that combining information from multiple modalities performed better than the single modality in the task of distinguishing between malignant and benign lesions. This work has been reported in SPIE Medical Imaging Conference, 2009 [20], please refer to Appendix F for more details.

KEY RESEARCH ACCOMPLISHMENTS

- Built a multi-modality database, which includes full-field digital mammograms, breast ultrasound images and breast DCE-MR images.
- Developed a dual-stage lesion segmentation method for FFDM images, which outperformed the performances of our previous developed region-growing method and the RGI-based segmentation method.
- Evaluated the proposed dual-stage segmentation method for the task of assessing the likelihood of malignancy of a mass lesion on FFDM images, which yielded improved classification performance over that with region-growing method.
- Investigated feature correlation with both linear correlation coefficient and mutual information. The results demonstrate that the features representing large-scale information of lesions usually have better correlation performance and linear correlation coefficient is an appropriate metric characterizing the dependence between features from different views.
- Developed texture features and distance feature, which have been proven to be useful in differentiating corresponding and non-corresponding image pairs.
- Developed a computerized correlative feature analysis (CFA) framework to identify the correspondence between lesions imaged in different images. The two-step classification scheme not only effectively utilizes the information regarding correlation between feature pairs, but also efficiently combines multiple classifiers into a final decision.
- Evaluated the proposed correlative feature analysis on two sets of pair-wise mammographic views, i.e. CC versus ML and CC versus MLO. The results show that the proposed correlative feature analysis is effective and robust across different view pairs.
- Investigated feature correlation with both Pearson correlation coefficient and Spearman rank order correlation coefficient. The results demonstrate that Spearman rank order correlation coefficient could effectively extract non-linear relationship between features obtained from mammographic images and MR images.
- Developed geometric features to correlate lesions in 3D MR images and lesions in 2D projection MG images, which have been proven to be useful in distinguishing corresponding lesions from non-corresponding ones.
- Evaluated the proposed correlative feature analysis across MR images and different mammographic views. The results show that the proposed method can achieve statistically improved performance compared to the use of single feature in the task of differentiating corresponding and non-corresponding lesions.

- Conducted a pilot study on computerized diagnosis of breast lesions with mammography and DCE-MRI. The results showed that combining information from multiple modalities performed better than the single modality in the task of distinguishing between malignant and benign lesions.

REPORTABLE OUTCOMES

Peer-reviewed Journal Papers

- [1] **Y. Yuan**, M. L. Giger, H. Li, K. Suzuki, and C. A. Sennett, "A dual-stage method for lesion segmentation on digital mammograms," *Medical Physics*, 34(11), 4180-4193 (2007).
- [2] **Y. Yuan**, M. L. Giger, H. Li, and C. A. Sennett, "Correlative feature analysis on FFDM," *Medical Physics*, 35(12), 5490-5500 (2008).
- [3] H. Li, M. L. Giger, **Y. Yuan**, W. Chen, K. Horsch, L. Lan, A. R. Jamieson, C. A. Sennett, and S. A. Jensen, "Evaluation of computer-aided diagnosis on a large clinical full-field digital mammographic dataset," *Academic Radiology*, 15, 1437-1445 (2008).

Conference Proceeding Papers

- [1] **Y. Yuan**, M. L. Giger, K. Suzuki, H. Li, and A. R. Jamieson, "A two-stage method for lesion segmentation on digital mammograms," *Proceedings of SPIE, Medical Imaging 2006: Image processing*, vol. 6144, 61443W (2006).
- [2] H. Li, M. L. Giger, **Y. Yuan**, K. Suzuki, J. A. Jamieson, and C. A. Sennett, "Comparison of computerized image analyses for digitized mammograms and full-field digital mammography images," *Proceedings of the 8th International Workshop on Digital Mammography*, vol. 4046, 569-575 (2006).
- [3] M. L. Giger, **Y. Yuan**, H. Li, K. Drukker, W. Chen, L. Lan, and K. Horsch, "Progress in breast CADx," *Proceedings of IEEE International Symposium on Biomedical Imaging 2007*, 508-511 (2007).
- [4] **Y. Yuan**, M. L. Giger, H. Li, L. Lan, and C. A. Sennett, "Identifying corresponding lesions from CC and MLO views via correlative feature analysis," *Proceedings of the 9th International Workshop on Digital Mammography*, vol. 5116, 323-328 (2008).
- [5] H. Li, M. L. Giger, **Y. Yuan**, L. Lan, and C. A. Sennett, "Performance of CADx on a large clinical database of FFDM images," *Proceedings of the 9th International Workshop on Digital Mammography*, vol. 5116, 510-514 (2008).
- [6] **Y. Yuan**, M. L. Giger, H. Li, and C. A. Sennett, "Correlative feature analysis of FFDM images," *Proceedings of SPIE, Medical Imaging 2008: Computer-aided diagnosis*, vol. 6915, 69151L (2008).
- [7] **Y. Yuan**, M. L. Giger, H. Li, and C. A. Sennett, "Breast lesion classification using mammography and DCE-MRI," *Proceedings of SPIE, Medical Imaging 2009: Computer-aided diagnosis*, vol. 7260, 72600O (2009).

Conference Presentations and Abstracts

- [1] **Y. Yuan**, M. L. Giger, K. Suzuki, H. Li, and A. R. Jamieson, "A two-stage method for lesion segmentation on digital mammograms," Poster presented at *SPIE International Symposium: Medical Imaging 2006*, San Diego, California February, 2006. (Honorable Mention Poster Award)
- [2] **Y. Yuan**, M. L. Giger, H. Li, K. Suzuki, A. R. Jamieson, and C. A. Sennett, "Comparison of image segmentation algorithms on digitized mammograms and FFDM images for CAD," *Medical Physics* 33: 2195-2196, 2006. Presented at the *48th Annual Meeting of American Association of Physicists in Medicine*, Orlando, Florida, July, 2006.
- [3] **Y. Yuan**, M. L. Giger, H. Li, and C. A. Sennett, "Computer-based feature correlation on multiple-view FFDM images," Poster presented at the *92nd Scientific Assembly and Annual Meeting of Radiological Society of North America*, Chicago, Illinois, December 2006.
- [4] H. Li, M. L. Giger, **Y. Yuan**, W. Chen, C. A. Sennett, L. Lan, A. R. Jamieson, K. Suzuki, and S. Arkani, "Conversion of screen-film mammographic CADx for FFDM," *The 92nd Scientific Assembly and Annual Meeting of Radiological Society of North America*, Chicago, Illinois, December 2006.
- [5] W. Chen, M. L. Giger, G. M. Newstead, S. Jansen, K. Chiang, L. Lan, **Y. Yuan**, A. R. Jamieson, and H. Li, "Breast cancer diagnosis using DCE-MRI: Role of computer in interpretation of 4D data," Scientific exhibit presented at the *93rd Scientific Assembly and Annual Meeting of Radiological Society of North America*, Chicago, Illinois, December 2007.
- [6] **Y. Yuan**, M. L. Giger, H. Li, and C. A. Sennett, "Correlative feature analysis of FFDM images," *Presented at SPIE International Symposium: Medical Imaging 2008*, San Diego, California, February, 2008.
- [7] **Y. Yuan**, M. L. Giger, H. Li, and C. A. Sennett, "Correlative feature analysis for multi-modality breast CAD: Identifying the corresponding lesions from different mammographic views," Poster presented at *Era of Hope 2008 - Department of Defense Breast Cancer Research Program Meeting*, Baltimore, Maryland, June, 2008.
- [8] **Y. Yuan**, M. L. Giger, H. Li, L. Lan, and C. A. Sennett, "Identifying corresponding lesions from CC and MLO views via correlative feature analysis," Presented at the *9th International Workshops on Digital Mammography*, Tucson, Arizona, July, 2008.
- [9] H. Li, M. L. Giger, **Y. Yuan**, L. Lan, and C. A. Sennett, "Performance of CADx on a large clinical database of FFDM images," *The 9th International Workshops on Digital Mammography*, Tucson, Arizona, July, 2008.
- [10] **Y. Yuan**, M. L. Giger, H. Li, L. Lan, and C. A. Sennett, "Comparison of image segmentation methods on classification performance of FFDM CAD," *Medical Physics*, 35: 2941, 2008. Presented at the *50th Annual Meeting of American Association of Physicists in Medicine*, Houston, Texas, July, 2008.
- [11] **Y. Yuan**, M. L. Giger, H. Li, and C. A. Sennett, "Breast lesion classification using mammography and DCE-MRI," Presented at *SPIE International*

Symposium: Medical Imaging 2009, Lake Buena Vista, Florida, February, 2009.

Honors and Awards

- 2009 The Lawrence Lanzl Graduate Student Fellowship in Recognition of Exceptional Research Ability, Graduate Programs in Medical Physics, the University of Chicago
- 2006, 2008 Women's Board Travel Awards in the Division of Biological Science, the University of Chicago
- 2006 Honorable Mention Poster Award, SPIE Medical Imaging, San Diego, CA
- 2006 Young Investigator Presentation Award, AAPM Midwest Chapter 2006

CONCLUSIONS

The recipient of the Predoctoral Traineeship Award has finished all the required courses towards his Ph.D. degree. These trainings have proven useful for the recipient to achieve the proposed research goals.

During the research, we have collected and maintained a multi-modality database including full-field digital mammograms, breast ultrasound images and breast MR images, which is suitable for the proposed research on correlative feature analysis for multi-modality images. We have developed a dual-stage lesion segmentation for FFDM images. Our results suggest that the proposed method could improve the segmentation performance and yield improved classification performance over that with previous-developed region growing method.

In this project, we developed a computerized method to identify the corresponding lesions across mammographic views. The investigation of feature correlation showed that linear correlation coefficient is an appropriate metric characterizing the dependence between features from different views. We also developed texture features and distance feature, which have been prove to be useful in differentiating corresponding and non-corresponding image pairs. The proposed two-step correlative feature analysis not only effectively utilizes the information regarding correlation between feature pairs, but also efficiently combines multiple classifiers into a final decision. The results of a further investigation with different mammographic views have shown that the proposed correlative feature analysis is effective and robust.

In order to extend the proposed correlative feature analysis to multi-modality images, we have compared Pearson correlation coefficient and Spearman rank order correlation coefficient. Our results demonstrated that Spearman rank order correlation coefficient could effectively extract non-linear relationship between features obtained from mammographic images and MR images. By including the newly developed geometric features, we evaluated the performance of the proposed correlative feature analysis across MR images and various mammographic views. The results showed that the proposed method can achieve statistically improved performance compared to the use of single feature in the task of differentiating corresponding and non-corresponding lesions.

At last, we conducted a pilot study on computerized diagnosis of breast lesions with mammography and DCE-MRI. The results showed that combining information from multiple modalities performed better than the single modality in the task of distinguishing between malignant and benign lesions.

Overall, we have achieved the goals for this research and laid down a solid foundation for recipient's continuation of breast cancer research. The future work will focus on the research that evaluates how the proposed correlative feature analysis will facilitate radiologists for the task of identifying corresponding lesions in different images.

REFERENCES

- [1] S. H Heywang-Kobrunner, D. D Dershaw, and I. Schreer, *Diagnostic breast imaging: Mammography, sonography, magnetic resonance imaging, and interventional procedures*, Thieme Medical Publishers, 2nd Edition, 2001.
- [2] K. Drukker, K. Horsch, and M. L. Giger, "Multimodality computerized diagnosis of breast lesions using mammography and sonography", *Acad. Radiol.*, vol.12, 970-979, 2005.
- [3] J. L. Jesneck, J. Y. Lo, and J. A. Baker, "Breast mass lesions: computer-aided diagnosis methods with mammographic and sonographic descriptors," *Radiology*, vol. 244, pp. 390-398, 2007.
- [4] G. M. Newstead, "Role of MR in breast imaging," RSNA Categorical Course in Breast Imaging, 287-293, 1999.
- [5] Y. Yuan, M. L. Giger, H. Li, K. Suzuki and C. Sennett, "A dual-stage method for lesion segmentation on digital mammograms", *Med. Phys.*, vol. 34, pp. 4180-4193, 2007.
- [6] Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick and P. Lu, "Analysis of spiculation in the computerized classification of mammographic masses," *Med. Phys.*, vol. 22, pp. 1569-1579, 1995.
- [7] M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Trans. Med. Imaging*, vol 17, pp. 510-517, 1998.
- [8] Y. Yuan, M. L. Giger, H. Li, L. Lan, and C. A. Sennett, "Comparison of image segmentation methods on classification performance of FFDM CAD," *Med. Phys.*, vol 35, pp. 2941, 2008.
- [9] Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," *Acad. Radiol.*, vol 5, pp. 155-168, 1998.
- [10] C. M. Bishop, *Neural networks for pattern recognition*. Oxford, UK: Oxford University, 1995.
- [11] C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.* Vol. 21, pp. 720-733, 1986.
- [12] Y. Yuan, M. L. Giger, H. Li and C. Sennett, "Correlative feature analysis on FFDM", *Med. Phys.*, vol 35, pp. 5490-5500, 2008
- [13] Y. Yuan, M. L. Giger, H. Li and C. Sennett, "Computer-based feature correlation on multiple-view FFDM images", Radiological Society of North America, Chicago, Illinois, 2006
- [14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. Wiley, 2000.
- [15] P. A. Lachenbruch, *Discriminant analysis*. Hafner, London, 1975.

- [16] Y. Yuan, M. L. Giger, H. Li, L. Lan and C. Sennett, "Identifying corresponding lesions from CC and MLO views via correlative feature analysis," IWDM 2008, LNCS 5116, 323-328, 2008.
- [17] W. Chen, M. L. Giger, U. Bick, and G. M. Newstead, "Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI," *Med. Phys.*, vol. 33, pp. 2878-2887, 2006.
- [18] W. Chen, M. L. Giger, H. Li, U. Bick, G. M. Newstead. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. *Magnetic Resonance in Medicine*, vol 58, pp. 562-571, 2007.
- [19] G. W. Corder, and D. I. Foreman, *Nonparametric statistics for non-statisticians: A step-by-step approach*. Wiley, 2009.
- [20] Y. Yuan, M. L. Giger, H. Li, and C. A. Sennett, "Breast lesion classification using mammography and DCE-MRI," Proceedings of SPIE, Medical Imaging 2009: Computer-aided diagnosis, vol. 7260, 72600O, 2009

APPENDICES

- **Appendix A: Y. Yuan**, M. L. Giger, H. Li, K. Suzuki and C. Sennett, “A dual-stage method for lesion segmentation on digital mammograms”, *Med. Phys.*, vol. 34, pp. 4180-4193, 2007.
- **Appendix B: Y. Yuan**, M. L. Giger, H. Li, L. Lan, and C. A. Sennett, “Comparison of image segmentation methods on classification performance of FFDM CAD,” *Med. Phys.*, 35: 2941, 2008
- **Appendix C: Y. Yuan**, M. L. Giger, H. Li and C. Sennett, “Correlative feature analysis on FFDM”, *Med. Phys.*, vol 35, pp. 5490-5500, 2008
- **Appendix D: Y. Yuan**, M. L. Giger, H. Li and C. Sennett, “Computer-based feature correlation on multiple-view FFDM images”, Radiological Society of North America, Chicago, Illinois, 2006
- **Appendix E: Y. Yuan**, M. L. Giger, H. Li, L. Lan and C. Sennett, “Identifying corresponding lesions from CC and MLO views via correlative feature analysis,” IWDM 2008, LNCS 5116, 323-328, 2008.
- **Appendix F: Y. Yuan**, Y. Yuan, M. L. Giger, H. Li, and C. A. Sennett, “Breast lesion classification using mammography and DCE-MRI,” Proceedings of SPIE, Medical Imaging 2009: Computer-aided diagnosis, vol. 7260, 72600O, 2009

A dual-stage method for lesion segmentation on digital mammograms

Yading Yuan,^{a)} Maryellen L. Giger, Hui Li, Kenji Suzuki, and Charlene Sennett

Department of Radiology, Committee on Medical Physics, The University of Chicago,
5841 South Maryland Avenue-MC 2026, Chicago, Illinois 60637

(Received 10 June 2007; revised 4 September 2007; accepted for publication 4 September 2007;
published 15 October 2007)

Mass lesion segmentation on mammograms is a challenging task since mass lesions are usually embedded and hidden in varying densities of parenchymal tissue structures. In this article, we present a method for automatic delineation of lesion boundaries on digital mammograms. This method utilizes a geometric active contour model that minimizes an energy function based on the homogeneities inside and outside of the evolving contour. Prior to the application of the active contour model, a radial gradient index (RGI)-based segmentation method is applied to yield an initial contour closer to the lesion boundary location in a computationally efficient manner. Based on the initial segmentation, an automatic background estimation method is applied to identify the effective circumstance of the lesion, and a dynamic stopping criterion is implemented to terminate the contour evolution when it reaches the lesion boundary. By using a full-field digital mammography database with 739 images, we quantitatively compare the proposed algorithm with a conventional region-growing method and an RGI-based algorithm by use of the area overlap ratio between computer segmentation and manual segmentation by an expert radiologist. At an overlap threshold of 0.4, 85% of the images are correctly segmented with the proposed method, while only 69% and 73% of the images are correctly delineated by our previous developed region-growing and RGI methods, respectively. This resulting improvement in segmentation is statistically significant. © 2007 American Association of Physicists in Medicine. [DOI: [10.1118/1.2790837](https://doi.org/10.1118/1.2790837)]

Key words: Mass lesion segmentation, geometric active contour model, computer-aided diagnosis, breast cancer

I. INTRODUCTION

Breast cancer is the most common malignancy in American women and the second most common cause of death from malignancy in this population. According to the American Cancer Society, about 178,480 women in the United States will be found to have invasive breast cancer in 2007, and about 40,460 women will die from the disease this year.¹ Although some imaging modalities, such as magnetic resonance imaging (MRI)^{2,3} and sonography,^{4,5} are currently being investigated to improve sensitivity and specificity of breast cancer diagnosis, x-ray mammography is still the most prevalent imaging procedure for the early detection of breast cancer.

Lesion segmentation, which extracts the lesion from the surrounding tissues, is an essential step in the computerized analysis of mammograms. As mass lesions are usually embedded and hidden in varying densities of parenchymal structures, the task of lesion segmentation is not trivial. Many researchers have developed computer algorithms for this task. Huo *et al.*⁶ employed a region-growing method to find the contour, in which abrupt changes in size and circularity were used as the rules of segmentation. Kupinski *et al.*⁷ segmented the mass by applying either a radial gradient index (RGI) model or a probabilistic model to the lesion, multiplied by a constraint function. Petrick *et al.*⁸ introduced a segmentation algorithm that combines a density-weighted contrast enhancement filter and a region growing method. Li *et al.*⁹ employed a multiresolution Markov random field

model to detect tumors in mammographic images. Timp *et al.*¹⁰ employed both edge based information as well as a priori knowledge about the gray level distribution of the region of interest (ROI) around the mass, and obtained an optimal contour using dynamic programming. To segment lesions, Guliato *et al.*¹¹ proposed two fuzzy sets related methods—one employing a region growing after fuzzy-sets-based preprocessing, and the other using a fuzzy region-growing method that takes into account the uncertainty present around the boundaries of tumor. Li *et al.*¹² presented a statistical model for enhanced segmentation and extraction of a suspicious mass area from mammographic images. In their study, a morphological operation is derived to enhance disease patterns of suspected masses by eliminating unrelated back-ground clutter, and a model-based image segmentation is performed to localize the suspected mass areas using stochastic relaxation labeling.

Originally introduced by Kass,¹³ active contour models (or snakes) have attracted much attention as image segmentation techniques. An active contour model minimizes an energy functional along a deformable contour, which is influenced by both internal and external terms. The internal energy controls the smoothness and elasticity of the contour, while the external energy attracts the evolving contour to deform toward salient image features, such as edges. Although the active contour model has been used for segmenting objects in a wide range of medical applications,^{14–19} to the best of our knowledge, few works have applied this

model to the task of lesion segmentation in mammographic images. Brake *et al.*²⁰ segmented mass lesions by a discrete active contour method, whose external energy was determined by the image gradient magnitude. Sahiner *et al.*²¹ applied an active contour model that incorporated edge and region analysis, in which the contour energy was minimized by a greedy algorithm. In their work, however, the contour was represented by the vertices of an N -points polygon and each vertex was tracked during the process, which makes it difficult for the contour to adapt to a change of topology, such as splitting or merging parts.

Differing from the segmentation methods mentioned above, in this study we develop an automatic lesion segmentation algorithm that employs a geometric active contour model to extract lesions. Geometric active contour models^{22,23} represent contours as a level set of a higher-dimensional scalar function.²⁴ The contours are obtained only after complete evolution, thereby allowing the model to handle the topological changes naturally. As mass lesions usually have weak edges, we use a region-based active contour model²⁵ that is based on global image information, and is less sensitive to noise and the initial contour. In order to improve the computational efficiency and suppress the influence of unrelated structures, our previous RGI-based segmentation method⁷ is applied first to delineate an initial contour, which is relatively close to the actual margin, and to estimate the effective background. We then exploit a dynamic stopping criterion, which is solely based on the property of the given image, to terminate the evolving procedure automatically.

The organization of this paper is as follows: Section II introduces the database used for this study. Section III describes the proposed segmentation method. Section IV presents the results, and Sections V and VI give a discussion and conclusion, respectively.

II. MATERIALS

In this study, we used a full-field digital mammography (FFDM) database, which consists of 139 benign (327 mammograms) and 148 malignant (412 mammograms) lesions. All of the images were collected from the University of Chicago Hospitals and obtained from GE Senographe 2000D systems (GE Medical Systems, Milwaukee, WI) with a spatial resolution of $95\ \mu\text{m} \times 95\ \mu\text{m}$. The masses were identified and outlined by an expert breast radiologist based on visual criterion and biopsy-proven reports. These outlines were used as the “gold standard” for calibrating parameters and evaluating performance. The distributions of effective projection diameter, which is defined as the effective diameter of the area inside the radiologist’s manually delineated contours, are shown in Fig. 1.

III. METHODS

The main aspects of the proposed segmentation method include an initial RGI segmentation,⁷ background estimation

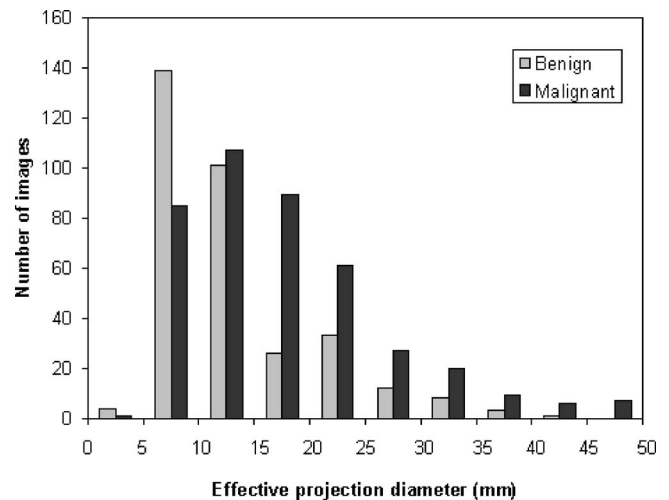


FIG. 1. Distribution of lesions' effective diameters obtained from the FFDM database.

and trend correction, and an active contour segmentation based on level sets. Figure 2 shows the flow chart of the overall implementation.

III.A. Active contour model

The active contour model²⁵ relies on an intrinsic property of image segmentation: For an image formed by two regions, each segmented region should be as homogeneous as possible. Mathematically, this model can be expressed by the following energy function:

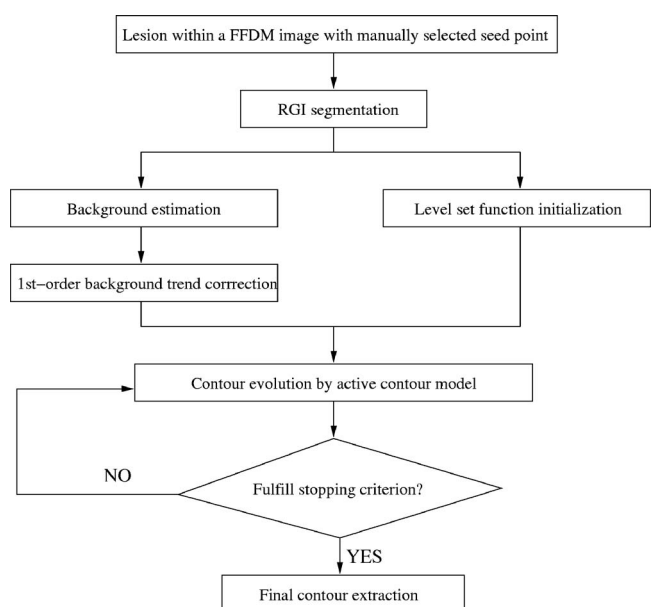


FIG. 2. Schematic diagram of the proposed dual-stage lesion segmentation algorithm.

$$\begin{aligned}
E(c_1, c_2, C) = & \mu \cdot \text{length}(C) \\
& + \lambda_1 \cdot \int_{\text{inside}(C)} |f_0(x, y) - c_1|^2 dx dy \\
& + \lambda_2 \cdot \int_{\text{outside}(C)} |f_0(x, y) - c_2|^2 dx dy, \quad (1)
\end{aligned}$$

where $\mu \geq 0$, $\lambda_1, \lambda_2 > 0$ are fixed weight parameters, C is the evolving contour and $\text{length}(C)$ is a regularizing term that prevents the final contour from converging to a small area due to noise, and c_1 and c_2 are mean values inside and outside of C , respectively. Note that many other active contour models are edge-based as opposed to the gray-level based method used here.

Equation (1) can be represented and solved by level set theory.²⁶ Level set theory, in which the two-dimensional evolving contour C is represented implicitly as the zero level set of a three-dimensional Lipschitz function $\phi(x, y)$, i.e., $C = \{(x, y) \in \Omega : \phi(x, y) = 0\}$, evolves the contour by updating the level set function $\phi(x, y)$ at fixed coordinates through iterations instead of tracking the contour itself. The initial level set function $\phi(x, y)$ is usually defined as the signed distance function:

$$\phi(x, y; t = 0) = \pm d, \quad (2)$$

where d is the distance from (x, y) to $C(t=0)$, where $C(t=0)$ corresponds to the initial contour. The plus (minus) sign is chosen if the point (x, y) is inside (outside) the initial contour $C(t=0)$.

With the evolution of the contour, the level set function ϕ cannot be held as a signed distance function, nor can it be kept smooth. In order to maintain a smooth level set function, and thus ensure numerical stability of evolution, it is necessary to reinitialize the evolving level set function to a signed distance function periodically. However, reinitialization is a computationally consuming procedure as it evolves solving the partial differential equation $\phi_t = \text{sign}(\phi_t)(1 - \|\nabla \phi_t\|)$, where $\nabla \phi_t$ corresponds to the gradient of the level set function. In addition, most reinitializing schemes tend to move the contour to some degree due to numerical errors.²⁷

A signed distance function ϕ , however, has the intrinsic property that $\|\nabla \phi\| = 1$. Thus, it is more natural to incorporate this property into the contour evolution instead of using the independent reinitializing procedure previously described. Thus, we can introduce another regularizing term²⁸ in the active contour model in Eq. (1):

$$\begin{aligned}
E(c_1, c_2, C) = & \mu \cdot \text{length}(C) + \nu \cdot \frac{1}{2} \int_{\Omega} (1 - \|\nabla \phi_t\|)^2 dx dy \\
& + \lambda_1 \cdot \int_{\text{inside}(C)} |f_0(x, y) - c_1|^2 dx dy \\
& + \lambda_2 \cdot \int_{\text{outside}(C)} |f_0(x, y) - c_2|^2 dx dy, \quad (3)
\end{aligned}$$

where ν is a weighted parameter and Ω represents the whole image space.

By replacing C with $\phi(x, y)$ in the energy functional in Eq. (3) and introducing the regularized versions of the Heaviside function $H_\epsilon(\phi) = \frac{1}{2}[1 + 2/\pi \arctan(\phi/\epsilon)]$ along with the corresponding Dirac measure $\delta_\epsilon(\phi) = d/d\phi H_\epsilon(\phi) = \epsilon \cdot [\pi \cdot (\epsilon^2 + \phi^2)]^{-1}$, as given by Chen and Vese in,²⁵ Eq. (3) can be expressed as:

$$\begin{aligned}
E_\epsilon(c_1, c_2, \phi) = & \mu \cdot \int_{\Omega} \delta_\epsilon(\phi(x, y)) \|\nabla \phi(x, y)\| dx dy \\
& + \nu \cdot \frac{1}{2} \int_{\Omega} (1 - \|\nabla \phi(x, y)\|)^2 dx dy \\
& + \lambda_1 \cdot \int_{\Omega} |f_0(x, y) - c_1|^2 H_\epsilon(\phi(x, y)) dx dy \\
& + \lambda_2 \cdot \int_{\Omega} |f_0(x, y) - c_2|^2 [1 - H_\epsilon(\phi(x, y))] dx dy, \quad (4)
\end{aligned}$$

where the first integral controls the length of the contour and the second integral helps to smooth the level set function and thus avoid the need for reinitialization.

By fixing c_1 and c_2 and minimizing E_ϵ in terms of ϕ at each iteration, the associated Euler–Lagrange equation can be derived as:

$$\begin{aligned}
\delta_\epsilon(\phi) \cdot [\mu \cdot \kappa - \lambda_1 \cdot (f_0 - c_1)^2 + \lambda_2 \cdot (f_0 - c_2)^2] \\
+ \nu \cdot \text{div} \left[\left(1 - \frac{1}{\|\nabla \phi\|} \right) \cdot \nabla \phi \right] = 0, \quad (5)
\end{aligned}$$

where

$$\kappa = \text{div} \left(\frac{\nabla \phi}{\|\nabla \phi\|} \right) \quad (6)$$

represents the curvature of the contour C , and also now incorporates the regularizing term from Li et al.²⁸ This derivation, combining the aspect of active contour without edges and level set without reinitialization, is given in the Appendix I. Using the gradient descent method, we can solve ϕ in Eq. (5) iteratively by letting ϕ be a function of iteration t and replace the zero on the right-hand side of Eq. (5) by the time derivative of ϕ . Thus, we obtain a partial differential equation as:

$$\begin{aligned}
\frac{\partial \phi}{\partial t} = & \delta_\epsilon(\phi) \cdot [\mu \cdot \kappa - \lambda_1 \cdot (f_0 - c_1)^2 + \lambda_2 \cdot (f_0 - c_2)^2] \\
& + \nu \cdot \text{div} \left[\left(1 - \frac{1}{\|\nabla \phi\|} \right) \cdot \nabla \phi \right]. \quad (7)
\end{aligned}$$

The time derivative $\partial \phi / \partial t$ was approximated by a forward finite difference:

$$\frac{\partial \phi}{\partial t} = \frac{\phi^{n+1} - \phi^n}{\Delta t}, \quad (8)$$

while considering the numerical stability of the PDE solution, the curvature κ was approximated by a discretizing

scheme that combines both forward and backward finite differences, as suggested in Ref. 29.

$$\kappa = \Delta^x \left(\frac{\Delta_+^x \phi_{i,j}^n}{((\Delta_+^x \phi_{i,j}^n)^2 + (m(\Delta_+^x \phi_{i,j}^n, \Delta_-^x \phi_{i,j}^n)^2)^{1/2}} \right) + \Delta^y \left(\frac{\Delta_+^y \phi_{i,j}^n}{((\Delta_+^y \phi_{i,j}^n)^2 + (m(\Delta_+^y \phi_{i,j}^n, \Delta_-^y \phi_{i,j}^n)^2)^{1/2}} \right), \quad (9)$$

where

$$\Delta_{\mp}^x = \mp (\phi_{i\mp 1,j} - \phi_{i,j}), \quad (10)$$

and similarly for $\Delta_{\mp}^y \phi_{i,j}$;

$$m(a,b) = \left(\frac{\text{sng}(a) + \text{sng}(b)}{2} \right) \min(|a|, |b|). \quad (11)$$

III.B. Contour initialization

The energy function in Eq. (3) depends on the evolving curve C in a complex way. It is not guaranteed to be quadratic or even convex, and one might find a local minimum of the energy function somewhere in the neighborhood of the initial contour. Thus, initializing the contour is a nontrivial task for active contour models. Since lesion sizes vary, it is difficult to find fixed parameters (such as the radius of a circle) with which to initialize the contour for an entire database. Hence, we use our previous RGI-based segmentation method⁷ to estimate the initial boundary of a lesion.

The RGI-based segmentation algorithm⁷ incorporates prior knowledge that mass lesions are roughly compact, and thus, the original image $f(x,y)$ is multiplied with a two-dimensional constraint function $G(x,y; \mu_x, \mu_y, \sigma^2)$ to yield a preprocessed image $h(x,y)$ as

$$h(x,y) = f(x,y) \times G(x,y; \mu_x, \mu_y, \sigma^2), \quad (12)$$

where $G(x,y; \mu_x, \mu_y, \sigma^2)$ is a Gaussian function centered at the manually indicated seed point (μ_x, μ_y) , and with variance σ^2 . The multiplication with the Gaussian function reduces the contribution of structures beyond the lesion, and thus, σ is set to 15 mm to accommodate most mammographic lesion sizes. We have found that the segmentation performance is not strongly dependent on the choice of σ . Larger lesions can also be segmented even though the small deviations around the margin of the lesion are usually not delineated well.

Starting from the given seed point (μ_x, μ_y) , a series of gray level thresholds are then applied to the preprocessed image $h(x,y)$ to yield multiple contours. For each contour, an RGI value is calculated, where RGI is defined as

$$\text{RGI}(\mu_x, \mu_y, C_i) = \frac{\sum_{(x,y) \in C_i} \left(\nabla h(x,y) \cdot \frac{\hat{r}(x,y)}{\|\hat{r}(x,y)\|} \right)}{\sum_{(x,y) \in C_i} \|\nabla h(x,y)\|}, \quad (13)$$

where C_i is the set of points on the i th contour, $\nabla h(x,y)$ is the gradient vector of $h(x,y)$ at point (x,y) , and

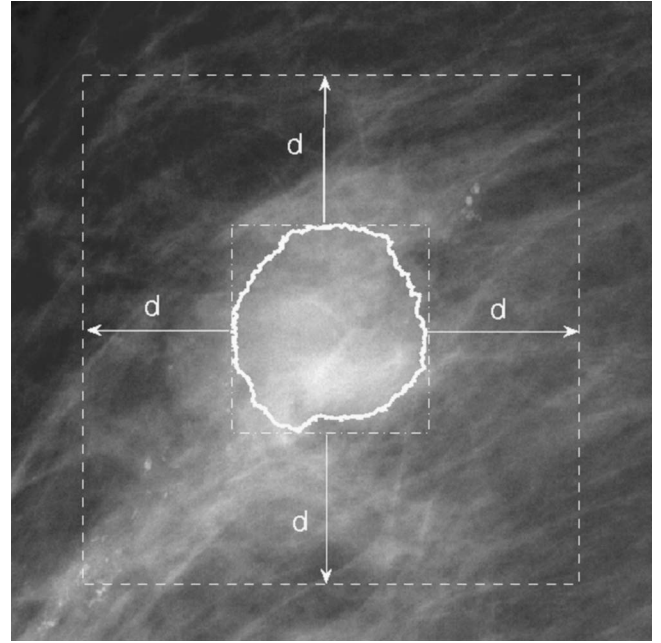


FIG. 3. Illustration of defining the effective background. In this figure, the solid line represents the initial contour obtained by RGI segmentation and the dashed rectangle is the circumscribed rectangle of this initial contour. The effective background is defined as the region inside the dashed rectangle excluding the region within the initial contour. An automatic scheme is employed to determine the best d .

$\hat{r}(x,y)/\|\hat{r}(x,y)\|$ is the normalized radial vector, the direction of which is calculated at position (x,y) with respect to the seed point (μ_x, μ_y) . Of these contours, the one yielding the maximum RGI value is chosen as the contour that best delineates the lesion in the initial step.

RGI represents the average proportion of the gradients in the radially outward direction. The strategy of choosing maximum RGI works well for benign lesions as most have circularlike shapes and smooth margins. However, for malignant lesions, because of irregular shapes and spiculate margins, the resulting contours are usually undergrown. Nevertheless, RGI provides a good initial contour for the following evolution driven by active contour model.

III.C. Background estimation

In the active contour model, contour evolution relies on the competition between the region inside the contour (foreground) and that outside the contour (background). The presence of structure noises, such as lymph nodes, parenchyma, and localization markers, complicates the background in mammograms. RGI segmentation provides not only the initial contour, but also a means of estimating the effective background surrounding the lesion. In our study, the effective background is defined as the set of pixels within a given distance d (pixels) from the circumscribed rectangle of the initial contour, as shown in Fig. 3.

Distance d plays an important role in determining the effective background. On one hand, a large d yields a large

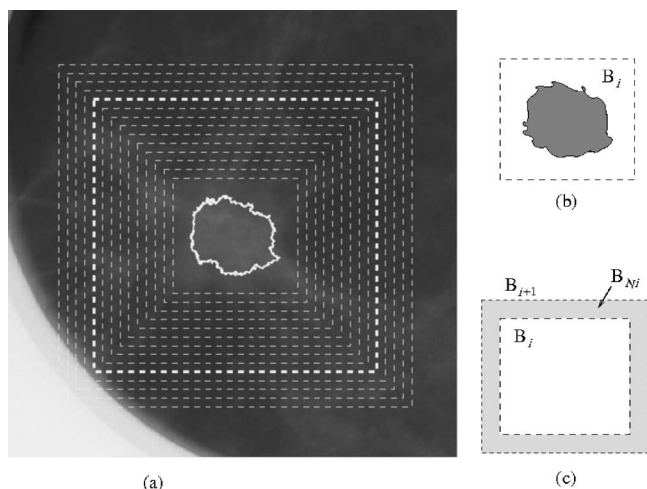


FIG. 4. The illustration of determining the distance d : (a) a mammogram with a series of distances d_i , in which the thick dashed rectangle represents the computer-selected distance d ; (b) B_i : the i th background candidate corresponding to d_i ; and (c) B_{Ni} : the i th net background increase. Background is defined as the set of pixels within a given distance d_i (pixel) from the circumscribed rectangle of the initial contour.

region and thus better statistics on the background. On the other hand, a small d would not be contaminated by nearby structures. In this study, an automatic scheme was developed to determine the best distance d from a series of candidates.

For a series of distances d_i , $i=1, \dots, L$, two series of regions can be determined, as Fig. 4(a) shows. One series of regions are background candidates B_i [Fig. 4(b)], and the other series are net increases of background B_{Ni} [Fig. 4(c)], where $B_{Ni}=B_{i+1}-B_i$, $i=1, \dots, L-1$. Our method is based on the following two principles: With the expansion of background, (1) the mean gray value of B_i , i.e., $\text{mean}(B_i)$, should decrease as more areas with lower gray level are included and (2) the standard deviation of B_{Ni} , i.e., $\text{std}(B_{Ni})$, should not change substantially for relatively smooth background. By monitoring $\text{mean}(B_i)$ and $\text{std}(B_{Ni})$ with increasing d_i , two

potential distance candidates are obtained. One candidate is defined as the distance at which $\text{mean}(B_i)$ reaches a minimum value, and the other candidate is defined as the distance at which $\text{std}(B_{Ni})$ demonstrates the maximum increase, as shown in Fig. 5. At last, the final distance is chosen as the minimum of these two candidates. As for the example in Fig. 4, the distance is automatically determined $d=110$ (pixels).

III.D. Background trend correction

Due to the nonuniformity of the background distribution, some pixels in the background have similar gray values as in the lesion, which hinders the segmentation performance of the active contour model. Thus, a two-dimensional background trend correction was employed prior to segmentation. The background trend is estimated by fitting a two-dimensional surface with a least-squares method to the gradual change in the background pixel values within the extracted background estimation region. Here, we used a first-order polynomial function, i.e., $f(x,y)=a+b \cdot x+c \cdot y$, to describe the two-dimensional surface as higher order polynomial functions will estimate mass lesion instead. Figure 6 demonstrates the significance of the background trend correction when a nonuniform background is present.

III.E. Dynamic stopping criterion

To stop the evolution of a contour, a predetermined threshold is often used. Various metrics can be used to check convergence of evolution, such as the change of level set function ϕ ³⁰ and the change of length of contour.³¹ The contour evolution can also be terminated when the area inside the contour differs from the initial one by a given value.³² In our initial study, we had defined a stopping criterion of relative foreground change (RFC), which is the ratio between the change of foreground and the area of foreground. Comparing with the stopping criterion of change of contour length used in,³¹ RFC has two advantages: (1) RFC is a relative measure and thus is more suitable for lesions with various sizes and

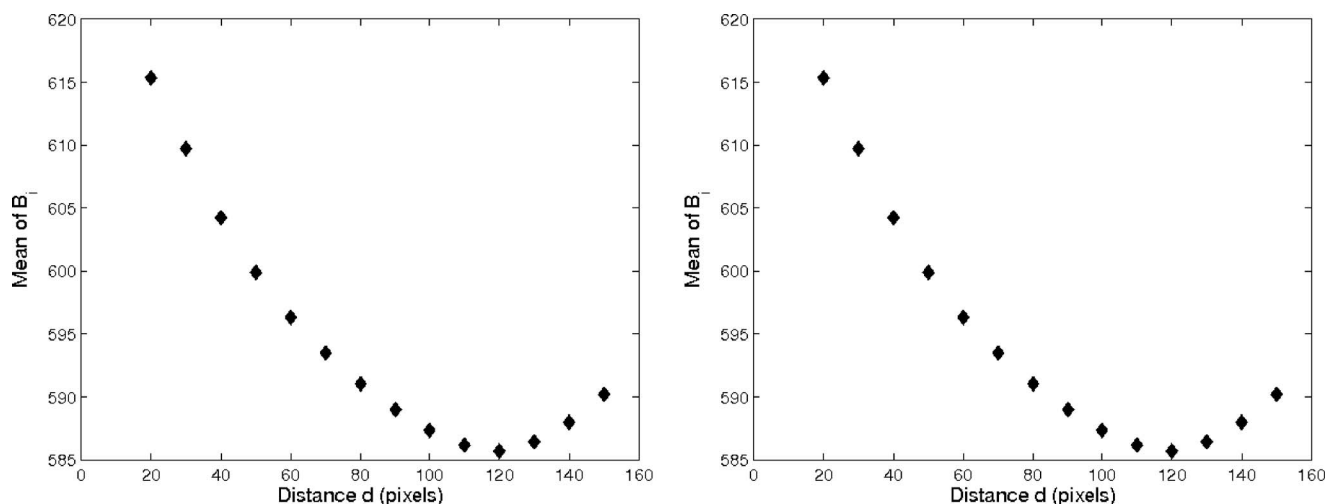


FIG. 5. Left: The trend of mean value of B_i , the i th background candidate with respect to distance d . Right: The trend of standard deviation of B_{Ni} , the i th net background increase with respect to B_i and B_{i+1} .

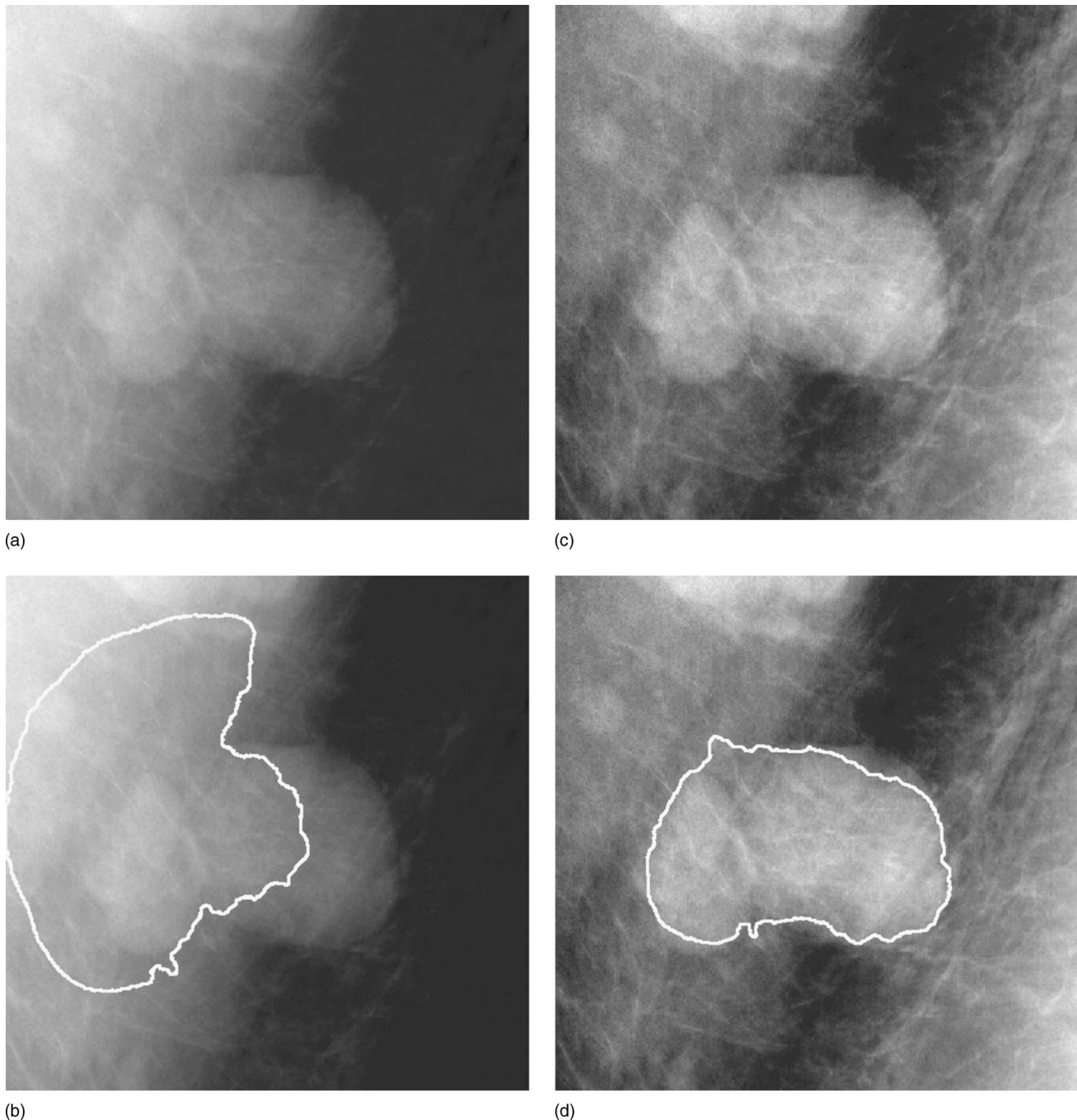


FIG. 6. An example of the effect of background trend correction on segmentation. (a) the original ROI; (b) segmentation result of (a); (c) the processed ROI after background trend correction; and (d) segmentation result of (c).

(2) RFC is more computationally efficient as the acquisition of contour in³¹ brings additional computation. No matter what strategy is used, it is necessary to set some threshold in advance. However, due to varying sizes of lesions, as well as sizes of background obtained from automatic background estimation, it is difficult to find a fixed parameter for controlling convergence.

In our preliminary work,³³ we developed a dynamic method to terminate contour evolution automatically. In that work, as the contour evolves, the mean values of both foreground and background will decrease gradually. As the fore-

ground is generally more homogeneous than the background, the rate of foreground mean change is less than that of background mean change. However, as the evolving contour crosses the lesion margin, the foreground mean will decrease faster than the background mean. Thus, during dynamic contouring, the difference between the rate of foreground mean change and that of background mean change is tracked, and contour evolution is terminated when the decrease of foreground mean value is more rapid than that of the background mean value. This method provides a way to terminate contour evolution free of predefined threshold. However, it ne-

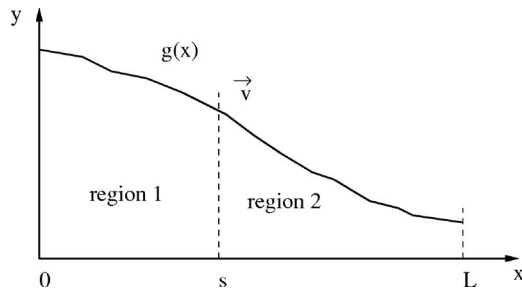


FIG. 7. The illustration of determining the stopping point. $g(x)$ is a decreasing function defined on $[0, L]$ and $s \in [0, L]$ is a moving point with speed of $\vec{v}$.

glects the influence of the sizes of both the foreground and background, and thus ceases contour evolution earlier than expected.

In order to address this problem, we modified the previous method, which we present here in one dimension. As Fig. 7 shows, $g(x)$ is a decreasing function defined on the interval $[0, L]$, and point s is moving within $[0, L]$ at the speed of $\vec{v}$. s also splits $[0, L]$ into two regions. For simplicity, the region $[0, s]$ is named region 1, and $[s, L]$ is region 2. Then, the mean values of regions 1 and 2 are:

$$c_1 = \frac{\int_0^s g(x) dx}{s}, \quad c_2 = \frac{\int_s^L g(x) dx}{L-s}.$$

The slope of c_1 is

$$\begin{aligned} \frac{dc_1}{dt} &= \frac{dc_1}{ds} \cdot \frac{ds}{dt}, \\ &= \frac{d}{ds} \left(\frac{\int_0^s g(x) dx}{s} \right) \cdot \vec{v}, \\ &= \frac{g(s) - c_1}{s} \cdot \vec{v}. \end{aligned}$$

Here, we use the fact that $\vec{v} = ds/dt \cdot \hat{v}$, where $\hat{v}$ is the outward unit vector. Similarly, the slope of c_2 is

$$\frac{dc_2}{dt} = \frac{g(s) - c_2}{L-s} \cdot \vec{v}.$$

Thus, the difference between these two slopes is

$$\Delta v = \frac{dc_1}{dt} - \frac{dc_2}{dt} = \left(\frac{g(s) - c_1}{s} + \frac{g(s) - c_2}{L-s} \right) \cdot \vec{v}. \quad (14)$$

As the discussed above, as s moves within the object, we have $\Delta v > 0$. As s moves across the edge, Δv will become negative. When $\Delta v = 0$, we have $g(s) = s/L \cdot c_2 + (L-s/L) \cdot c_1 > 1/2(c_1 + c_2)$, as in general $L-s > s$ and $c_1 > c_2$. However, if only the speed terms driven by image property in Eq. (7) are considered, the evolution should stop at s_0 such that $g(s_0) = 1/2(c_1 + c_2)$. Because of the influence of sizes, s will stop moving quickly if the criterion in Eq. (14) is used.

In order to eliminate the influence of size, a weighted difference between slope of c_1 and that of c_2 is introduced as

$$\Delta v_w = \frac{s}{L-s} \cdot \frac{dc_1}{dt} - \frac{dc_2}{dt} = \frac{1}{L-s} \cdot [2 \cdot g(s) - c(c_1 + c_2)] \cdot \vec{v}. \quad (15)$$

It can be shown that Δv_w goes to zero at the desired contour s_0 , where $g(s_0) = 1/2(c_1 + c_2)$.

The one-dimensional case, described above, can be extended to a two-dimensional one. During the contour evolution, the weighted difference between the mean slope of foreground and that of background is monitored, and the contour evolution is terminated when the weighted slope difference converges to zero.

III.F. Implementation

In order to calibrate parameters in the proposed segmentation method, ten digitized screen-film mammograms (SFM) with spatial resolution of $100 \mu\text{m} \times 100 \mu\text{m}$ were analyzed. The calibrated segmentation method was then applied to the entire FFDM database for independent performance evaluation.

In our study, we kept both λ_1 and λ_2 in Eq. (7) to one (i.e., $\lambda_1 = \lambda_2 = 1$) since the contribution of the homogeneities of inside and outside the contour should be equally considered. Other parameters in Eq. (7) were chosen as follows: $\epsilon = 1$ and $\Delta t = 0.1$, where ϵ influences the Heavyside function and Δt controls how quickly the level set function changes. Note that μ controls the smoothness of the final contour. However, if one wants to depict the fine details of the object, one should choose a small μ . On the contrary, if one wants to obtain a smoother contour, one should set a large μ . As some of our computer-extracted features, such as spiculation, characterize the fine details of the lesion margin, we chose a fairly small value of μ , i.e., 0.001×1023^2 , which also allows for the use of the 10-bit data. To ensure numerical stability, the coefficient ν must satisfy $\nu \cdot \Delta t < 1/4$,²⁸ so we set $\nu = 2$ in our study. The maximum number of iterations is set to 500.

III.G. Performance evaluation

The performance of the proposed segmentation algorithm was assessed by comparing the computer-delineated contours with the outlines drawn by an expert breast radiologist. Besides visually evaluating the agreement of computer-segmented results with radiologist's manually contoured lesion margins, a quantitative measure was used to evaluate the segmentation performance. For a particular lesion, the area overlap ratio (AOR) between manual segmentation and computer segmentation is defined as

$$\text{AOR} = \frac{\text{Area}(M \cap C)}{\text{Area}(M \cup C)}, \quad (16)$$

where M is the manually segmented contour and C is the computer-segmented contour. AOR ranges from zero to one, with zero in the case of no overlap and 1.0 in the case of a perfect match. For the entire database, a series of AOR

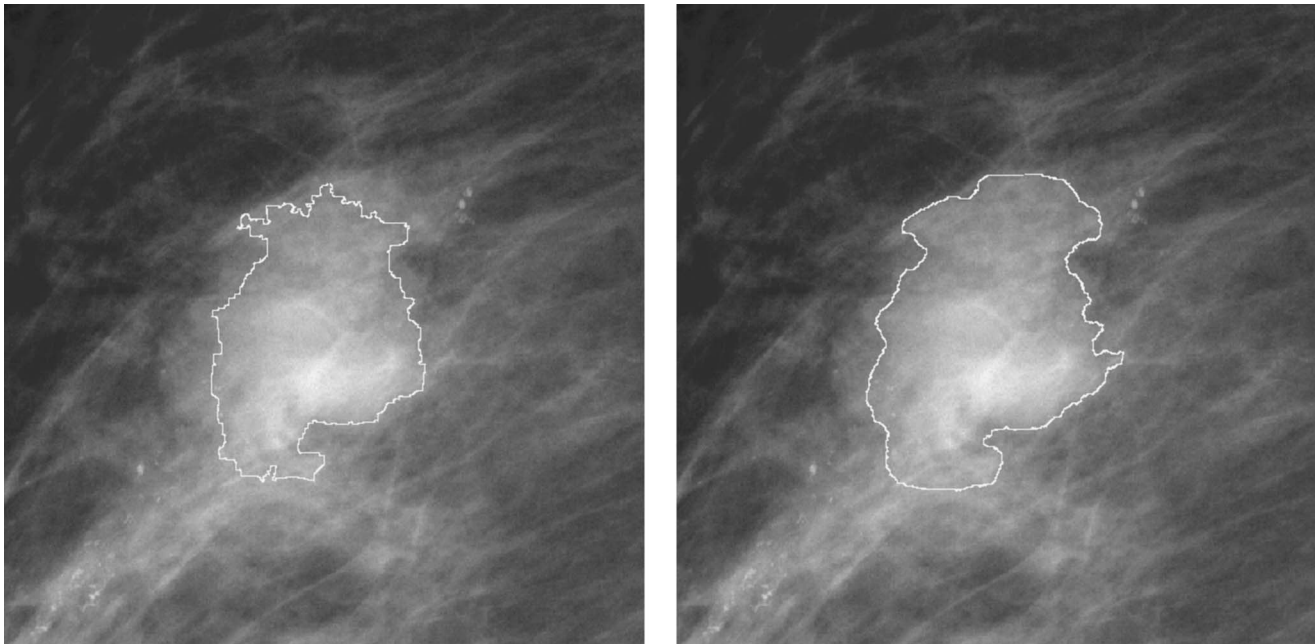


FIG. 8. An example of the effect of level set smoothness to the final segmentation results. Left: Segmentation without level set smoothness; Right: segmentation with level set smoothness.

thresholds were obtained and at each AOR threshold one the percentage of lesions “correctly” segmented was calculated by counting the number of lesions with AOR greater than that threshold.

IV. RESULTS

IV.A. Evaluation of level set smoothness

In our study, a new term $E_S = \int_{\Omega} (1 - \|\nabla \phi_i\|) dx dy$ is added to the original active contour model in,²⁵ thus we initially evaluate the usefulness of this term. Two sets of final contours were extracted from the entire FFDM database; one was obtained with E_S and the other without. The results show that E_S cannot only provide a smoother contour, but also can push the contour closer to the lesion margin with less iterations, yielding a mean number of iterations of 160 compared to the mean number of iterations of 327 without E_S . In the example shown in Fig. 8, the left figure shows the segmentation result without smoothing level set function, which took 500 iterations. Meanwhile, for the result with smoothing level set function shown in the right figure, it took only 248 iterations to converge.

IV.B. Evaluation of dynamic stopping criterion

We investigated our new stopping criterion based on the weighted slope difference between foreground mean and background mean (Δv_w), and compared it to the unweighted slope difference method as well as the relative foreground change (RFC). The RFC thresholds to terminate contour evolution were set as 0.05, and 0.01, respectively. During the evolution, we recorded the contours using these four stopping criteria and obtained the AOR with radiologist’s outlines.

Figure 9 shows plots of the fraction of correctly segmented lesions at various AOR threshold for the four stopping criteria (Δv_w , Δv , $\text{RFC}_{0.05}$, and $\text{RFC}_{0.01}$) on the FFDM databases. For benign images, all the criteria yielded similar segmentation performances since the initial contours, obtained by RGI segmentation, are close to the true lesion margins. However, as RGI segmentation is inferior for malignant lesions, Δv_w does perform better among all the stopping criteria.

Table I summarizes the statistical comparison (Holm t test)³⁴ among these four criteria, given the mean and standard deviation of AOR for each criterion. In terms of AOR, the weighted slope difference method is statistically better than the unweighted slope difference method, and the convergence rate at $\text{RFC}=0.05$ (overall significant level $\alpha^T=0.05$). However, we failed to show a statistically significant difference between the weighted slope difference method and the convergence rate at $\text{RFC}=0.01$. Nevertheless, if the number of iterations is taken into account, the mean number of iterations for weighted slope difference is 156, while it is 280 for $\text{RFC}_{0.01}$. The weighted slope difference is more efficient than $\text{RFC}_{0.01}$.

IV.C. Comparative evaluation of the segmentation method

The segmentation algorithm was compared with our previously reported region-growing⁶ and RGI-based segmentation⁷ methods. Figure 10 shows several examples of lesion segmentations using these three segmentation methods. The result of the proposed method visually demonstrates a better agreement with the radiologist’s outline of the lesion.

Figure 11 shows the fraction of lesions correctly segmented at various overlap threshold levels. At the overlap

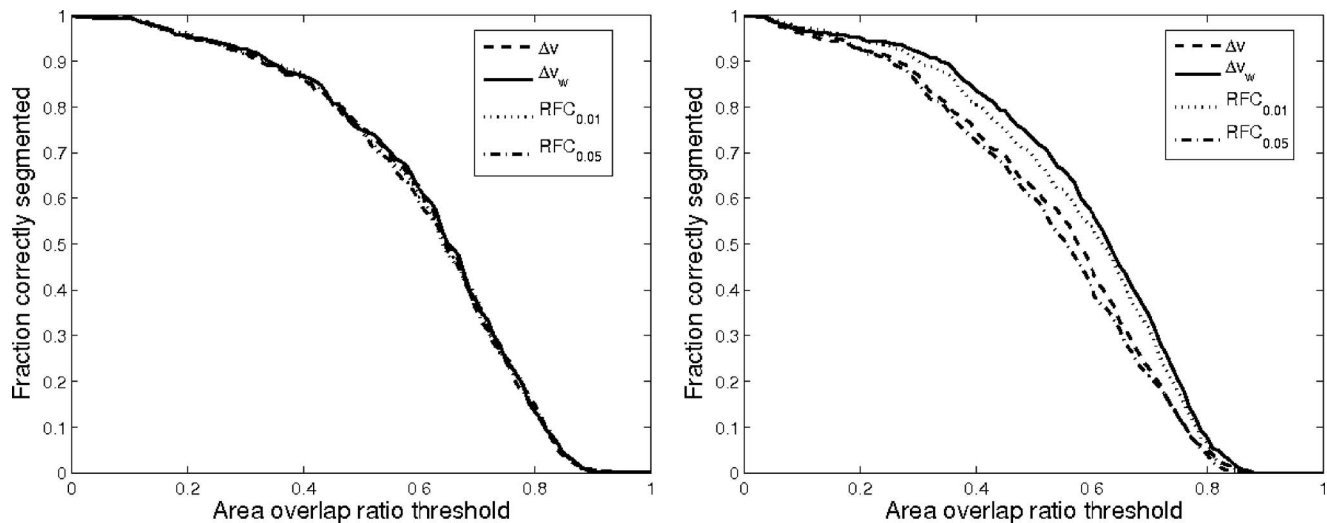


FIG. 9. Segmentation performance of four different stopping criteria in terms of AOR on a clinical FFDM database. In both plots, Δv_w is the weighted slope difference between foreground mean value and background mean value, in which foreground is the area within the evolving contour and background is the area outside contour; Δv is the unweighted slope difference between these two mean values. $\text{RFC}_{0.01}$ stands for a stopping criterion that terminates contour from evolution when the RFC is not greater than 0.01. Similarly, $\text{RFC}_{0.05}$ stops the contour evolution when RFC is not greater than 0.05. Left: evaluated on 327 benign images; Right: Evaluated on 412 malignant images. The results show that the weighted slope difference is statistically superior to unweighted slope difference and convergence rate at $\text{RFC}=0.05$ on malignant images.

threshold of 0.4, for benign lesions, 87% of the images are correctly segmented with the proposed method, while 72% and 81% of the images are correctly segmented by the region-growing and RGI-based methods, respectively. For malignant lesions, 84% of the images are correctly segmented with the proposed method, while 66% and 67% of the images are correctly segmented by region-growing and RGI-based methods, respectively.

Table II gives the statistical comparison (Holm t test)³⁴ for AOR means from the three segmentation methods. The improvement of AOR with the proposed method was found to be statistically significant (overall significant level $\alpha^T=0.05$).

V. DISCUSSION

We developed a dual-stage segmentation method to efficiently segment mass lesions from the parenchymal surround in FFDM images. Our proposed method includes a geometric active contour model, which includes analysis of homogeneities both inside and outside of the evolving contour. The application of RGI-based segmentation to provide initial contour not only improves the computational efficiency, but also provides a method with which to estimate the effective background about the lesion and to suppress unrelated pixel

TABLE I. Statistical comparison of the performance of four stopping criteria in the dual-stage segmentation in terms of AOR, and p -values are given for the comparison of the weighted slope difference with any other stopping criterion. The significant level α_i for the individual paired t test is calculated using Holm's procedure (overall $\alpha^T=0.05$). Same convention as Fig. 9.

		Δv_w	Δv	$\text{RFC}_{0.01}$	$\text{RFC}_{0.05}$
Benign	mean $\pm$ std	0.61 $\pm$ 0.19	0.61 $\pm$ 0.19	0.61 $\pm$ 0.19	0.61 $\pm$ 0.19
	p -value	—	0.856	0.801	0.601
	sig. lev. (α_i)	—	—	—	—
Malignant	mean $\pm$ std	0.59 $\pm$ 0.19	0.53 $\pm$ 0.20	0.57 $\pm$ 0.19	0.52 $\pm$ 0.20
	p -value	—	<0.001	0.192	<0.001
	sig. lev. (α_i)	—	0.05	—	0.025
All	mean $\pm$ std	0.60 $\pm$ 0.19	0.57 $\pm$ 0.20	0.59 $\pm$ 0.19	0.56 $\pm$ 0.20
	p -value	—	0.002	0.25	<0.001
	sig. lev. (α_i)	—	0.05	—	0.025

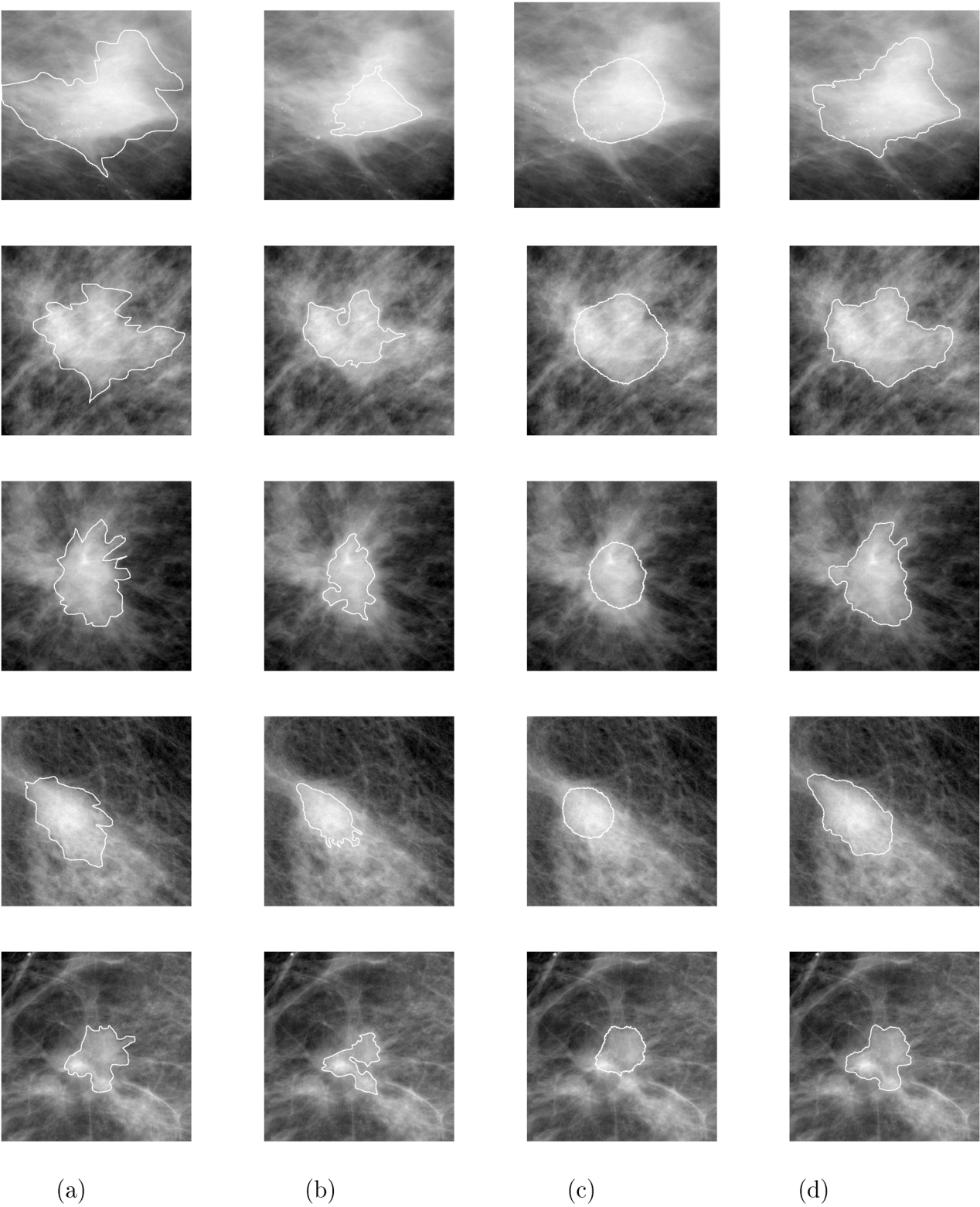


FIG. 10. Segmentation results for five malignant lesion examples: (a) radiologist's outline, (b) region-growing, (c) RGI-based segmentation, and (d) the proposed dual-stage segmentation method.

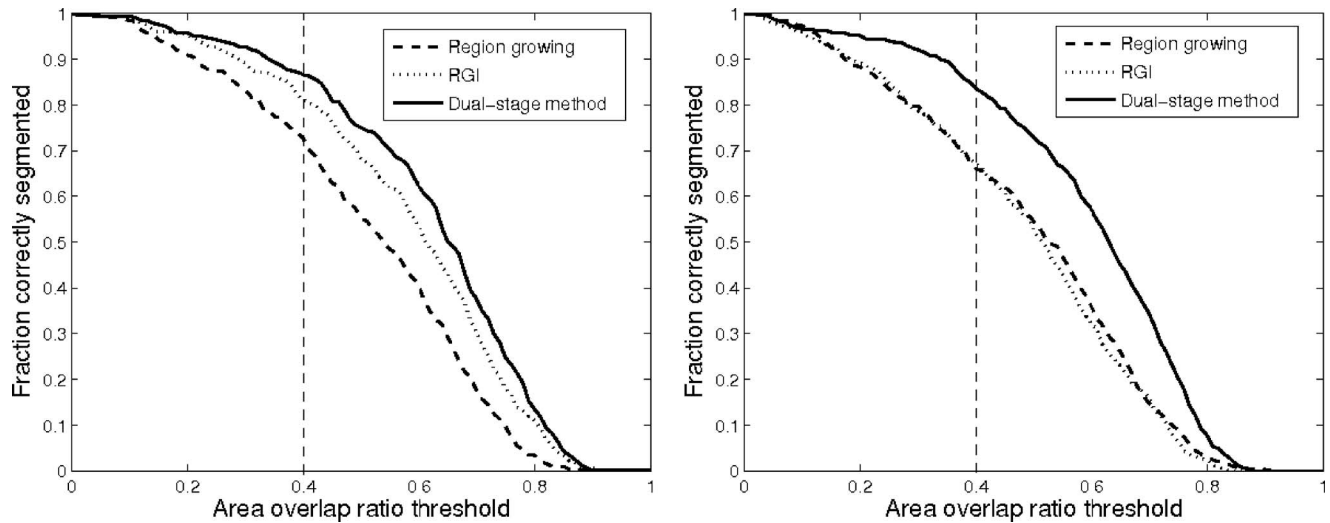


FIG. 11. Performance of three different segmentation methods in terms of AOR on a clinical FFDM database. Left: Evaluated on 327 benign images; Right: Evaluated on 412 malignant images. The results show that the dual-stage segmentation method is statistically superior to both region-growing and RGI-based method.

values. Also, our automatic stopping criterion is lesion-specific and does not rely on fixed iterations.

As the results show, the term E_S in the active contour model plays an important role in effective and efficient segmentation. As $\|\nabla\phi\| > 1$, $\text{div}[(1 - (1/\|\nabla\phi\|))\nabla\phi]$ will evolve the level set function ϕ toward reducing $\|\nabla\phi\|$, thus to smooth ϕ . The larger the gradient magnitude of level set function, the more it will be smoothed. While as $\|\nabla\phi\| < 1$, $\text{div}[(1 - (1/\|\nabla\phi\|))\nabla\phi]$ will evolve the level set function toward increasing $\|\nabla\phi\|$ to maintain the gradient of the level set function to some level. This mechanism ensures the level set function, and thus the final contour, is relatively smooth. Meanwhile, as $\|\nabla\phi\|$ is restricted in magnitude, the foreground has the potential to grow faster.

It should be noticed that the weighted slope difference Δv_w is always non-negative as long as $g(x)$ is a decreasing

function. In the active contour model, if only the speed term driven by image property is considered, the speed of contour can be simplified as:

$$\begin{aligned}\vec{v} &= [(g(s) - c_2)^2 - (g(s) - c_1)^2] \cdot \hat{v}, \\ &= (c_1 - c_2) \cdot [2 \cdot g(s) - (c_1 + c_2)] \cdot \hat{v},\end{aligned}$$

where $\hat{v}$ is the outward unit vector. Inserting $\vec{v}$ into Eq. (15), we have:

$$\Delta v_w = \frac{1}{L - s} \cdot (c_1 - c_2) \cdot [2 \cdot g(s) - (c_1 + c_2)]^2 \cdot \hat{v} \geq 0.$$

If $\vec{v}$ is driven by another image property, such as edge information, this relationship still holds. When $g(s) > 1/2(c_1 + c_2)$, i.e., s is within the object, the contour will move outward to the edge, thus, we have $\Delta v_w \geq 0$. While if $g(s) < 1/2(c_1 + c_2)$, i.e., s is out of object, it will move inward to the edge, we will also have $\Delta v_w \geq 0$. So the weighted slope difference also provides a general mechanism for terminating contour evolution with other active contour models.

In this study, we empirically compared the segmentation performance of the proposed method with our previously reported region growing⁶ and RGI-based⁷ segmentation methods. However, it is impossible for us to perform empirical comparisons between our method and those reviewed in the introduction section, as we do not have codes of those methods. Timp's method¹⁰ uses polar coordinate and restricts the mass sizes within a certain range; thus one would expect their method to work better for lesions with circularlike margins. However, for lesions with irregular shapes or very large sizes, their method may have difficulty. Our dual stage segmentation method is able to handle this situation by further evolving the contour via the active contour model. For the fuzzy-set-based methods developed by Guliato *et al.*,¹¹ both of them need to present some thresholds such as the gray-level threshold in the first method and the maximum allowed

TABLE II. Statistical comparison of the three lesion segmentation algorithms. Performance is given by average AOR, and p -values are given for the comparison of the dual-stage segmentation with the previous region-growing and RGI-based method. The significant level α_i for the individual paired t test is calculated using Holm's procedure (overall $\alpha^T=0.05$).

		Dual-stage segmentation	RGI	Region- growing
Benign	mean $\pm$ std	0.61 $\pm$ 0.19	0.58 $\pm$ 0.19	0.51 $\pm$ 0.20
	p -value	—	0.01	<0.001
	seg. lev. (α_i)	—	0.05	0.025
Malignant	mean $\pm$ std	0.59 $\pm$ 0.19	0.48 $\pm$ 0.20	0.49 $\pm$ 0.20
	p -value	—	<0.001	<0.001
	seg. lev. (α_i)	—	0.025	0.05
All	mean $\pm$ std	0.60 $\pm$ 0.19	0.52 $\pm$ 0.20	0.50 $\pm$ 0.20
	p -value	—	<0.001	<0.001
	seg. lev. (α_i)	—	0.05	0.025

difference between the value of the pixel being analyzed and the mean of the subregion in the second method, which prevents these methods from being applied in a large database. Their two thresholds were manually selected case by case in their evaluation using a database with 47 mammograms. On the other hand, our method is flexible in that no threshold needs to be set in advance.

In our preliminary study,³⁵ we compared two radiologists' outlines with a digitized screen-film mammograms (SFM) database, which consisted of 29 benign (51 mammograms) and 55 malignant (96 mammograms) lesions. At an overlap threshold of 0.4, 96.6% of lesion images were correctly segmented by one radiologist in comparison with the other. This result indicates that the radiologists highly agreed on the lesion margins for SFM. We could expect that the radiologists would also agree on the lesion margins for FFDM as the manufacturer has preprocessed the FFDM images to make them appear to the radiologist as traditional-looking SFM mammographs.

When we developed the proposed segmentation algorithm, the FFDM database was being constructed, so our method was initially calibrated and tested with the SFM database.³³ After building the FFDM database, we randomly picked three groups of FFDM images, each of which consisted of five benign and five malignant images, and evaluated the segmentation performance using the proposed method calibrated with SFM images. The results were similar with what we had obtained with SFM images. Thus, we believe that the parameters obtained by SFM also work with FFDM images, which was subsequently validated by the independent evaluation with the entire FFDM database.

Our results could be partially explained by the preprocessing of FFDM images, which is performed by the manufacturers. After preprocessing, the gray-level range and contrast of FFDM images become similar to those of SFM images, which ensures the possibility of applying parameters from SFM images to FFDM images as gray-level range and contrast are two key components used in our proposed lesion segmentation method. Our results also show the robustness of the proposed method as it mainly uses the global information of images.

VI. CONCLUSION

In this article, we present a new lesion segmentation method based on a geometric active contour model, which includes an initial RGI segmentation, background estimation, background trend correction, and a dynamic stopping criterion. Evaluation with a large number of FFDM images has shown that the proposed method is statistically superior to our previous region-growing and RGI-based algorithms in terms of overlap ratios obtained in comparison with experts' manual outlines. At an overlap threshold of 0.4, 85% of the images are correctly segmented by the proposed method, while only 69% and 73% of the images are correctly segmented by our previous region-growing and RGI-based methods, respectively.

ACKNOWLEDGMENTS

This work was supported in part by the US Army Breast Cancer Research Program (BCRP) Predoctoral Traineeship Award (W81XWH-06-1-726), by United States Public Health Service (USPHS) (Grant No. CA89452), and by a grant from the U.S. Army Medical Research and Materiel Command (Grant No. DAMD 98-1209), and by Cancer Center Support (Grant No. 5-P30CA14599). M. L. Giger is a shareholder in R2 Technology, Inc. (Sunnyvale, CA), a Hologic Company. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actually or potential significant financial interest which would reasonably appear to be directly and significantly affected by the research activities.

APPENDIX

In this part, we provide the details of the derivation from energy function (4) to the associated Euler–Lagrange equation (5). For convenience, we restate Eq. (4) here as

$$E_{\epsilon}(c_1, c_2, \phi) = \int_{\Omega} \left[\mu \cdot \delta_{\epsilon}(\phi(x, y)) \|\nabla \phi(x, y)\| + \frac{v}{2} \cdot (1 - \|\nabla \phi(x, y)\|)^2 + \lambda_1 \cdot |f_0(x, y) - c_1|^2 H_{\epsilon}(\phi(x, y)) + \lambda_2 \cdot |f_0(x, y) - c_2|^2 (1 - H_{\epsilon}(\phi(x, y))) \right] dx dy. \quad (A1)$$

We define $F(\phi, \nabla \phi, x, y)$ as

$$F(\phi, \nabla \phi, x, y) = \mu \delta_{\epsilon}(\phi) \|\nabla \phi\| + \frac{v}{2} (1 - \|\nabla \phi\|)^2 + \lambda_1 |f_0 - c_1|^2 H_{\epsilon}(\phi) + \lambda_2 |f_0 - c_2|^2 (1 - H_{\epsilon}(0)). \quad (A2)$$

For simplicity, we have omitted the independent variables (x, y) of ϕ and f_0 . According to calculus of variations, the scalar function $\phi(x, y)$ that minimizes $E_{\epsilon}(c_1, c_2, \phi)$ solves the PDE:

$$\frac{d}{dx} \left(\frac{\partial F}{\partial \phi_x} \right) + \frac{d}{dy} \left(\frac{\partial F}{\partial \phi_y} \right) - \frac{\partial F}{\partial \phi} = 0. \quad (A3)$$

Taking the partial derivative of F with respect to ϕ_x , ϕ_y , and ϕ , respectively, we have:

$$\frac{\partial F}{\partial \phi_x} = \mu \delta_{\epsilon}(\phi) \frac{\phi_x}{\|\nabla \phi\|} + v \left(\phi_x - \frac{\phi_x}{\|\nabla \phi\|} \right),$$

$$\frac{\partial F}{\partial \phi_y} = \mu \delta_{\epsilon}(\phi) \frac{\phi_y}{\|\nabla \phi\|} + v \left(\phi_y - \frac{\phi_y}{\|\nabla \phi\|} \right),$$

$$\frac{\partial F}{\partial \phi} = \mu \|\nabla \phi\| \delta'_\epsilon(\phi) + [\lambda_1(f_0 - c_1)^2 - \lambda_2(f_0 - c_2)^2] \delta_\epsilon(\phi), \quad (\text{A4})$$

where $\delta'_\epsilon = d\delta/d\phi$ and we use the relation $\|\nabla \phi\| = \sqrt{\phi_x^2 + \phi_y^2}$.

The partial derivative of $\partial F/\partial \phi_x$ with respect to x is

$$\begin{aligned} \frac{d}{dx} \left(\frac{\partial F}{\partial \phi_x} \right) &= \mu \delta'_\epsilon(\phi) \frac{\phi_x^2}{\|\nabla \phi\|} + \mu \delta_\epsilon(\phi) \frac{d}{dx} \left(\frac{\phi_x}{\|\nabla \phi\|} \right) \\ &\quad + v \frac{d}{dx} \left[\phi_x - \frac{\phi_x}{\|\nabla \phi\|} \right]. \end{aligned} \quad (\text{A5})$$

Similarly, we have

$$\begin{aligned} \frac{d}{dy} \left(\frac{\partial F}{\partial \phi_y} \right) &= \mu \delta'_\epsilon(\phi) \frac{\phi_y^2}{\|\nabla \phi\|} + \mu \delta_\epsilon(\phi) \frac{d}{dy} \left(\frac{\phi_y}{\|\nabla \phi\|} \right) \\ &\quad + v \frac{d}{dy} \left[\phi_y - \frac{\phi_y}{\|\nabla \phi\|} \right]. \end{aligned} \quad (\text{A6})$$

Inserting Eqs. (A4)–(A6) back to (A3), we obtain

$$\begin{aligned} 0 &= \mu \delta'_\epsilon(\phi) \left[\frac{\phi_x^2}{\|\nabla \phi\|} + \frac{\phi_y^2}{\|\nabla \phi\|} - \|\nabla \phi\| \right] + \mu \delta_\epsilon(\phi) \\ &\quad \times \left[\frac{d}{dx} \left(\frac{\phi_x}{\|\nabla \phi\|} \right) + \frac{d}{dy} \left(\frac{\phi_y}{\|\nabla \phi\|} \right) \right] + v \left\{ \frac{d}{dx} \left[\phi_x - \frac{\phi_x}{\|\nabla \phi\|} \right] \right. \\ &\quad \left. + \frac{d}{dy} \left[\phi_y - \frac{\phi_y}{\|\nabla \phi\|} \right] \right\} - \delta_\epsilon(\phi) [\lambda_1(f_0 - c_1)^2 \\ &\quad - \lambda_2(f_0 - c_2)^2]. \end{aligned} \quad (\text{A7})$$

By noticing that

$$\begin{aligned} \frac{\phi_x^2}{\|\nabla \phi\|} + \frac{\phi_y^2}{\|\nabla \phi\|} &= \|\nabla \phi\|, \\ \frac{d}{dx} \left(\frac{\phi_x}{\|\nabla \phi\|} \right) + \frac{d}{dy} \left(\frac{\phi_y}{\|\nabla \phi\|} \right) &= \operatorname{div} \left(\frac{\nabla \phi}{\|\nabla \phi\|} \right), \end{aligned}$$

and

$$\begin{aligned} \frac{d}{dx} \left[\phi_x - \frac{\phi_x}{\|\nabla \phi\|} \right] + \frac{d}{dy} \left[\phi_y - \frac{\phi_y}{\|\nabla \phi\|} \right] \\ = \operatorname{div} \left[\left(1 - \frac{1}{\|\nabla \phi\|} \right) \nabla \phi \right], \end{aligned}$$

we finally obtain the compact form of Eq. (A7) as

$$\begin{aligned} 0 &= \delta_\epsilon(\phi) \left[\mu \cdot \operatorname{div} \left(\frac{\nabla \phi}{\|\nabla \phi\|} \right) - \lambda_0(f_0 - c_1)^2 + \lambda_2(f_0 - c_2)^2 \right] \\ &\quad + v \cdot \operatorname{div} \left[\left(1 - \frac{1}{\|\nabla \phi\|} \right) \nabla \phi \right]. \end{aligned} \quad (\text{A8})$$

^{a)} Author to whom correspondence should be addressed. Electronic mail: yading@uchicago.edu

¹ A. J. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, and M. J. Thun, "Cancer statistics, 2007," *Ca-Cancer J. Clin.* **57**, 43–66 (2007).

² G. M. Newstead, "Role of MR in breast imaging," In *RSNA Categorical Course in Breast Imaging* (Chicago, 1999), pp. 287–293.

³ S. G. Orel and M. D. Schnall, "MR imaging of the breast for the detec-

tion, diagnosis, and staging of breast cancer," *Radiology* **220**, 13–30 (2001).

⁴ A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney, "Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions," *Radiology* **196**, 123–134 (1995).

⁵ K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, "Computerized diagnosis of breast lesions on ultrasound," *Med. Phys.* **29**, 157–164 (2002).

⁶ Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick, and P. Lu, "Analysis of speculation in the computerized classification of mammographic masses," *Med. Phys.* **22**, 1569–1579 (1995).

⁷ M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Trans. Med. Imaging* **17**, 510–517 (1998).

⁸ 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).

⁹ H. D. Li, M. Kallergi, L. P. Clarke, and V. K. Jain, "Markov random field for tumor detection in digital mammography," *IEEE Trans. Med. Imaging* **14**, 565–576 (1995).

¹⁰ S. Timp and N. Karssemeijer, "A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography," *Med. Phys.* **31**, 958–971 (2004).

¹¹ D. Guliato, R. M. Rangayyan, W. A. Carnielli, J. A. Zuffo, and J. E. L. Desautels, "Segmentation of breast tumors in mammograms using fuzzy sets," *J. Electron. Imaging* **12**, 369–378 (2003).

¹² H. Li, Y. Wang, K. J. R. Liu, S. B. Lo, and M. T. Freedman, "Computerized radiographic mass detection—Part 1: Lesion site selection by morphological enhancement and contextual segmentation," *IEEE Trans. Med. Imaging* **20**, 289–301 (2001).

¹³ M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.* **1**, 321–331 (1987).

¹⁴ T. Cootes, A. Hill, C. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image Vis. Comput.* **12**, 355–366 (1994).

¹⁵ M. S. Atkins and B. Mackiewicz, "Fully automatic segmentation of the brain in MRI," *IEEE Trans. Med. Imaging* **17**, 98–107 (1998).

¹⁶ A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum, "A geometric snake model for segmentation of medical imagery," *IEEE Trans. Med. Imaging* **16**, 199–209 (1997).

¹⁷ C. Xu, D. L. Pham, and J. L. Prince, "Medical image segmentation using deformable models," in *Handbook of Medical Imaging: Medical Imaging Processing and Analysis*, edited by M. Sonka and M. J. Fitzpatrick (2000), vol. 2 of *Proc. SPIE*, pp. 129–174.

¹⁸ J. Yang, L. H. Staib, and J. S. Duncan, "Neighbor-constrained segmentation with level set based 3D deformable models," *IEEE Trans. Med. Imaging* **23**, 940–948 (2004).

¹⁹ F. Liu, B. Zhao, and P. K. Kijewski, "Liver segmentation for CT images using GVF snake," *Med. Phys.* **32**, 3699–3706 (2005).

²⁰ G. M. Brake and N. Karssemeijer, "Segmentation of suspicious densities in digital mammograms," *Med. Phys.* **28**, 259–266 (2001).

²¹ B. Sahiner, N. Petrick, H. P. Chan, L. M. Hadjiiski, C. Paramagul, M. A. Helvie, and M. N. Gurcan, "Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on classification," *IEEE Trans. Med. Imaging* **20**, 1275–1284 (2001).

²² V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.* **22**, 61–79 (1997).

²³ R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 158–175 (1995).

²⁴ S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulation," *J. Comput. Phys.* **79**, 12–49 (1988).

²⁵ T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.* **10**, 266–277 (2001).

²⁶ J. A. Sethian, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science* (Cambridge University Press, Cambridge, 1999).

²⁷ S. J. Osher and R. P. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces* (Springer, New York, 2002).

²⁸ C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without reinitialization: A new variational formulation," In *Proc. 2005 IEEE CVPR*

- (San Diego, 2005), pp. 1:430–436.
- ²⁹L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *J. Phys. D* **60**, 259–268 (1992).
- ³⁰R. Kimmel, *Numerical Geometry of Images: Theory, Algorithms, and Applications* (Springer, New York, 2003).
- ³¹K. N. Chaudhury and K. R. Ramakrishnan, “Stability and convergence of the level set method in computer vision,” *Pattern Recogn. Lett.* **28**, 884–893 (2007).
- ³²Y. Yu and J. A. Molloy, “Segmentation of the prostate from suprapubic ultrasound images,” *Med. Phys.* **31**, 3474–3484 (2004).
- ³³Y. Yuan, M. L. Giger, K. Suzuki, H. Li, and A. R. Jamieson, “A two-stage method for lesion segmentation on digital mammograms,” in *Medical Imaging: Image processing*, edited by J. M. Reinhardt and J. P. W. Pluim (2006), vol. 6144 of *Proc. SPIE*, pp. 61443W:1–5.
- ³⁴S. A. Glantz, *Primer of Biostatistics* (McGraw-Hill, New York, 2002).
- ³⁵H. Li, M. L. Giger, Y. Yuan, L. Lan, K. Suzuki, A. R. Jamieson, and C. Sennett, “Comparison of computerized image analysis for digitized screen film mammograms and full field digital mammography images,” in *Digital Mammography*, edited by S. M. Astley, M. Brady, C. Rose, and R. Zwiggelaar (2006), vol. LNCS4046 of *IWDM 2006*, pp. 569–575.

Appendix B

Comparison of image segmentation methods on classification performance of FFDM CAD

Section I: Segmentation methods

The classification performance of two segmentation methods were investigated and compared in this study.

1) Region growing method^[1]

In the region growing method, a two-dimensional background correction and histogram equalization are first applied to the original image data. Gray level thresholding is subsequently performed on the processed image to yield contours. By monitoring the size and shape of the evolving contour with each incremental threshold step, the final lesion contour (i.e., lesion margin), corresponding to an abrupt transition from small size to larger size, and from high circularity to low circularity, is automatically selected.

2) Dual-stage segmentation^[2]

In the dual-stage segmentation method, a radial gradient index (RGI) based segmentation is used to yield an initial contour close to the lesion boundary location in a computationally efficient manner. This initial segmentation also provides a base to identify the effective circumstance of the lesion via an automatic background estimation method. Then a region-based active contour model is utilized to evolve the contour to the lesion boundary. This active contour model relies on an intrinsic property of image segmentation that each segmented region should be as homogeneous and possible for an image formed by two regions. Instead of empirically determined criteria such as fixed iteration times, a dynamic stopping criterion is implemented to terminate the contour evolution when it reaches the lesion boundary.

Section II: Segmentation results

The performance of segmentation was initially evaluated by comparing the computer segmentation with manual outlines delineated by an expert breast radiologist. Figure 1 shows a malignant example of lesion segmentation, which indicates that the dual-stage

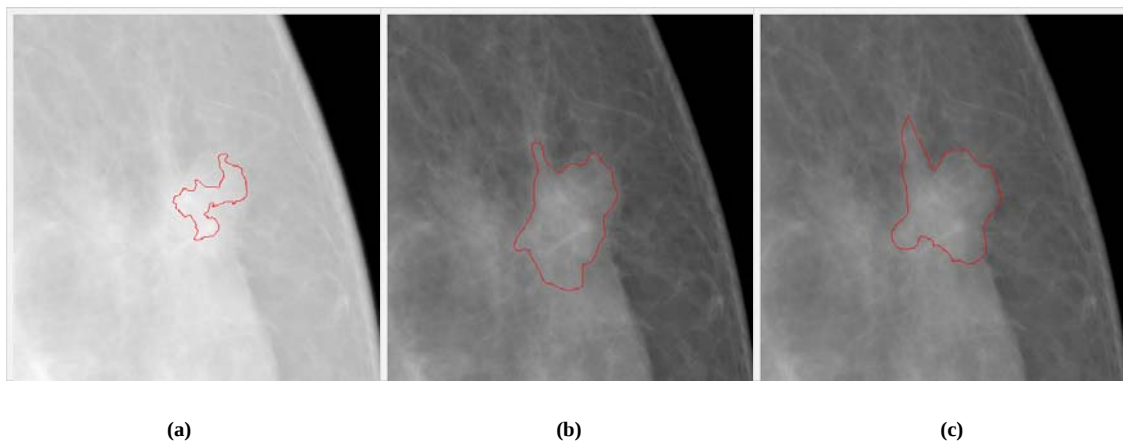


Fig. 1 A malignant example of lesion segmentation via different methods: (a) region growing, (b) dual-stage method, and (c) radiologist's outline

Comparison of image segmentation methods on classification performance of FFDM CAD

method visually demonstrates a much closer agreement with the radiologist's outline of the lesion.

Section III: Classification performance

By stepwise feature selection with Wilks lambda criterion, three features, including normalized radial gradient (NRG) of ROI, NRG of lesion and gradient texture, were selected from 15 features^[1] being extracted from the lesions segmented by the dual-stage method, which yielded an AUC of 0.78. Margin sharpness, gradient texture, two lesion margin spiculation features and the average gray level were selected for region growing method, yielding an AUC of 0.72. The difference is statistically significant. Figure 2 shows the ROC curves resulting from evaluation of these two groups of features.

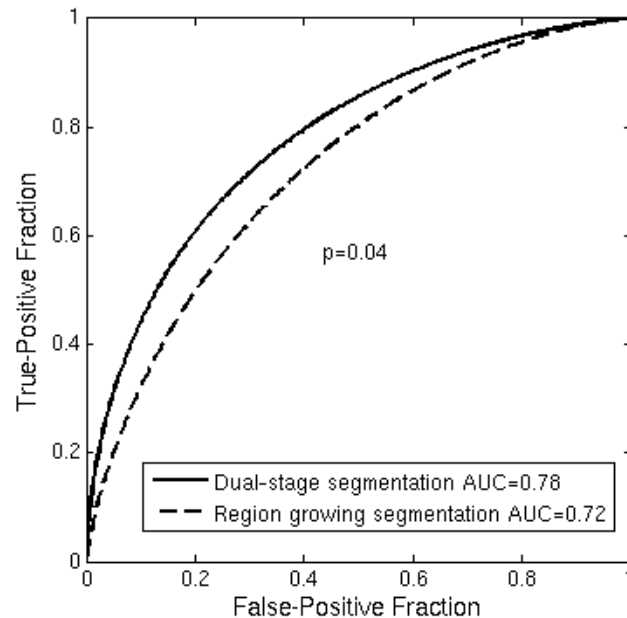


Fig.2 ROC curves of leave-one-out by lesion for the feature subset from the dual-stage segmentation method (solid line) and the feature subset from the region growing method (dash line).

Section IV: References

- [1] Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick and P. Lu, "Analysis of spiculation in the computerized classification of mammographic masses," *Med. Phys.* 22, pp. 1569-1579, 1995.
- [2] Y. Yuan, M. L. Giger, H. Li, K. Suzuki, and C. Sennett, "A dual-stage method for lesion segmentation on digital mammograms," *Med. Phys.* 34, pp. 4180-4193, 2007.
- [3] C. E. Metz, B. A. Herman, and J. Shen, "Maximum likelihood estimation of receiver operating characteristic ROC curves from continuously-distributed data," *Stat. Med.* 17, pp. 1033-1053, 1998.

Correlative feature analysis on FFDM

Yading Yuan,^{a)} Maryellen L. Giger, Hui Li, and Charlene Sennett

Department of Radiology, Committee on Medical Physics, The University of Chicago, 5841 South Maryland Avenue, MC 2026 Chicago, Illinois 60637

(Received 18 April 2008; revised 9 September 2008; accepted for publication 1 October 2008; published 13 November 2008)

Identifying the corresponding images of a lesion in different views is an essential step in improving the diagnostic ability of both radiologists and computer-aided diagnosis (CAD) systems. Because of the nonrigidity of the breasts and the 2D projective property of mammograms, this task is not trivial. In this pilot study, we present a computerized framework that differentiates between corresponding images of the same lesion in different views and noncorresponding images, i.e., images of different lesions. A dual-stage segmentation method, which employs an initial radial gradient index (RGI) based segmentation and an active contour model, is applied to extract mass lesions from the surrounding parenchyma. Then various lesion features are automatically extracted from each of the two views of each lesion to quantify the characteristics of density, size, texture and the neighborhood of the lesion, as well as its distance to the nipple. A two-step scheme is employed to estimate the probability that the two lesion images from different mammographic views are of the same physical lesion. In the first step, a correspondence metric for each pairwise feature is estimated by a Bayesian artificial neural network (BANN). Then, these pairwise correspondence metrics are combined using another BANN to yield an overall probability of correspondence. Receiver operating characteristic (ROC) analysis was used to evaluate the performance of the individual features and the selected feature subset in the task of distinguishing corresponding pairs from noncorresponding pairs. Using a FFDM database with 123 corresponding image pairs and 82 noncorresponding pairs, the *distance* feature yielded an area under the ROC curve (AUC) of 0.81 ± 0.02 with leave-one-out (by physical lesion) evaluation, and the feature metric subset, which included *distance*, *gradient texture*, and ROI-based *correlation*, yielded an AUC of 0.87 ± 0.02 . The improvement by using multiple feature metrics was statistically significant compared to single feature performance. © 2008 American Association of Physicists in Medicine. [DOI: [10.1118/1.3005641](https://doi.org/10.1118/1.3005641)]

Key words: computer-aided diagnosis, full-field digital mammography, correlative feature analysis, lesion segmentation, feature selection

I. INTRODUCTION

Breast cancer is a leading cause of mortality in American women, with an estimated 182 460 new cancer cases and 40 480 deaths in the United States in 2008.¹ Nevertheless, between the years 1990 to 2003, there has been a steady decrease in the annual death rate due to female breast cancer.² This decrease largely reflects improvements in early detection and treatment. Currently, x-ray mammography is the most prevalent imaging procedure for the early detection of breast cancer.³

During mammographic screening, multiple projection views, such as craniocaudal (CC), mediolateral oblique (MLO), and mediolateral (ML) views, are usually obtained. Researchers have analyzed images from these different views to increase the performance of computer-aided detection. Paquerault *et al.*⁴ developed a two-view matching method that computes a correspondence score for each possible region pair in CC and MLO views, and merged it with a single-view detection score to improve lesion detectability. To reduce the number of false positive detections, Zheng *et al.*⁵ identified a matching strip of interest on the ipsilateral view based on the projected distance to the nipple and searched for a region within the strip and paired it with the

original region. Engeland *et al.*⁶ built a cascaded multiple-classifier system, in which the last stage computes suspiciousness of an initially detected region conditional on the existence and similarity of a linked candidate region in the other view.

It has also been well recognized that multiple views can improve the diagnosis of breast cancer in the computerized analysis of mammograms,^{7–10} since different projections provide complementary information about the same physical lesion. To merge information from images of different views, an essential step is to verify that these images actually represent the same physical lesion.

We present a dual-stage correlative feature analysis (CFA) method to address the task of classifying corresponding images of lesions as seen in different views. In this method, mass lesions are initially segmented automatically from the surrounding parenchyma. Then various features, including distance, morphological, and textural features, are extracted from the mass lesion on each of the two views. For a given pair of images, one from each view, each pair of computer-extracted features is merged through a Bayesian artificial neural network (BANN) to obtain correspondence metrics. The correspondence metrics are then merged with a second

BANN to yield an estimate of the probability that the two lesions on different mammographic images are of the same physical lesion. This CFA method is different from conventional image registration methods in the following two aspects: (1) The task of image registration is to align two images known to represent the same object, while CFA is to assess the probability that the given two images represent the same object. (2) The key point of image registration is to determine a geometrical transformation that minimizes some cost functions defined by intensities, contours, and mutual information,^{11–13} in which various geometrical landmarks, such as control points and inherent image landmarks (nipple, curves, regions and breast skin),^{14–16} are identified and matched. The proposed CFA technique is feature based, which is motivated by the studies on fusion of two-view information for computer-aided detection,^{4–6} as well as our prior research on the task of automated classification of breast lesions, i.e., in the determination of benign and malignant breast lesions based on computer-extracted features.^{17,18}

Differing from the studies on computer-aided detection, however, our purpose is to identify the corresponding lesions from different views, and ultimately improve the performance of computer-aided diagnosis. Therefore, the noncorresponding pairs in our study will be lesion-lesion pairs, as compared to the lesion-parenchyma or parenchyma-parenchyma noncorresponding pairs in lesion detection task. In a correspondence study between two mammographic views for the lesion diagnosis task, Gupta *et al.*¹⁹ investigated the correlation between corresponding texture features from two different views, and suggested that one could include features from an additional view only if they were less correlated with features from the existing view, i.e., providing more complementary information. Our study, however, does not discuss methods to merge information from different views, but rather focuses on classifying the correspondence between lesions instead.

II. MATERIALS AND METHODS

The main aspect of the proposed correlative feature analysis includes automatic lesion segmentation, computerized feature extraction, feature selection, and an estimation of the probability that two given images represent the same physical lesion. Figure 1 shows the schematic diagram of the proposed method.

II.A. Database

The full-field digital mammography (FFDM) database in our study consists of 135 biopsyproven mass lesions acquired at the University of Chicago Hospitals, in which lesions were collected under an approved institutional review board (IRB) protocol. Of the 135 lesions, 67 are benign with 123 mammograms and 68 are malignant with 139 mammograms. All the images were obtained from GE Senographe 2000D systems (GE Medical Systems Milwaukee, WI) with a spatial resolution of $100 \times 100 \mu\text{m}^2$ in image plane. The masses were identified and outlined by an expert breast radiologist based on visual criterion and biopsy-proven reports.

Based on the correspondence of lesions identified by the radiologist, we constructed 123 corresponding pairs and 82 noncorresponding pairs. Each pair consists of a CC view and a ML view. Figure 2 shows an example case with multiple lesions seen on mammograms in CC and ML views. Considering the most realistic scenario of lesion mismatch in clinical practice, the noncorresponding pairs were constructed from cases of the same patients but different physical lesions. Since in our database only 28 patients had two or more lesions in the same breast, the noncorresponding dataset, which includes all possible lesion combinations from the different views, is limited. Table I lists the detailed information regarding the corresponding and noncorresponding datasets.

II.B. Lesion segmentation

In our study, a dual-stage method,²⁰ on which we have already reported, was employed to automatically extract lesions from the normal breast tissue. In this method, a radial gradient index (RGI) based segmentation²¹ is used to yield an initial contour in a computationally efficient manner. This initial segmentation also provides a base to identify the effective circumstance of the lesion via an automatic background estimation method. Then a region-based active contour model^{22,23} is utilized to evolve the contour further to the lesion margin. The active contour model relies on an intrinsic property of image segmentation in that each segmented region (i.e., the lesion region and the parenchymal background region) should be as homogeneous as possible. Thus, the contour evolution tries to minimize the following energy function:

$$\begin{aligned}
 E(c_1, c_2, C) = & \mu \cdot \text{Length}(C) + \nu \cdot \frac{1}{2} \int_{\Omega} (1 - \|\nabla \phi_l\|)^2 dx dy \\
 & + \lambda_1 \cdot \int_{\text{inside}(C)} |f_0(x, y) - c_1|^2 dx dy \\
 & + \lambda_2 \cdot \int_{\text{outside}(C)} |f_0(x, y) - c_2|^2 dx dy, \quad (1)
 \end{aligned}$$

where $\mu \geq 0$, $\nu \geq 0$, $\lambda_1, \lambda_2 > 0$ are fixed weight parameters, C is the evolving contour, and $\text{Length}(C)$ is a regularizing term that prevents the final contour from converging to a small area due to noise. Ω represents the entire image space and $\int_{\Omega} (1 - \|\nabla \phi_l\|)^2 dx dy$ is an additional regularizing term that provides a smoother contour and pushes the contour closer to the lesion margin with less iterations. c_1 and c_2 are mean values inside and outside of C , respectively. The minimization of this energy function can be achieved by level set theory²⁴ and Calculus of Variations, in which the two-dimensional evolving contour C is represented implicitly as the zero level set of a three-dimensional function $\phi(x, y)$, i.e., $C = \{(x, y) \in \Omega : \phi(x, y) = 0\}$. Instead of empirically determined criteria such as fixed iteration times, a dynamic stopping criterion is implemented to automatically terminate the contour evolution when it reaches the lesion boundary.

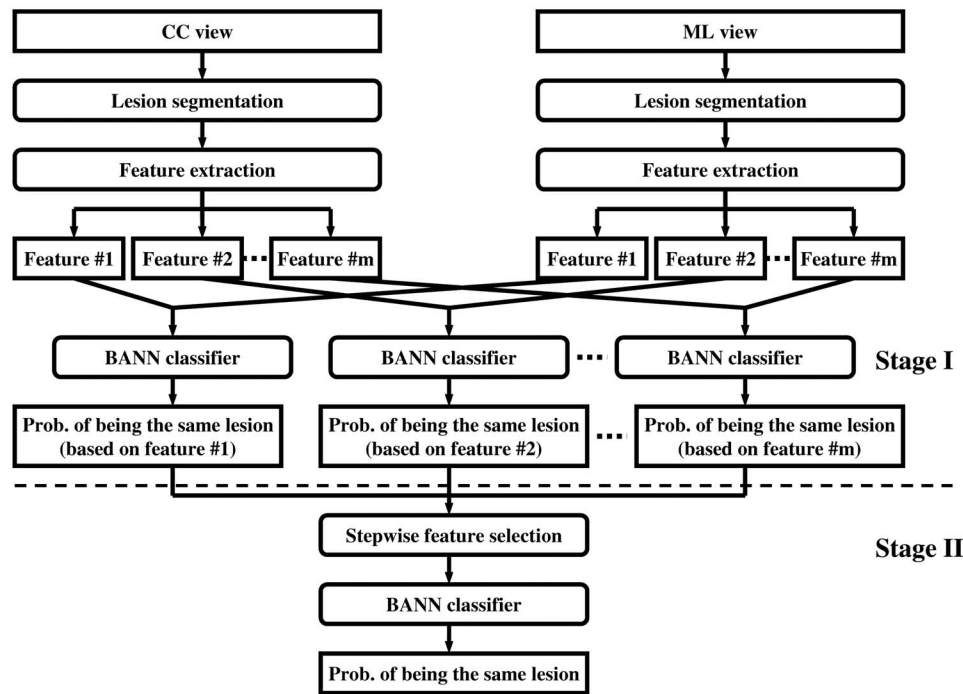


FIG. 1. Schematic diagram of the proposed correlative feature analysis.

II.C. Computerized feature extraction

In this study, our primary interest is to investigate the potential usefulness of various computer-extracted features in the task of differentiating corresponding image pairs from noncorresponding ones. Features in this study are grouped into three categories: (I) margin and density features, (II) texture features based on gray-level co-occurrence matrix (GLCM), and (III) a distance feature. The features in the first two categories have been described in detail elsewhere^{25–27} and are only summarized here.

II.C.1. Margin and density features

Margin and density of a mass are two important properties used by radiologists when assessing the probability of malignancy of mass lesions. The margin of a mass can be characterized by its sharpness and spiculation. The *margin sharpness* is calculated as the average of the gradient magnitude along the margin of the mass.²⁵ The *margin spiculation* is measured by the full width at half maximum (FWHM) of the normalized edge-gradient distribution calculated for a neighborhood of the mass with respect to the radial direction, and by the normalized radial gradient (NRG).²⁵ Three features were extracted to characterize different aspects of the density of a lesion. *Gradient texture* is the standard deviation of the gradient within a mass lesion. *Average gray value* is obtained by averaging the gray level values of each pixel within the segmented region of mass lesion, and *contrast* measures the difference between the average gray level of the segmented region and that of the surrounding parenchyma. Furthermore, an *equivalent diameter* feature was also used in this study, which is defined as the diameter of a circle yielding the same area as the segmented lesion.

II.C.2. Texture features

The calculation of texture features in our study is based on the gray-level co-occurrence matrix (GLCM).^{4,19,26,27} For an image with G gray levels, the corresponding GLCM is of size $G \times G$, where each element of the matrix is the joint probability ($p_{r,\theta}(i,j)$) of the occurrence of gray levels i and j in two paired pixels with an offset of r (pixels) along the direction θ in the image.

Fourteen texture features were extracted from the GLCM matrix, including *contrast*, *correlation*, *difference entropy*, *difference variance*, *energy*, *entropy*, *homogeneity*, *maximum correlation coefficient*, *sum average*, *sum entropy*, *sum variance*, *variance*, and two *information measures of correlation*. These features quantify different characteristics of a lesion, such as homogeneity, gray-level dependence, brightness, variation, and randomness.

In our study, texture features were extracted from the lesion and the associated region of interest (ROI). A ROI includes a lesion and its surrounding neighborhood, which was determined by an automatic estimation method developed in our prior study.²⁰ Here, an effective neighborhood is defined as the set of pixels within a distance d (pixels) from the circumscribed rectangle of the segmented lesion, as shown in Fig. 3. It should be noted that this neighborhood estimation is similar to that used earlier in the lesion segmentation, however, here the ROI is centered to the edge of the segmented lesion. Furthermore, a two-dimensional linear background trend correction was employed after the ROI extraction to eliminate the low-frequency background variations in the mammographic region.²⁰

For each region, four GLCMs were constructed along four different directions of 0° , 45° , 90° , and 135° , and a

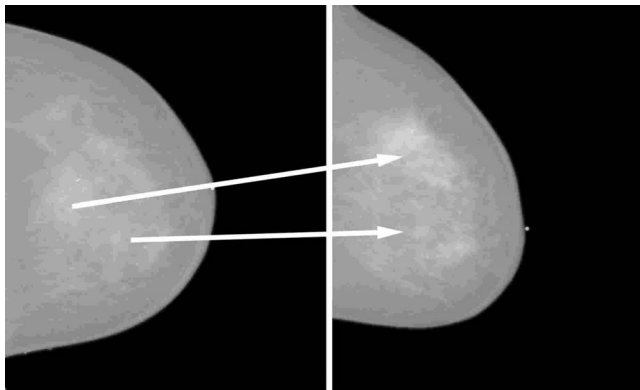


FIG. 2. An example of two lesions in the same breast as seen in CC view (left) and ML view (right). The arrow indicates the correspondence of the same physical lesion in different views.

nondirectional GLCM was obtained by summing all the directional GLCMs. Texture features were computed from each nondirectional GLCM, resulting in a total of 28 texture features. To avoid sparse GLCMs for smaller ROIs, the gray level range of all the image data was scaled down to 6 bits, resulting in GLCMs of size 64×64 . The offset r was empirically determined to be 16 (pixels).

II.C.3. Distance feature

In clinical practice, radiologists commonly use the distance from the nipple to the center of a lesion to correlate the lesion in different views.^{4,5} It is generally believed that this distance remains fairly constant. Thus, a distance feature in our study measures the Euclidean distance between the nipple location and the mass center of the lesion. Figure 4 shows the high correlation between the distance features of the same lesions in CC and ML views, with a correlation coefficient of 0.88. For this figure, the nipple locations were manually identified.

In mammographic images, nipple markers are commonly used. These present as bright markers on the mammograms (as shown in Fig. 5), and, thus, an automatic nipple localization scheme was developed to locate those markers. The scheme includes several processing stages. Initially, gray-level thresholding is employed to the entire mammogram to

TABLE I. The number of lesion/image pairs in corresponding and noncorresponding datasets. The noncorresponding pairs were constructed from cases of the same breasts but different physical lesions.

	Corresponding dataset	Noncorresponding dataset
Benign		
Images	112	72
Lesions	56	39
Malignant		
Images	134	64
Lesions	67	19
Mixed		
Images	—	28
Lesions	—	14

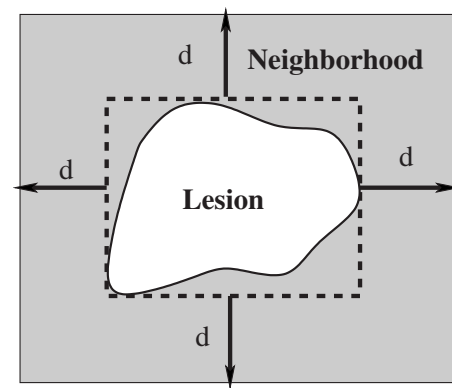


FIG. 3. Lesion neighborhood illustration.

extract the breast region from the air background. Then, another gray-level threshold is applied to the breast region, yielding several nipple marker candidates. The breast skin boundary is obtained by subtracting a morphologically eroded²⁸ breast region from its original region. To reduce the number of falsely identified nipple markers, area and circularity constraints are imposed on each candidate, and those candidates with area within a given range and circularity above a certain threshold are kept for the final step. The area range and circularity threshold were empirically determined with ten randomly selected images in this study. The nipple marker is finally chosen as the one closest to the breast boundary. For those cases in which there is no nipple marker or the marker is neglected erroneously by the above scheme, the nipple location is roughly estimated as the point on the breast skin boundary with the largest distance to the chest wall.

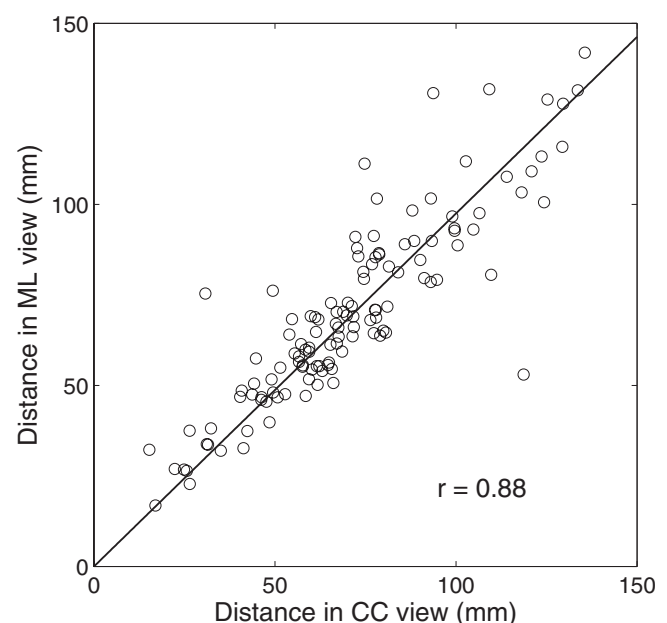


FIG. 4. The correlation between distance features of the same lesions in CC and ML views. The distance feature is defined as the Euclidean distance between the nipple location and the mass center of the lesion. Here, the nipple location is manually identified.

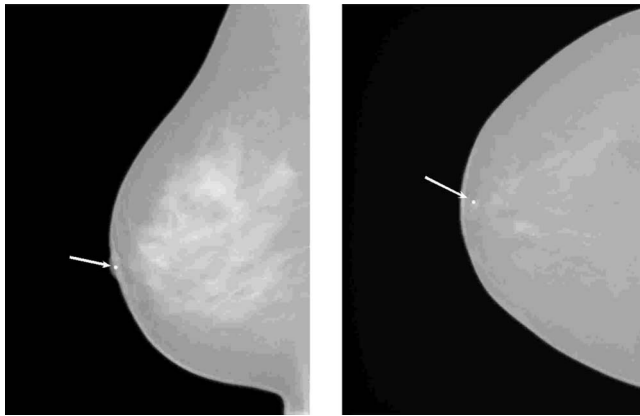


FIG. 5. Two examples of nipple markers. Nipple markers are bright spots close to the breast skin boundary, as indicated by arrows.

II.D. Feature selection and classification

For each pairwise set of features extracted from lesions in different views, a Bayesian Artificial Neural Network (BANN) classifier^{29,30} was employed to merge each feature pair into a correspondence metric, which is an estimate of the probability that the two lesion images are of the same physical lesion, i.e., stage I as shown in Fig. 1. For example, Fig. 6(a) shows the distributions of three features (*distance*, *diameter*, and *texture*) generated from breast lesions taken in different views for corresponding and noncorresponding datasets. The histograms in Fig. 6(b) demonstrate, for the corresponding and noncorresponding datasets, the distribution of these correspondence metrics output from the first BANN.

Linear stepwise feature selection³¹ with Wilks lambda criterion was employed on all feature-based correspondence metrics to select a subset of metrics for the final task of distinguishing corresponding pairs from noncorresponding ones. Note that instead of using lesion features directly, the *correspondence metrics* obtained from the first BANNs are used as inputs in the feature selection. BANNs were then retrained with the selected *correspondence metrics* to yield an overall estimate of probability of correspondence, i.e., the second BANN stage as shown in Fig. 1.

II.E. Evaluation

Receiver operating characteristic (ROC) analysis^{32,33} was used to assess the performance of the individual feature-based correspondence metrics and the overall performance in the task of distinguishing corresponding image pairs from noncorresponding ones. The area under the maximum likelihood-estimated binormal ROC curve (AUC) was used as an index of performance. ROCKIT software (version 1.1 b, available at http://xray.bsd.uchicago.edu/krl/KRL_R0C/software_index6.htm)³⁴ was used to determine the *p* value of the difference between two AUC values, and the Holm *t* test³⁵ for multiple tests of significance was employed to

evaluate the statistical significance. Leave-one-out by lesion analysis was used in all performance evaluations. This method removes all images of a lesion while training with all other images. The trained classifier is then run on the images of the lesion removed. In the case of correspondence analysis, images of all pairs, both corresponding and noncorresponding pairs, are removed in the training to eliminate bias.

III. RESULTS

III.A. Segmentation

Figure 7 shows two examples of lesion segmentation using the dual-stage segmentation method. A measure of area overlap ratio (AOR) was used to quantitatively evaluate the segmentation performance, which is defined as the intersection of human outline and computer segmentation over the union of them. At the overlap threshold of 0.4, 81% of the images are correctly segmented.

III.B. Nipple localization method

Figure 8 shows the correlation between distance features calculated with manually identified nipples and those calculated with computer-identified nipples. These two distance features are highly correlated with correlation coefficient of 0.996 ($p\text{-value} < 10^{-4}$). Both of these two distance features have an AUC value of 0.81 ± 0.02 in the task of distinguishing between corresponding and noncorresponding image pairs.

III.C. Performance of single-feature correspondence metrics

We calculated the correlation coefficient *r* for the corresponding dataset, the *r'* for the noncorresponding datasets, and their associated *p-value* for features extracted from two view images. Table II shows the results for features with $r \geq 0.5$. It also shows the AUC values and the associated standard errors (se) representing the performance of the correspondence metrics of these individual features in the task of differentiating the corresponding lesion pairs from noncorresponding ones, with the lesions automatically delineated by the segmentation algorithm. The results show that all three categories have potential for the classification task. The results also show that the performance of pairwise image classification as learned by a BANN is determined by both the correlation of corresponding pairs and that of noncorresponding pairs.

We also investigated the effect of lesion segmentation on the performance of each individual feature-based correspondence metric. Table III shows the AUC values and the associated standard error (se) of the 18 features extracted from lesions delineated by a radiologist and by the dual-stage segmentation algorithm, respectively. Also shown are the 95% confidence intervals (C. I.) of the difference of AUCs obtained from radiologist-outlined lesions (AUC_R) and the computer-segmented lesions (AUC_C), i.e., $\Delta AUC = AUC_R - AUC_C$. For 5 of the 18 features, manual segmentation yielded statistically significant higher AUC values than com-

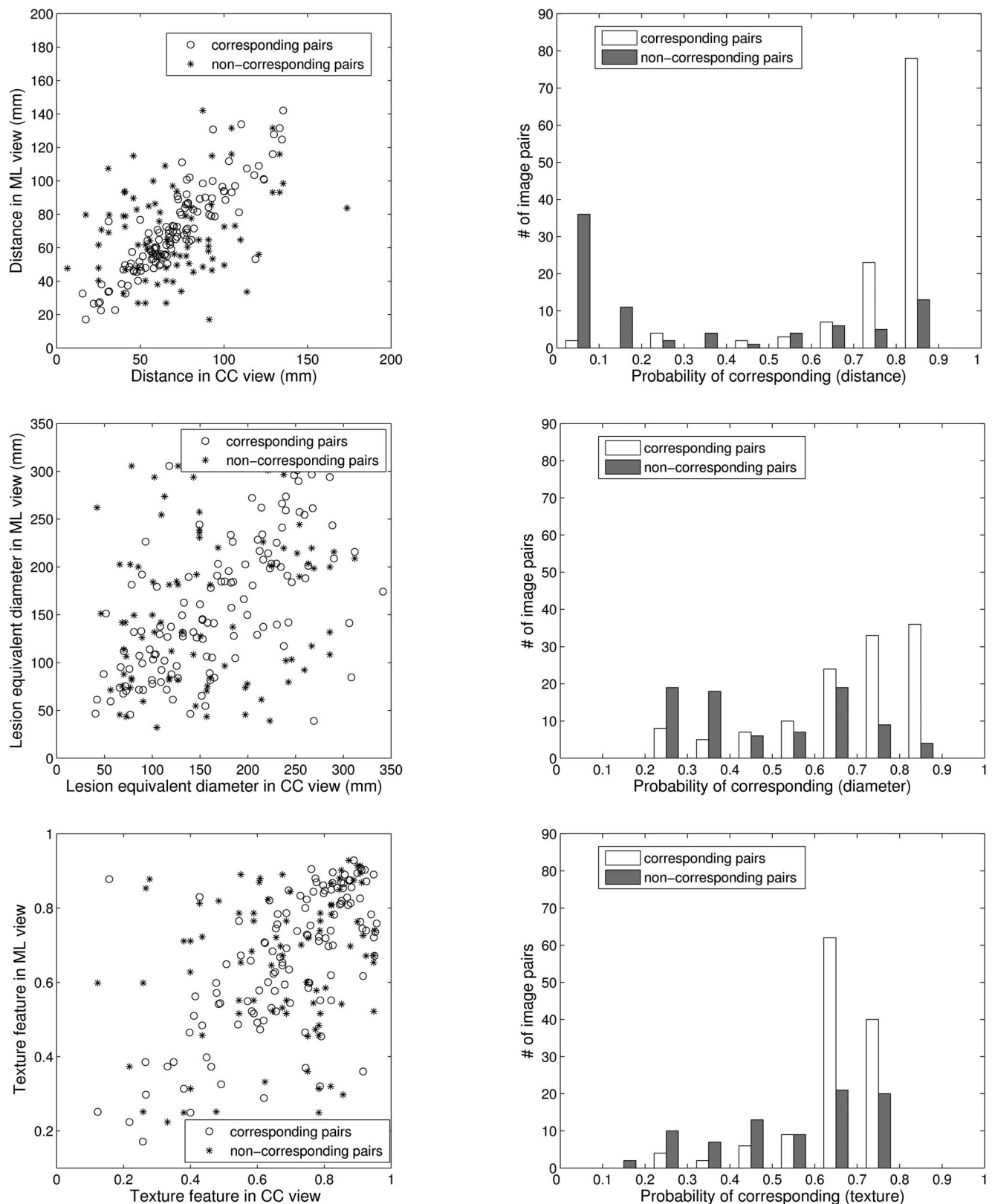


FIG. 6. (a) The scatter plots of three features (distance, diameter, and texture) generated from lesions seen on CC and ML views. (b) The distribution of the output correspondence metrics of these features obtained from the first BANN stage.

puter segmentation (overall significant level $\alpha^T=0.05$),³⁵ and we failed to show significant differences between manual segmentation and computer segmentation for the remaining features.

III.D. Multiple features performance

Two sets of individual feature-based correspondence metrics were selected by stepwise feature selection³¹—one set

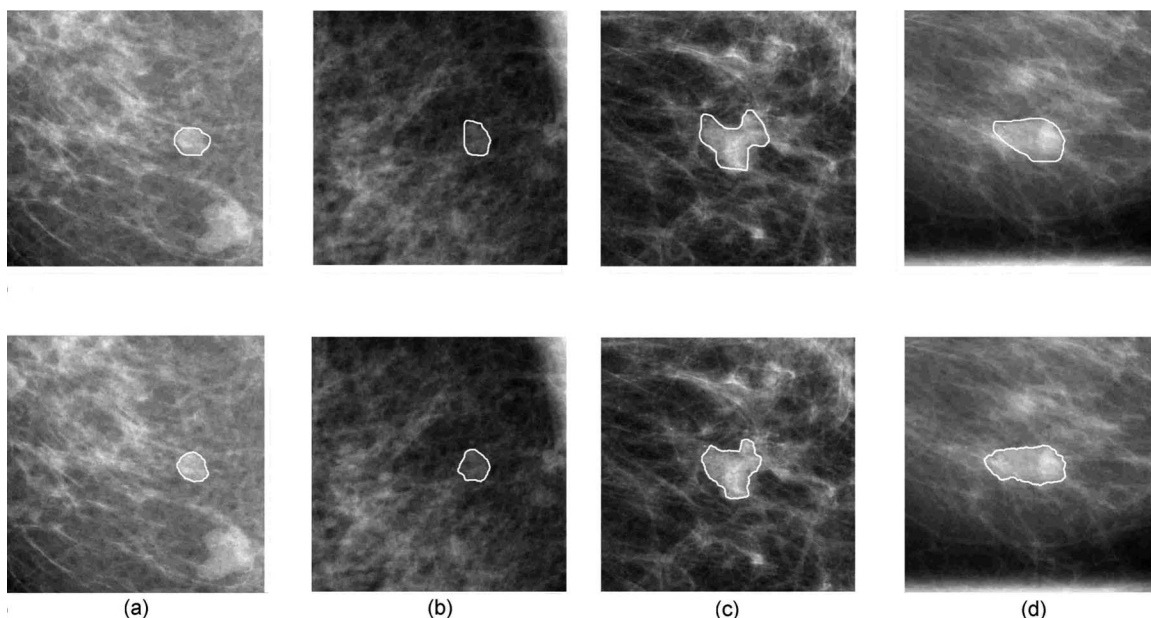


FIG. 7. Segmentation results for a benign lesion and a malignant lesion. The solid lines in the upper four images depict the lesion margin as outlined by a radiologist, and the solid lines in the bottom four images are segmentation results from our previously-reported automatic dual-stage method (Ref. 20). (a) CC view of a benign lesion, (b) the corresponding ML view of the benign lesion, (c) CC view of a malignant lesion, and (d) the corresponding ML view of the malignant lesion.

for each of the two segmentation methods, as shown on Table IV. The subset selected from the feature-based correspondence metrics based on manually segmented lesions included *distance* ($F_{III,1}$), *equivalent diameter* ($F_{I,3}$), and *gradient texture* ($F_{I,1}$). The subset selected from computer-segmented lesions included *distance* ($F_{III,1}$), ROI-based *correlation* ($F_{II,6}$), and *gradient texture* ($F_{I,1}$). The leave-one-out (by lesion) validation using BANN to merge the selected

correspondence metrics yielded an AUC of 0.89 for manual segmentation and 0.87 for computer segmentation, respectively. We failed to show a statistically significant difference between the performances of these two metric subsets ($p=0.35$). The improvement by using multiple-feature-based correspondence metrics was statistically significant compared to that of single feature-based correspondence metric performance, as shown in Table IV.

Since the *distance* feature performed best among the individual features for differentiating corresponding and non-corresponding image pairs, we evaluated the performance of the proposed correlative feature analysis method with the *distance* feature excluded. Using the remaining 17 features extracted from the computer-segmented lesions, a feature-based correspondence metric subset was obtained by step-wise feature selection, which included *equivalent diameter* ($F_{I,3}$), ROI-based *correlation* ($F_{II,6}$), and ROI-based *sum of variance* ($F_{II,14}$). The leave-out-out (by lesion) validation using BANN yielded an AUC of 0.71 ± 0.03 . The difference as compared to the performance of *distance* feature is statistically significant ($p=0.005$). This result indicates that the *distance* feature is dominant but not sufficient for the overall performance of the proposed method.

IV. DISCUSSION

In this study, we presented a correlative feature analysis framework to assess the probability that a given pair of two mammographic images is of the same physical lesion. Our results demonstrate that this framework has potential to distinguish between corresponding and noncorresponding lesion pairs. It is very important to note that our method is feature based, which employs two BANN classifiers to estimate the

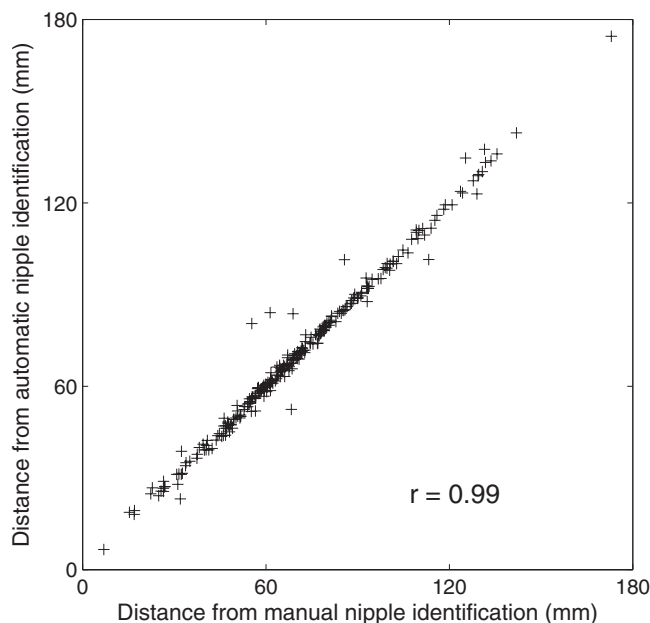


FIG. 8. The correlation between distance features calculated from manually identified nipple locations and those calculated from computer-identified nipple locations.

TABLE II. Performance of the correspondence metrics from computer-extracted lesion features that yielded $r \geq 0.5$ in differentiating corresponding image pairs from noncorresponding ones. r is the correlation coefficient for the corresponding dataset and r' is for the noncorresponding dataset. The value after “ $\pm$ ” is the standard error (se) associated with each AUC.

	Corresponding pairs		Noncorresponding pairs		AUC $\pm$ se
	r	p value	r'	p value	
I. Density and morph. features					
$F_{I,1}$: Gradient texture	0.53	<0.001	0.27	0.01	0.56 ± 0.03
$F_{I,2}$: Average gray level	0.58	<0.001	-0.10	0.39	0.54 ± 0.03
$F_{I,3}$: Equivalent diameter	0.62	<0.001	0.14	0.22	0.66 ± 0.03
II. Texture features					
* Lesion based					
$F_{II,1}$: Correlation	0.56	<0.001	0.13	0.25	0.65 ± 0.03
$F_{II,2}$: Info. corr. 1	0.50	<0.001	0.06	0.61	0.67 ± 0.03
$F_{II,3}$: Info. corr. 2	0.53	<0.001	0.09	0.40	0.67 ± 0.03
$F_{II,4}$: Max. corr.	0.53	<0.001	0.11	0.35	0.66 ± 0.03
** ROI based					
$F_{II,5}$: Contrast	0.58	<0.001	0.16	0.15	0.54 ± 0.03
$F_{II,6}$: Correlation	0.67	<0.001	0.24	0.03	0.56 ± 0.03
$F_{II,7}$: Diff. variance	0.61	<0.001	0.20	0.07	0.53 ± 0.03
$F_{II,8}$: Entropy	0.51	<0.001	0.15	0.17	0.56 ± 0.03
$F_{II,9}$: Info. corr. 1	0.62	<0.001	0.16	0.15	0.61 ± 0.03
$F_{II,10}$: Info. corr. 2	0.62	<0.001	0.14	0.21	0.57 ± 0.03
$F_{II,11}$: Max. corr.	0.61	<0.001	0.11	0.33	0.55 ± 0.03
$F_{II,12}$: Sum. Average	0.63	<0.001	0.27	0.01	0.59 ± 0.03
$F_{II,13}$: Sum. Entropy	0.53	<0.001	0.16	0.15	0.57 ± 0.03
$F_{II,14}$: Sum. Variance	0.61	<0.001	0.41	<0.001	0.50 ± 0.03
III. Distance feature					
$F_{III,1}$: Distance	0.88	<0.001	0.23	0.04	0.81 ± 0.02

TABLE III. Performance of 18 single feature-based correspondence metrics obtained from radiologist-outlined (AUC_R) and computer-segmented (AUC_C) lesions, respectively. The value after “ $\pm$ ” is the standard error (se) associated with each AUC. The two-tailed p -value and 95% C.I. of ΔAUC were calculated by ROCKIT. The “Sig. level” column represents the significance level of individual tests adjusted with Holm t test (overall significant level $\alpha^T=0.05$) and the tests with asterisks (*) indicate significant difference using the adjusted significance level. The features have the same convention as Table II.

Feature	$AUC_R \pm se$	$AUC_C \pm se$	p value	Sig. level	95% C.I. of ΔAUC
$F_{I,1}$	0.65 ± 0.03	0.56 ± 0.03	0.04	0.0045	[0.004, 0.20]
$F_{I,2}$	0.53 ± 0.03	0.54 ± 0.03	0.76	—	[-0.07, 0.05]
$F_{I,3}^*$	0.78 ± 0.03	0.66 ± 0.03	0.001	0.0031	[0.05, 0.19]
$F_{II,1}$	0.71 ± 0.03	0.65 ± 0.03	0.06	—	[-0.01, 0.13]
$F_{II,2}$	0.68 ± 0.03	0.67 ± 0.03	0.66	—	[-0.05, 0.09]
$F_{II,3}$	0.69 ± 0.03	0.67 ± 0.03	0.48	—	[-0.04, 0.09]
$F_{II,4}$	0.70 ± 0.03	0.66 ± 0.03	0.20	—	[-0.02, 0.11]
$F_{II,5}$	0.57 ± 0.03	0.54 ± 0.03	0.30	—	[-0.03, 0.10]
$F_{II,6}$	0.61 ± 0.03	0.56 ± 0.03	0.01	0.0042	[0.01, 0.10]
$F_{II,7}$	0.61 ± 0.03	0.53 ± 0.03	0.009	0.0038	[0.02, 0.15]
$F_{II,8}$	0.58 ± 0.03	0.56 ± 0.03	0.44	—	[-0.03, 0.07]
$F_{II,9}^*$	0.69 ± 0.03	0.61 ± 0.03	0.002	0.0036	[0.03, 0.14]
$F_{II,10}^*$	0.65 ± 0.03	0.57 ± 0.03	4×10^{-4}	0.0029	[0.04, 0.13]
$F_{II,11}^*$	0.66 ± 0.03	0.55 ± 0.03	$<10^{-5}$	0.0028	[0.06, 0.15]
$F_{II,12}$	0.62 ± 0.03	0.59 ± 0.03	0.34	—	[-0.03, 0.09]
$F_{II,13}$	0.58 ± 0.03	0.57 ± 0.03	0.90	—	[-0.05, 0.05]
$F_{II,14}^*$	0.59 ± 0.03	0.50 ± 0.03	0.001	0.0031	[0.05, 0.18]
$F_{III,1}$	0.81 ± 0.02	0.81 ± 0.02	0.73	—	[-0.01, 0.01]

TABLE IV. Performances of the overall correlative feature analysis method using leave-one-out (by lesion) validation, as well as the comparison with the distance feature alone. This table also shows the comparison between the overall performances of merged features obtained from radiologist-outlined and computer-segmented lesions. Same convention as Table III.

Lesion segmentation	Feature set	AUC $\pm$ se		p value	95% C.I. of Δ AUC
Radiologist outlined	$F_{III,1}$	0.81 ± 0.02	}	3×10^{-4}	[0.04, 0.12]
	$F_{III,1}, F_{I,3}, F_{I,1}$	0.89 ± 0.02		0.35	[-0.02, 0.06]
Computer segmented	$F_{III,1}, F_{II,6}, F_{I,1}$	0.87 ± 0.02	}	0.01	[0.01, 0.08]
	$F_{III,1}$	0.81 ± 0.02			

relationships (linear or nonlinear) between computer-extracted features of a lesion in different views. The supervised-learning manner not only makes the relationship flexible to each feature, but also avoids the sophisticated geometrically deformable models and the corresponding computationally demanding optimizations that are used in geometric breast registrations.^{36,37}

In our study, we excluded those features characterizing subtle information of a lesion, such as spiculation, margin sharpness, and normalized radial gradient (NRG). These features have been used in the task of distinguishing between benign and malignant lesions for mammographic images.^{9,25} However, as the lesion details are usually sensitive to positions, it is expected that the associated features are less correlated in different views. Nevertheless, our ultimate aim is to improve the diagnostic performance of CAD systems with multiple images, in which complementary information provided by different images are desired; therefore, those features corresponding to lesion details would be used in the later step of the overall CAD scheme for differentiating between malignant and benign lesions.

In addition, as shown in Table III, improving lesion segmentation can improve the performance of the computer in differentiating corresponding and noncorresponding image pairs. This is expected since more accurate segmentation yields more reliable computer-extracted features with which to characterize the lesion and the two-view correspondence.

A two-stage procedure was employed to address the problem of estimating the probability of correspondence for a pair of lesion images in different views. Stage I deals with the pairwise features and estimates the probability of correspondence based on individual lesion features. Stage II merges the correspondence metrics estimated in stage I from various individual lesion features to yield an overall probability of correspondence. To illustrate the superiority of the proposed two-stage method to a one-stage method that combines the multiple paired features directly, we compared the performances of the two methods with the four features of *distance*, *lesion equivalent diameter*, *lesion-based correlation*, and *lesion-based information correlation*, all of which performed best among the 18 individual features extracted from computer-segmented lesions. The two-stage scheme yielded an AUC of 0.83 while the one-stage scheme yielded an AUC of 0.67, with difference being statistically significant ($p < 10^{-4}$). The inferior performance of the one-stage scheme

can be mainly explained by the fact that a single BANN classifier lacks the ability to deal with features in a pairwise way, thus the information regarding correlation between feature pairs cannot be efficiently utilized.

In order to evaluate how the pathology of lesions affects the performance of the proposed method, the entire dataset was split into benign and malignant subsets, as described in Table I. As noted earlier, the correlation value between pairwise features, and not the feature value itself, plays a crucial role in the task of distinguishing corresponding and noncorresponding image pairs, thus we compared the correlation coefficients between image pairs for benign and malignant lesions, respectively. We failed to observe significant difference for most of features between benign and malignant lesions, as shown in Fig. 9. The results indicate that the pairwise feature analysis may be independent of pathology.

Due to the database size, there are two limitations in this preliminary study. First, the proposed correlative feature analysis was only applied on CC versus ML views, however, pairing other views, such as CC versus MLO and ML versus MLO, is also commonly used in clinical practice. Thus, in further study, we will evaluate the computerized analysis on those view pairs and investigate how the different pairwise views affect the performance of proposed analysis. Second, for noncorresponding pairs, lesions could be with either same pathology (i.e., both malignant or both benign) or different pathology (i.e., one malignant and one benign). Specifically, we are more interested in noncorresponding lesions of different pathology since integrating information from lesions with different pathology would hinder more the performance of CAD systems. However, we regarded the noncorresponding lesion pairs as a whole in this study as there are only 28 image pairs with different pathology. The performance of the proposed analysis on noncorresponding lesion pairs with different pathology, and the mismatching effects on the CAD performance are interesting research questions for our future study.

V. CONCLUSION

In this paper, we have presented a novel two-BANN correlative feature analysis framework to estimate the probability that a given pair of two images is of the same physical lesion. Our investigation indicates that the proposed method is a promising way to distinguish between corresponding and

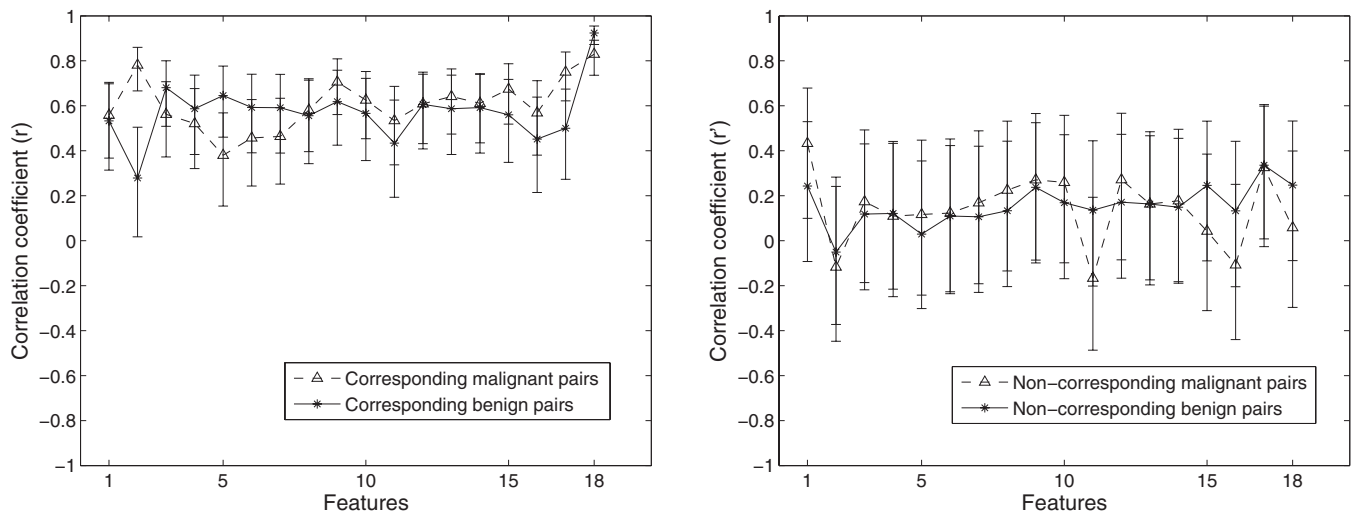


FIG. 9. Correlation coefficients between CC and ML views and associated 95% confidence intervals of the 18 features extracted from benign (solid) and malignant (dash) lesions. All the lesions were segmented via an automatic segmentation algorithm. Left: corresponding image pairs. Right: noncorresponding image pairs.

noncorresponding pairs. With leave-one-out (by lesion) cross validation, the distance-feature-based correspondence metric yielded an AUC of 0.81 and a feature correspondence metric subset, which includes *distance*, *gradient texture*, and ROI-based *correlation*, yielded an AUC of 0.87. The improvement by using multiple feature correspondence metrics was statistically significant compared to single feature metric performance. This method has the potential to be generalized and employed to differentiating corresponding and noncorresponding pairs from multi-modality breast imaging.

ACKNOWLEDGMENTS

This work was supported in part by US Army Breast Cancer Research Program (BCRP) Predoctoral Traineeship Award (W81XWH-06-1-0726), by United States Public Health Service (USPHS) Grant Nos. CA89452 and P50-CA125183, and by Cancer Center Support Grant (5-P30CA14599). MLG is a stockholder in, and receives royalties from, R2 Technology, Inc (Sunnyvale, CA), a Hologic Company. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actually or potential significant financial interest which would reasonably appear to be directly and significantly affected by the research activities.

^aElectronic mail: yading@uchicago.edu; Telephone: (773) 834-5101; Fax: (773) 702-0371.

¹A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, and M. J. Thun, "Cancer statistics, 2008," *Ca-Cancer J. Clin.* **58**, 71–96 (2008).

²A. J. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, and M. J. Thun, "Cancer statistics, 2007," *Ca-Cancer J. Clin.* **57**, 43–66 (2007).

³S. H. Heywang-Kobrunner, D. D. Dershaw, and I. Schreer, *Diagnostic Breast Imaging: Mammography, Sonography, Magnetic Resonance Imaging, and Interventional Procedures*, 2nd ed. (Thieme Medical Publisher, Stuttgart, 2001).

⁴S. Paquerault, N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Improvement of computerized mass detection on mammograms: fusion of two-view information," *Med. Phys.* **29**, 238–247 (2002).

⁵B. Zheng, J. K. Leader, G. S. Abrams, A. H. Lu, L. P. Wallace, G. S. Maitz, and D. Gur, "Multiview-based computer-aided detection scheme

for breast masses," *Med. Phys.* **33**, 3135–3143 (2006).

⁶S. van Engeland and N. Karssemeijer, "Combining two mammographic projections in a computer aided mass detection method," *Med. Phys.* **34**, 898–905 (2007).

⁷Y. Jiang, R. M. Nishikawa, D. E. Wolverton, C. E. Metz, M. L. Giger, R. A. Schmidt, C. J. Vyborny, and K. Doi, "Malignant and benign clustered microcalcifications: Automated feature analysis and classification," *Radiology* **198**, 671–678 (1996).

⁸H. P. Chan, B. Sahiner, K. L. Lam, N. Petrick, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces," *Med. Phys.* **25**, 2007–2019 (1998).

⁹Z. Huo, M. L. Giger, and C. J. Vyborny, "Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis," *IEEE Trans. Med. Imaging* **20**, 1285–1292 (2001).

¹⁰B. Liu, C. E. Metz, and Y. Jiang, "An ROC comparison of four methods of combining information from multiple images of the same patient," *Med. Phys.* **31**, 2552–2563 (2004).

¹¹D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Phys. Med. Biol.* **46**, R1–R45 (2001).

¹²J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imaging* **22**, 986–1004 (2003).

¹³F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imaging* **16**, 187–198 (1997).

¹⁴N. Vujovic and D. Brzakovic, "Control points in pairs of mammographic images," *IEEE Trans. Inf. Theory* **6**, 1388–1399 (1997).

¹⁵S. van Engeland, P. Snoeren, J. Hendriks, and N. Karssemeijer, "A comparison of methods for mammogram registration," *IEEE Trans. Med. Imaging* **22**, 1436–1444 (2003).

¹⁶K. Marias, C. Behrenbruch, S. Parbhoo, A. Seifalian, and M. Brady, "A registration framework for the comparison of mammogram sequences," *IEEE Trans. Med. Imaging* **24**, 782–790 (2005).

¹⁷M. L. Giger, Z. Huo, M. A. Kupinski, and C. J. Vyborny, "Computer-aided diagnosis in mammography," *Proc. SPIE* **2**, 915–1004 (2000).

¹⁸Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick, and P. Lu, "Analysis of spiculation in the computerized classification of mammographic masses," *Med. Phys.* **22**, 1569–1579 (1995).

¹⁹S. Gupta and M. K. Markey, "Correspondence in texture features between two mammographic views," *Med. Phys.* **32**, 1598–1606 (2005).

²⁰Y. Yuan, M. L. Giger, H. Li, K. Suzuki, and C. Sennett, "A dual-stage method for lesion segmentation on digital mammograms," *Med. Phys.* **34**, 4180–4193 (2007).

²¹M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation

- on digital mammograms," *IEEE Trans. Med. Imaging* **17**, 510–517 (1998).
- ²²T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Inf. Theory* **10**, 266–277 (2001).
- ²³C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: A new variational formulation," in *Proc. 2005 IEEE CVPR*, San Diego, 2005, pp. 1:430–436.
- ²⁴S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulation," *J. Comput. Phys.* **79**, 12–49 (1988).
- ²⁵Z. 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).
- ²⁶R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **3**, 610–621 (1973).
- ²⁷W. Chen, M. L. Giger, H. Li, U. Bick, and G. M. Newstead, "Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images," *Magn. Reson. Med.* **58**, 562–571 (2007).
- ²⁸M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision* (PWS Publishing, Pacific Grove, CA, 1998).
- ²⁹C. M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, Oxford, 1995).
- ³⁰M. A. Kupinski, D. C. Edwards, M. L. Giger, and C. E. Metz, "Ideal observer approximation using Bayesian classification neural networks," *IEEE Trans. Med. Imaging* **20**, 886–899 (2001).
- ³¹P. A. Lachenbruch, *Discriminant Analysis* (Hafner, London, 1975).
- ³²C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.* **21**, 720–733 (1986).
- ³³C. E. Metz, B. A. Herman, and J. Shen, "Maximum likelihood estimation of receiver operating characteristic ROC curves from continuously-distributed data," *Stat. Med.* **17**, 1033–1053 (1998).
- ³⁴C. E. Metz, B. A. Herman, and C. A. Roe, "Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets," *Med. Decis Making* **18**, 110–121 (1998).
- ³⁵S. A. Glantz, *Primer of Biostatistics* (McGraw-Hill, New York, 2002).
- ³⁶F. Richard and L. Cohen, "A new image registration technique with free boundary constraints: application to mammography," *Comput. Vis. Image Underst.* **89**, 166–196 (2003).
- ³⁷Y. Kita, R. Highnam, and M. Brady, "Correspondence between different view breast X-rays using a simulation of breast deformation," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition* (1998), pp. 700–707.

Appendix D

TITLE:

Feature correlation on multiple-view FFDM images

AUTHORS:

Yading Yuan, Maryellen L. Giger, Hui Li and Charlene Sennett

PURPOSE:

The objective of this study is to evaluate the correlation performance of individual computerized features extracted from the full field digital mammograms (FFDM) of a lesion obtained in two different views. This research provides a guide for discriminating corresponding and non-corresponding lesion pairs within the CAD framework. It is also helpful for guiding the development of new features to improve the accuracy of image matching in disease diagnosis and prognosis.

METHOD AND MATERIALS:

One dataset (A) includes 103 biopsy proven cases (48 benign solid lesions and 55 malignant lesions), each of which has a craniocaudal (CC) and mediolateral (ML) view. Another dataset (B) includes 52 cases (24 benign solid lesions and 28 malignant lesions), each of which has a CC and mediolateral oblique (MLO) view. In order to evaluate the robustness of the correlation performance to lesion segmentation, besides the radiologist's outlines, three automatic segmentation methods were employed to extract the mass lesion from the surrounding tissues. The conventional region-growing method uses abrupt changes in size and circularity as the rules of segmentation. The radial gradient index (RGI) based method applies RGI model to the suspicious lesion multiplied by a constraint function. The region-based active contour model evolves the contour based on the homogeneities both inside and outside of the evolving contour. Fifteen computer-extracted features of each lesion were calculated in both views in order to quantify the characteristics of margin, shape, contrast and texture of the lesion. For each feature, correlation coefficient between the two views and the p-value of the derived correlation coefficient were obtained.

RESULTS:

With the human outline, the feature characterizing the diameter of lesion yielded the correlation efficient of 0.87 for dataset A and 0.88 for dataset B, both of which have p-values far less than 0.05. The features characterizing shape, contrast and texture showed better performance among the 15 individual features despite of segmentation methods, pathology and the type of view pairs. This is because the features representing large-scale information are less sensitive to the change of position than those representing small-

scale information, which results in the higher correlation between large-scale features from different views than that of small-scale features.

CONCLUSIONS:

Our investigation indicates that the features that characterize the large-scale information of lesion have higher correlation between the two view images. We are currently applying these features to develop automated image matching method to determine corresponding and non-corresponding lesion pairs.

Identifying Corresponding Lesions from CC and MLO Views via Correlative Feature Analysis

Yading Yuan, Maryellen Giger, Hui Li, Li Lan, and Charlene Sennett

Department of Radiology The University of Chicago
5841 South Maryland Avenue, MC 2026
Chicago, IL 60637 USA
yading@uchicago.edu

Abstract. In this study, we present a computerized framework to identify the corresponding image pair of a lesion in CC and MLO views, a prerequisite for combining information from these views to improve the diagnostic ability of both radiologists and CAD systems. A database of 126 mass lesions was used, from which a corresponding dataset with 104 pairs and a non-corresponding dataset with 95 pairs were constructed. For each FFDM image, the mass lesions were firstly automatically segmented via a dual-stage algorithm, in which a RGI-based segmentation and an active contour model are employed sequentially. Then, various features were automatically extracted from the lesion to characterize the spiculation, margin, size, texture and context of the lesion, as well as its distance to nipple. We developed a two-step strategy to select an effective subset of features, and combined it with a BANN to estimate the probability that the two images are of the same physical lesion. ROC analysis was used to evaluate the performance of the individual features and the selected feature subset for the task of distinguishing corresponding and non-corresponding pairs. With leave-one-out evaluation by lesion, the distance feature yielded an AUC of 0.78 and the feature subset, which includes distance, ROI-based energy and ROI-based homogeneity, yielded an AUC of 0.88. The improvement by using multiple features was statistically significant compared to single feature performance ($p < 0.001$)

1 Introduction

In mammographic screening, different projections provide complementary information about the same physical lesion, and thus, it has been well recognized that multiple views can improve the diagnosis of breast cancer in the computerized analysis of mammograms [1–3]. To merge information from images of different views, an essential step is to clarify if these images actually represent the same physical lesion, as Fig. 1 shows. In this study, we present a correlative feature analysis (CFA) framework to address the task of identifying corresponding images of lesions as seen with craniocaudal (CC) and mediolateral oblique (MLO) views.

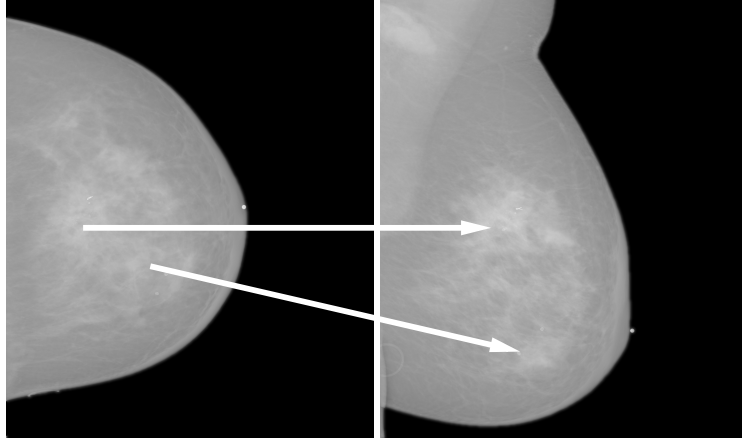


Fig. 1. An example of two lesions in the same breast seen in CC view (left) and MLO view (right). The arrow indicates the correspondence of the same physical lesion in different views.

2 Database

The full-field digital mammography (FFDM) database consists of 126 biopsied lesions obtained from GE Senographe 2000D systems with spatial resolution of $0.1mm \times 0.1mm$. The mass lesions were identified and outlined by an expert breast radiologist based on visual criterion and biopsy-proven reports. The distribution of lesion size, which is characterized as the equivalent diameter of the area inside the radiologist’s manually delineated contour, is shown in Fig. 2. Based on the correspondence of lesions identified by the radiologist, we constructed 104 corresponding and 95 non-corresponding image pairs. Each pair consists of a CC view and a MLO view. Considering the most realistic scenario of lesion mismatch in clinical practice, the non-corresponding pairs were built from cases of the same patients but different physical lesions.

3 Methods

3.1 Lesion segmentation

A dual-stage segmentation method was initially applied to extract mass lesions from the surrounding tissues [4]. This algorithm utilizes an active contour model that maximizes a cost function based on the homogeneities inside and outside of the evolving contour [5]. Prior to the application of the active contour model, a radial gradient index (RGI) based algorithm [6] is employed to yield an initial contour close to the lesion boundary location in a computationally efficient manner. Based on the initial segmentation, an automatic background estimation method is applied to identify the effective circumstance of the lesion. In addition, instead of empirically-determined criteria such as fixed iteration times, a

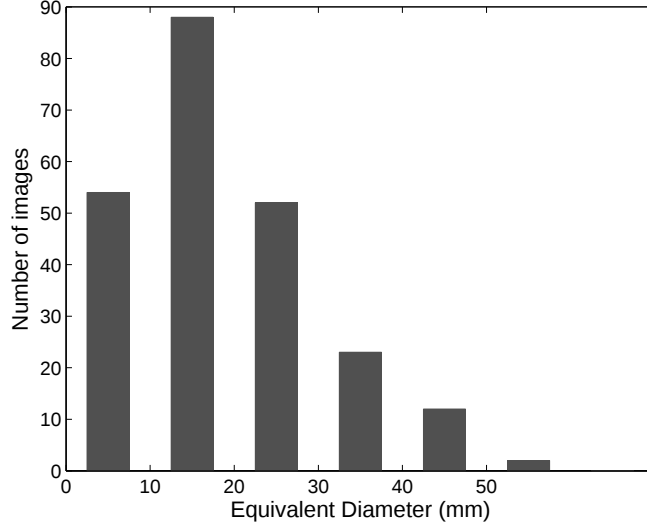


Fig. 2. Distribution of lesions' equivalent diameters obtained from the FFDM database.

dynamic stopping criterion is implemented to terminate the contour evolution when it reaches the lesion boundary.

3.2 Feature extraction

In our study the computer-extracted features were grouped into three categories. The first category includes features quantifying margin sharpness, spiculation, gradient, contrast and shape of a lesion [7]. The second category includes texture features extracted from two regions, i.e. the lesion and the entire encompassing ROI, respectively. For each region, a 2D gray-level co-occurrence matrix (GLCM) was constructed, and texture features were extracted to characterize homogeneity, gray-level dependence, brightness, variation and randomness [8]. We developed an automatic neighborhood estimation method to determine the effective circumstance of the lesion.

The third group includes a distance feature calculated as the Euclidean distance from the nipple location to the center of the lesion. Since nipple markers, which present as bright markers on the mammograms, are commonly used in mamographic images, we developed a nipple identification method by locating those markers automatically. This method includes several processing steps, as illustrated in Fig. 3. Initially, gray-level thresholding is employed on the entire mammogram to extract the breast region from the external-to-breast background. Then, another gray-level threshold is applied to the breast region, yielding several nipple marker candidates. The breast skin boundary is obtained by subtracting a morphologically eroded [9] breast region from its original region.

To reduce the number of falsely identified nipple markers, area and circularity constraints are imposed on each candidates, and those candidates with area within a given range and circularity above a certain threshold are kept for the final step. The area range and circularity threshold were empirically determined with 10 randomly selected images in this study. The nipple marker is finally chosen as the one closest to the breast boundary.

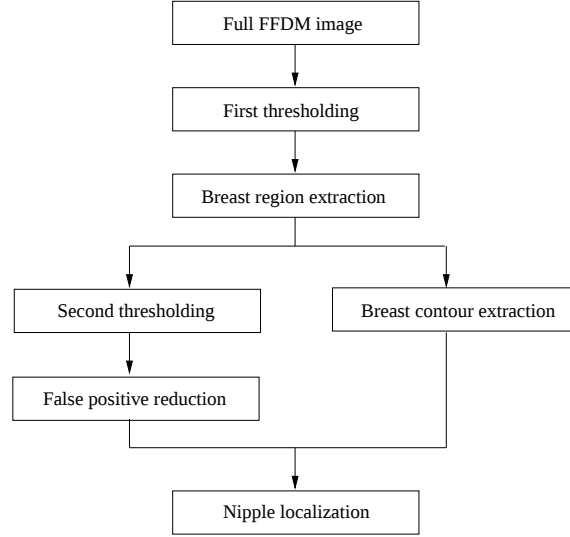


Fig. 3. Schematic diagram of the proposed automatic nipple identification algorithm.

3.3 Feature selection and classification

For each pair-wise set of features in representing the two different views, a Bayesian artificial neural network (BANN) classifier [10] was employed to merge each feature pair into a correspondence metric, which estimates the probability that the two images are of the same physical lesion, based on that specific feature. Next, an effective subset of correspondence metrics was selected via a linear stepwise feature selection [11] with a Wilks lambda criterion, and merged with another BANN to yield an overall estimate of the probability of correspondence.

3.4 Evaluation

The area under the receiver operating characteristic (ROC) curve (AUC) [12][13] was used as an index of performance of the individual features and the classifier outputs in the task of distinguishing between corresponding pairs from and

non-corresponding ones. The levels of statistical significance among individual features, and single feature versus merged multiple features, were calculated by ROCKIT software (version 1.1*b*).

4 Results

In a leave-one-out evaluation by lesion, the distance feature outperformed among all the other individual features, yielding an AUC of 0.78. The selected feature subset, which includes distance, ROI-based energy and ROI-based homogeneity, yielded an AUC of 0.88. The improvement by using multiple features was statistically significant compared to single feature performance ($p < 0.001$). ROC curves resulting from evaluation of the distance feature and the feature subset are shown in Fig. 4.

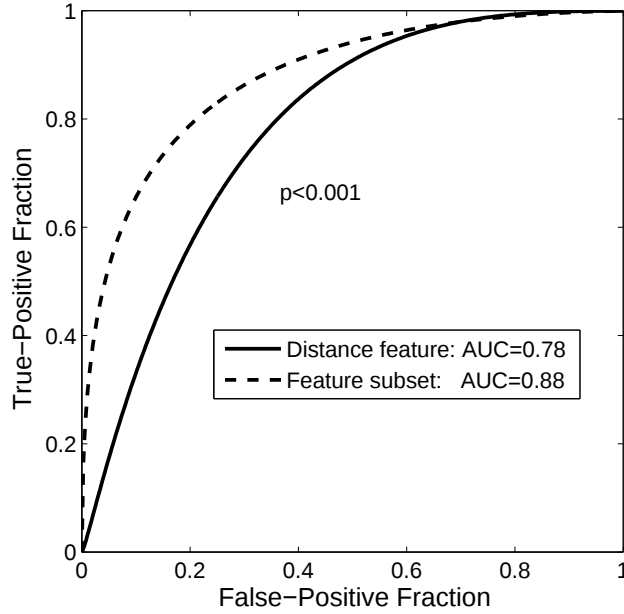


Fig. 4. ROC curves of leave-one-out evaluation by lesion for distance feature and the feature subset (distance feature, ROI-based energy and ROI-based homogeneity).

5 Discussion

We presented here a correlative feature analysis framework to assess the probability that a given pair of two mammographic images from CC and MLO view

is of the same physical lesion. Our results are promising into distinguishing between corresponding and non-corresponding lesion pairs. It is important to note that our method is feature-based, which employs a BANN classifier to estimate the relationship between computer-extracted features of a lesion in CC and MLO views. The supervised-learning manner cannot only make the relationship flexible to each feature, but also avoid the sophisticated geometrically-deformable models that are widely used in geometric breast registrations.

We are generalizing this framework and applying it to differentiate corresponding and non-corresponding pairs from multi-modality breast images, such as FFDM and breast MRI images.

Acknowledgment

This work was supported in part by US Army Breast Cancer Research Program (BCRP) Predoctoral Traineeship Award (W81XWH-06-1-0726), by United States Public Health Service (USPHS) Grant CA89452, and by a grant from the US Army Medical Research and Materiel Command grant (DAMD 98-1209), and by Cancer Center Support Grant (5-P30CA14599). M. L. Giger is a stockholder in R2 Technology, Inc (Sunnyvale, CA), a Hologic Company. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actually or potential significant financial interest with would reasonably appear to be directly and significantly affected by the research activities.

References

1. Huo, Z., Giger, M.L., Vyborny, C.J.: Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis. *IEEE Trans. Med. Imaging* **20** (2001) 1285–1292
2. Chan, H.P., Sahiner, B., Lam, K.L., Petrick, N., Helvie, M.A., Goodsitt, M.M., Adler, D.D.: Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces. *Med. Phys.* **25** (1998) 2007–2019
3. Liu, B., Metz, C.E., Jiang, Y.: An ROC comparison of four methods of combining information from multiple images of the same patient. *Med. Phys.* **31** (2004) 2552–2563
4. Yuan, Y., Giger, M.L., Li, H., Suzuki, K., Sennett, C.: A dual-stage method for lesion segmentation on digital mammograms. *Med. Phys.* **34** (2007) 4180–4193
5. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. Image Processing* **10** (2001) 266–277
6. Kupinski, M.A., Giger, M.L.: Automated seeded lesion segmentation on digital mammograms. *IEEE Trans. Med. Imaging* **17** (1998) 510–517
7. Huo, Z., Giger, M.L., Vyborny, C.J., Bick, U., Lu, P.: Analysis of spiculation in the computerized classification of mammographic masses. *Med. Phys.* **22** (1995) 1569–1579
8. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Trans. Syst. Man. Cybern.* **3** (1973) 610–621
9. Sonka, M., Hlavac, V., Boyle, R.: *Image processing, analysis, and machine vision*. PWS publishing, Pacific Grove, CA (1998)

10. Bishop, C.M.: Neural networks for pattern recognition. Oxford university, Oxford, U. K (1995)
11. Lachenbruch, P.A.: Discriminant analysis. Hafner, London, U. K (1975)
12. Metz, C.E.: ROC methology in radiologic imaging. Invest. Radiol. **21** (1986) 720–733
13. Metz, C.E., Herman, B.A., Shen, J.: Maximum likelihood estimation of receiver operating characteristic ROC curves from continously-distributed data. Stat. Med. **17** (1998) 1033–1053

Breast cancer classification with mammography and DCE-MRI

Yading Yuan^a, Maryellen L. Giger, Hui Li and Charlene Sennett

^aDepartment of Radiology, The University of Chicago, Chicago, IL USA 60637

ABSTRACT

Since different imaging modalities provide complementary information regarding the same lesion, combining information from different modalities may increase diagnostic accuracy. In this study, we investigated the use of computerized features of lesions imaged via both full-field digital mammography (FFDM) and dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) in the classification of breast lesions. Using a manually identified lesion location, i.e. a seed point on FFDM images or a ROI on DCE-MRI images, the computer automatically segmented mass lesions and extracted a set of features for each lesion. Linear stepwise feature selection was firstly performed on single modality, yielding one feature subset for each modality. Then, these selected features served as the input to another feature selection procedure when extracting useful information from both modalities. The selected features were merged by linear discriminant analysis (LDA) into a discriminant score. Receiver operating characteristic (ROC) analysis was used to evaluate the performance of the selected feature subset in the task of distinguishing between malignant and benign lesions. From a FFDM database with 321 lesions (167 malignant and 154 benign), and a DCE-MRI database including 181 lesions (97 malignant and 84 benign), we constructed a multi-modality dataset with 51 lesions (29 malignant and 22 benign). With leave-one-out-by-lesion evaluation on the multi-modality dataset, the mammography-only features yielded an area under the ROC curve (AUC) of 0.62 ± 0.08 and the DCE-MRI-only features yielded an AUC of 0.90 ± 0.05 . The combination of these two modalities, which included a spiculation feature from mammography and a kinetic feature from DCE-MRI, yielded an AUC of 0.94 ± 0.03 . The improvement of combining multi-modality information was statistically significant as compared to the use of mammography only ($p = 0.0001$). However, we failed to show the statistically significant improvement as compared to DCE-MRI, with the limited multi-modality dataset ($p = 0.22$).

Keywords: Breast cancer, mammography, DCE-MRI, computer-aided diagnosis

1. INTRODUCTION

Breast cancer is the most common malignancy in American women and the second most common cause of deaths from malignancy in the population (after lung cancer). According to the American Cancer Society, about 182,460 women in the United States will be found to have invasive breast cancer in 2008, and about 40,480 women will die from the disease this year.¹ Although there are currently limited methods for curing breast cancer, early detection by breast imaging plays an important role in reducing the mortality. Between the years 1991 and 2003, there has been a steady decrease in the annual death rate from female breast cancer, from 32.69 to 25.19 (per 100,000 population). This decrease largely reflects improvements in early detection and treatment.

Although mammography has achieved significant success and reduces the mortality from breast cancer by 15% – 35%,^{2,3} it is far from perfect: about 15% – 20% of cancers are missed, and 75% of lesions detected by mammography are benign, resulting in many unnecessary medical procedures, including biopsies.⁴ Consequently, some complementary imaging modalities, such as breast DCE-MRI and breast sonography, are being investigated to improve the sensitivity and specificity of breast cancer detection and diagnosis.

While the results of computer-aided diagnosis (CAD) systems for single imaging modality are encouraging, merging information across different modalities is recently attracting more attention. Because different imaging modalities provide complementary information regarding lesions, combining information from two or more modalities may increase the diagnostic accuracy. Several investigations have been conducted to combine information from mammography and sonography to improve the diagnosis of breast cancer.^{5–8} Their results showed that the performance of cancer classification was improved as compared to each individual modality. However, to the

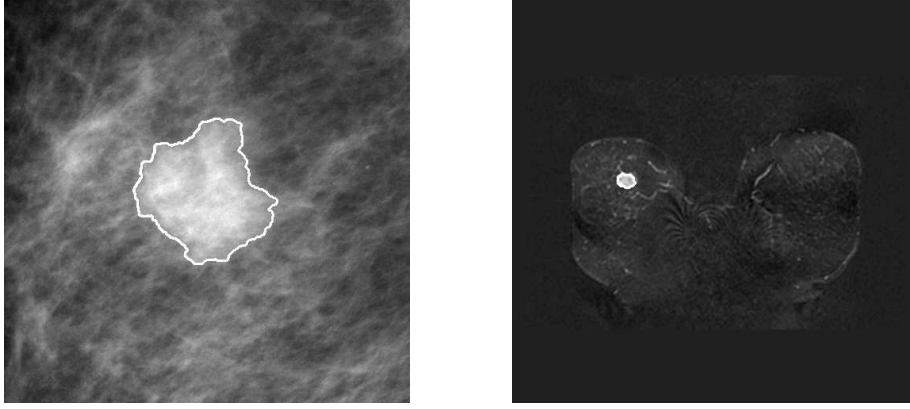


Figure 1. Example: a malignant lesion imaged by both mammography (left) and DCE-MRI (right). The solid lines in both images are the segmentation results from the dual-stage method⁹ and the fuzzy c-means based method,¹⁰ respectively.

best of our knowledge, there is limited research on combining information from mammography and DCE-MRI. Thus, in this study, we investigated the use of computerized features of lesions imaged via both mammography and DCE-MRI (Fig. 1) in the classification of breast lesions.

2. METHODS

Our computerized scheme consists of several steps: 1) automatic extraction of lesions on each modality images via computerized segmentation methods; 2) automatic extraction of various features (mathematical descriptors) from lesions; and 3) merging of information from different modalities and to estimate the probability of malignancy.

2.1. Lesion segmentation

For FFDM images, a dual-stage method was employed to automatically extract lesions from the normal breast tissue.⁹ In this method, a radial gradient index (RGI) based segmentation¹¹ is used to yield an initial contour close to the lesion boundary location in a computationally efficient manner. This initial segmentation also provides a base to identify the effective circumstance of the lesion via an automatic background estimation method. Then a region-based active contour model¹² is utilized to evolve the contour further to the lesion boundary. The active contour model relies on an intrinsic property of image segmentation that each segmented region should be as homogeneous as possible for an image formed by two regions. Instead of empirically determined criteria such as fixed iteration times, a dynamic stopping criterion is implemented to terminate the contour evolution when it reaches the lesion boundary.

For DCE-MRI images, a fuzzy c-means (FCM) clustering-based method was used for the segmentation of lesions in 3D space.¹⁰ This scheme includes six steps. An ROI is first selected by a human operator; then the post-contrast ROI series are enhanced by dividing the pixel value at each voxel by the value at the corresponding pre-contrast voxel. After the FCM clustering method is applied to partition the whole ROI into lesion and non-lesion parts, the lesion membership map is binarized with an empirically determined threshold. Then a 3D connected-component labeling operation is performed to reduce the false-positive voxels. Finally, a hole-filling operation is performed yielding the final segmented lesion.

2.2. Computerized feature extraction

For FFDM images, fifteen features were extracted from the segmented lesions, which characterize spiculation, margin sharpness, shape and density of the lesions.¹³ In our FFDM database, the number of mammograms available for each physical lesion was different, ranging from 1 to 13. Thus, for each feature, we determined a representative value of a lesion as its average value over all the mammograms of that particular lesion.

For DCE-MRI images, another 15 features were extracted from the lesions in 3D space, which included spiculation features, shape features, enhancement-kinetics-based features and enhancement-variance dynamics

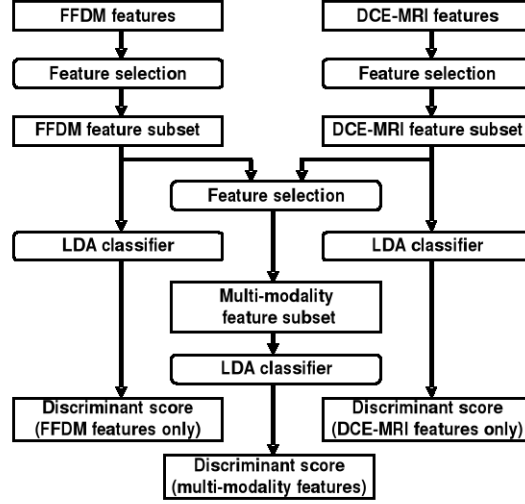


Figure 2. Flow chart of feature selection and classification. .

features.¹⁴ The characteristic kinetic curves were generated from the “most-enhanced” regions, which were automatically identified by a FCM clustering method.¹⁵

2.3. Feature selection and classification

In this study, the feature selection procedure included two steps, as shown in Fig. 2. Linear stepwise feature selection,¹⁶ with Wilke’s lambda criterion, was firstly performed on single modalities, yielding one feature subset for each modality. Then, these selected features served as the input to another feature selection procedure when extracting useful information from both modalities.

Linear discriminant analysis (LDA)¹⁷ was employed to merge the selected features to a single discriminant score that is related to the estimated likelihood of malignancy.

2.4. Performance evaluation and statistical analysis

The performance of discriminant scores from both single modality and multiple modalities, in the task of differentiating malignant lesions from benign ones, was evaluated using receiver operating characteristic (ROC) analysis,^{18,19} with the area under the ROC curve (AUC) as a figure of merit. The level of statistical significance was calculated by ROCKIT software (version 1.1b). A leave-one-out-by-lesion evaluation was used to evaluate the performance of each classifier.

3. RESULTS

In this preliminary study, we used a FFDM database including 321 lesions (167 malignant and 154 benign), and a DCE-MRI database including 181 lesions (97 malignant and 84 benign). All the lesions are biopsy-proven. From these two databases, we constructed a multi-modality dataset of 51 lesions (29 malignant and 22 benign). Mammograms and DCE-MRI images are available for these lesions.

With the entire FFDM database, ROI-based normalized radial gradient (NRG), lesion-based NRG, and gradient texture were selected as an effective feature subset. The leave-one-out-by-lesion evaluation using LDA to merge the selected features yielded an AUC value of 0.62 ± 0.08 on the multi-modality dataset.

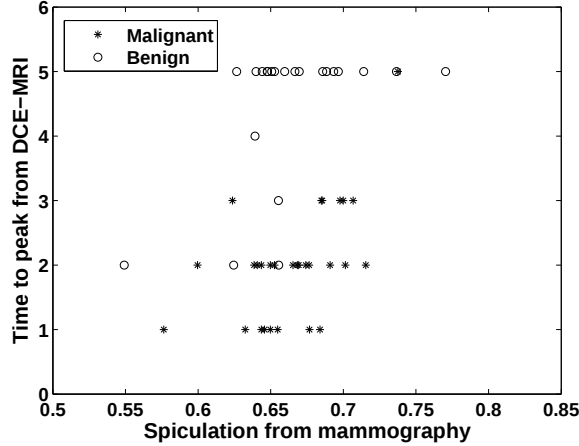


Figure 3. Scatter plot of the spiculation feature from mammography versus the peak location of the kinetic curve from DCE-MRI.

With the entire DCE-MRI database, margin sharpness, irregularity, peak location of the enhancement kinetics, and enhancement-variance uptake rate were selected as an effective feature subset. The leave-one-out-by-lesion evaluation using LDA yielded an AUC value of 0.90 ± 0.05 on the multi-modality dataset.

With the multi-modality dataset, the lesion-based NRG from mammography and the peak location of the enhancement kinetics from DCE-MRI were selected from the above 7 features. Figure 3 shows the scatter plot of these two features on the multi-modality dataset. The leave-one-out-by-lesion evaluation using LDA yielded an AUC of 0.94 ± 0.03 .

The improvement of combining multi-modality information was statistically significant as compared to the use of mammography only ($p = 0.0001$). However, although the performance of combining multi-modality information was better than the use of DCE-MRI only, we failed to show the statistically significant improvement, with the limited multi-modality dataset ($p = 0.22$).

4. CONCLUSION

In the proposed study, we investigated the performance of a computerized classification scheme with computer-extracted features based on mammography alone, DCE-MRI alone, and the combination of these two modalities. In mammography, spiculation and texture features were shown to be effective for breast cancer classification. In DCE-MRI images, margin sharpness, lesion shape and kinetic features were salient. In our previous studies, spiculation and kinetic features have been justified as the best features when distinguishing malignant and benign lesions for mammography and DCE-MRI, respectively. Our feature selection method correctly captured these two features when combining information across different modalities.

Our pilot results showed that combining information from multiple modalities performed better than the single modality in the task of distinguishing between malignant and benign lesions. We are currently expanding our multi-modality database and will evaluate the performance of combining information from multi-modalities on the larger database.

ACKNOWLEDGMENT

This work was supported in part by US Army Breast Cancer Research Program (BCRP) Predoctoral Traineeship Award (W81XWH-06-1-0726), by United States Public Health Service (USPHS) Grant CA89452 and P50-CA125183, and by DOE grant DE-FG02-08ER64578. M. L. Giger is a shareholder in R2 Technology, Inc (Sunnyvale, CA), a Hologic Company. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actually or potential significant financial interest with would reasonably appear to be directly and significantly affected by the research activities.

REFERENCES

1. A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, and M. J. Thun, "Cancer statistics, 2008," *CA Cancer J. Clin.* **58**, pp. 71–96, 2008.
2. L. L. Humphrey, M. Helfand, B. Chan, and S. H. Woolf, "Breast cancer screening: A summary of the evidence for the U.S. preventive services task force," *Ann Intern Med.* **137**, pp. 347–360, 2002.
3. Institute of Medicine, *Saving women's lives: integration and innovation: a framework for progress in early detection and diagnosis of breast cancer*, National Academic Press, Washington D.C., 2005.
4. S. H. Heywang-Kobrunner, D. D. Dershaw, and I. Schreer, *Diagnostic breast imaging: mammography, sonography, magnetic resonance imaging, and interventional procedures*, Thieme Medical Publisher, New York, New York, 2 ed., 2001.
5. K. Drukker, K. Horsch, and M. L. Giger, "Multimodality computerized diagnosis of breast lesions using mammography and sonography," *Acad. Radiol.* **12**, pp. 970–979, 2005.
6. J. L. Jesneck, J. Y. Lo, and J. A. Baker, "Breast mass lesions: computer-aided diagnosis methods with mammographic and sonographic descriptors," *Radiology* **244**, pp. 390–398, 2007.
7. K. Horsch, M. L. Giger, C. J. Vyborny, L. Lan, E. B. Mendelson, and R. E. Hendrick, "Multi-modality computer-aided diagnosis for the classification of breast lesions: observer study results on an independent clinical dataset," *Radiology* **240**, pp. 357–368, 2006.
8. K. Horsch, M. L. Giger, and C. E. Metz, "Potential effect of different radiologist reporting methods on studies showing benefit of CAD," *Acad Radiol* **15**, pp. 139–152, 2008.
9. Y. Yuan, M. L. Giger, H. Li, K. Suzuki, and C. Sennett, "A dual-stage method for lesion segmentation on digital mammograms," *Med. Phys.* **34**, pp. 4180–4193, 2007.
10. W. Chen, M. L. Giger, and U. Bick, "A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images," *Acad. Radiol.* **13**, pp. 63–72, 2006.
11. M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Trans. Med. Imaging* **17**, pp. 510–517, 1998.
12. T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Processing* **10**, pp. 266–277, 2001.
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. W. Chen, M. L. Giger, L. Lan, and U. Bick, "Computerized interpretation of breast MRI: investigation of enhancement-variance dynamics," *Med. Phys.* **31**, pp. 1076–1082, 2004.
15. W. Chen, M. L. Giger, U. Bick, and G. M. Newstead, "Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI," *Med. Phys.* **33**, pp. 2878–2887, 2006.
16. P. A. Lachenbruch, *Discriminant analysis*, Hafner, London, U. K, 1975.
17. R. Duda, P. Hart, and D. Stork, *Pattern classification*, Wiley-Interscience, Sons, USA, 2nd ed., 2001.
18. C. E. Metz, "ROC methology in radiologic imaging," *Invest. Radiol.* **21**, pp. 720–733, 1986.
19. C. E. Metz, B. A. Herman, and J. Shen, "Maximum likelihood estimation of receiver operating characteristic ROC curves from continously-distributed data," *Stat. Med.* **17**, pp. 1033–1053, 1998.