Multilevel learning-based segmentation of ill-defined and spiculated masses in mammograms

Yimo Tao^{a)}

Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, Virginia 22203 and Department of Radiology, ISIS Center, Georgetown University Medical Center, Washington, DC 20007

Shih-Chung B. Lo

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

Matthew T. Freedman

Department of Oncology, Georgetown University Medical Center, Washington, DC 20007

Erini Makariou

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

Jianhua Xuan

Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, Virginia 22203

(Received 12 March 2010; revised 10 August 2010; accepted for publication 27 August 2010; published 28 October 2010)

Purpose: A learning-based approach integrating the use of pixel-level statistical modeling and spiculation detection is presented for the segmentation of mammographic masses with ill-defined margins and spiculations.

Methods: The algorithm involves a multiphase pixel-level classification, using a comprehensive group of features computed from regional intensity, shape, and textures, to generate a mass-conditional probability map (PM). Then, the mass candidate, along with the background clutters consisting of breast fibroglandular and other nonmass tissues, is extracted from the PM by integrating the prior knowledge of shape and location of masses. A multiscale steerable ridge detection algorithm is employed to detect spiculations. Finally, all the object-level findings, including mass candidate, detected spiculations, and clutters, along with the PM, are integrated by graph cuts to generate the final segmentation mask.

Results: The method was tested on 54 masses (51 malignant and 3 benign), all with ill-defined margins and irregular shape or spiculations. The ground truth delineations were provided by five experienced radiologists. Area overlapping ratio of $0.689 (\pm 0.160)$ and $0.540 (\pm 0.164)$ were obtained for segmenting entire mass and margin portion only, respectively. Williams index of area and contour based measurements indicated that the segmentation results of the algorithm agreed well with the radiologists' delineation.

Conclusions: The proposed approach could closely delineate the mass body. Most importantly, it is capable of including mass margin and its spicule extensions which are considered as key features for breast lesion analyses. © 2010 American Association of Physicists in Medicine.

[DOI: [10.1118/1.3490477](https://doi.org/10.1118/1.3490477)]

Key words: breast cancer, mammography, breast masses, image segmentation

I. INTRODUCTION

Clinically, shape and margin characteristics of a mammographic mass are regarded as the two most important features for breast cancer diagnosis. It is known that malignant masses usually have ill-defined margins and irregular shape and/or spiculations; on the other hand, benign masses usually have well-defined margins. More specifically, a spiculated mass consists of a central mass body with “extensions,” hence the resulting stellate shape. From medical image analysis perspectives, image features (e.g., texture and intensity patterns) associated with the extended regions and ill-defined borders are important information for mass analyses.

Therefore, a segmentation algorithm that is tailored to delineate the mass body and periphery, including its irregular details or spiculations, is technically desirable.

Many algorithms have been proposed for automatic mammographic mass segmentation. Region growing is one of the most often used methods. Pohlman *et al.*¹ developed an adaptive region growing method, whose pixel aggregation criterion was determined from calculations made in 5×5 windows surrounding the pixel of interest. Kupinski and Giger² proposed two region growing approaches based on the radial gradient index and a probabilistic model. Kinnard *et al.*³ extended the probabilistic model based method by further analyzing the steepest change of cost functions. Their

FIG. 1. Flowchart of the proposed MLS technique. Major processes of the method as shown in rounded rectangles include pixel-level soft segmentation, object-level detection, and spiculation detection.

method was found to be able to further include some of the ill-defined mass boundaries. Besides region growing, many other techniques have also been investigated. Te Brake and Karssemeijer⁴ proposed a discrete dynamic contour model, which began as a set of vertices connected by edges (initial contour) and grew the subject according to internal and external forces. Li *et al.*⁵ developed a method that employed *k*-means classification to categorize pixels as belonging to the region of interest (ROI) or background. Sahiner *et al.*⁶ proposed a method consisting of segmentation initialization by pixel-intensity based clustering followed by region growing to improve boundary shape; then the initial result was further augmented by an active contour algorithm for shape refinement. Recently, Shi *et al.*⁷ proposed a level set approach for mass segmentation. The segmented masks were found to be able to improve the performance for discriminating benign and malignant masses in comparison with their previous work. However, the energy function defined in the level set approach was based on the image gradient information, which would be noisy for masses with ill-defined margins and spiculations. This might reduce the method's reliability and robustness in segmenting ill-defined masses. Domínguez and Nandi⁸ proposed a segmentation method based on contour tracing using dynamic programming. Although the method worked well for masses with circumscribed margins, it had difficulties in segmenting masses with less distinct contours. A review of mammographic mass segmentation algorithms could be found in the literature.⁹

While encouraging results have been obtained in the aforementioned works, fully automatic segmentation of

mammographic masses remains challenging, especially for those masses with ill-defined margins and spiculations. In this work, a multilevel learning-based segmentation (MLS) technique is proposed to specifically tackle the challenge of mass margins and associated extensions. The advantage of the MLS approach reflects in the following aspects: First, it is capable of accurately delineating the irregular shape of mass, including a substantial amount of margin portion and spicule extensions, which are considered as key features for breast lesion analyses.^{10,11} Second, the approach is technically robust in segmenting masses in the presence of dense glandular tissues and pectoral muscles. Specifically, it minimizes the risk of oversegmentation, which is one major shortcoming in traditional region growing based approaches.

II. MATERIALS AND METHODS

Instead of searching for a mass in mammogram, this study focused on developing algorithms that could accurately segment a suspicious mass. The proposed segmentation method would take each ROI through three major processes (shown as rounded rectangles in Fig. 1): (1) Pixel-level soft segmentation/labeling, (2) object-level mass and clutters detection, and (3) spiculation detection. The pixel-level mass and nonmass class segmentation and labeling works as follows: Given a ROI, the system would label each pixel with its probability associated with the mass configuration statistical model trained through a supervised learning. A pixel-level probability map (PM) for the whole image is thus obtained. After this, the object-level image classification/

FIG. 2. The distribution of the mass statistics of shape, margin, density, sizes, and subtlety ranking within the database.

detection module takes the PM along with *prior* information, namely, shape, size, and spatial distribution, to identify region of mass and clutters. In order to include irregular shape and spiculations, a spiculation detection module using a multiscale ridge detection algorithm is then employed to produce a binary image of spicules (named as “spiculation map”). Finally, graph cuts¹² algorithm is used to integrate the PM and all the object-level findings (i.e., mass region, clutters, and spiculation map) to produce the final segmentation.

II.A. Image database

In this study, we used a data set containing a total of 54 (51 malignant and 3 benign) ROIs. The spatial resolution of the image was sampled to $100\ \mu\text{m} \times 100\ \mu\text{m}$. The mass shape, margin, and density were measured by a senior radiologist according to the Breast Imaging Reporting and Data System (BI-RADS) atlas¹³ as shown in Figs. 2(a)–2(c). Figure 2(d) shows the size of these masses ranged from 5 to 50 mm. The size of a mass was measured as the longest dimension of the mass. The subtlety of these masses as shown in Fig. 2(e) were ranked on a scale from 1 (the easiest) to 10 (the most difficult) for determining the malignancy of the case. In this study, we collected a total of five radiologists’ delineation for each mass in the database. With the multiradiologists’ segmentation results, the final ground truth for each mass was then formed by setting the pixel as the fore-

ground (i.e., mass) where at least three radiologists reached a consensus. The remaining pixels were regarded as the background (i.e., nonmass).

II.B. Segmentation method

II.B.1. Pixel-level soft segmentation

In the first step, the segmentation of mammographic mass is processed by a pixel-level soft segmentation/labeling approach as illustrated in Fig. 3. Specifically, the probabilistic distributions of the mass appearance are learned through image subpatch level modeling (ISLM). Regional features of a subpatch [i.e., a subregion of interest (sROI)] centered at each pixel position $p(p \in \mathcal{R}^2)$ were computed across the image. These features include (1) texture features obtained from co-occurrence matrix¹⁴ and 2D local binary pattern computation,¹⁵ (2) shape features of vesselness and blobness⁷ computed through the eigenanalysis of Hessian matrix, and (3) a group of gray-level statistics features. In this study, a total of 30 dimensional features (denoted as x_p) were calculated on the preprocessed image for each scanned sROI. The image preprocessing steps were composed of background correction¹⁶ followed by morphological smoothing. Based on our ground truth maps of mass and nonmass regions, computed feature vectors were split into positive (i.e., mass) and negative (i.e., nonmass) classes and then fed into an offline classifier learning process.

FIG. 3. Flowchart of ISLM. For the training process, regional features of training subpatches are used for the pixel-level mass/nonmass classifier learning. For the testing process, the learned classifier is used to generate a mass possibility map for the whole testing image. A high intensity value in the mass probability map represents a high likelihood to be mass tissue. It is seen that very high intensity covers mass body and most of mass boundary with edge.

For the pixel-level classification task, it would take a large number ($\geq 100\,000$) of the training samples, i.e., feature samples computed from cropped subpatches. Linear discriminant analysis (LDA), along with Gaussian mixture model (GMM), was employed in this process for its strength in scalability as a classifier. More specifically, for each of the binary mass and nonmass class, LDA was first exploited to further project the original feature vector x_p into a vector with a lower dimension (of d). The distribution of samples in the projected subspace was then modeled using GMM consisting of k -Gaussian distributions as

$$f(x|\phi) = \sum_{i=1}^k \alpha_i \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp \left\{ -\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right\}, \quad (1)$$

where the parameter set $\phi = \{\alpha_i, \mu_i, \Sigma_i\}_{i=1}^k$ consists of the prior α_i , the mean μ_i , and the covariance matrix Σ_i for the i th distribution. Then, the expectation-minimization algorithm¹⁸ with multiple random initialization trials was used to determine ϕ . Note that a model selection was performed to choose the number k of Gaussian functions using Bayesian information criterion and the value of k was set at 3 and 4 for positive and negative classes, respectively.

As there were several different types of nonmass tissues in the mammogram, such as glandular tissues, ducts, etc., a multiphase classification approach was adopted to increase the classification accuracy.¹⁷ It started with positive class output PM from single phase and treated it as a new image. Next, another round of pixel-level feature extraction/selection and LDA-GMM training process was conducted using both the original image and the output image from the previous phase(s). All these intensity-texture-shape features, in the joint image and PM domain, were used to train a new classifier. An illustrative example of generated PM is shown

in Fig. 4. It is seen that the generated PM is able to enhance the mass tissue structure, while suppressing false responses from normal breast tissues.

II.B.2. Object-level labeling and detection

At this stage, the technical objective is to determine the mass region and other clutters in the intermediate PM output [denoted as $\text{prob}(p)$]. To suppress spurious responses, the multiscale blobness filtering with a larger smoothing kernel (than the pixel-level feature extraction step) was used to capture the mass shape. It was applied on each pixel to obtain a blobness likelihood map denoted as $\text{blm}(p)$. Then, a shape-prior refined probability map $\text{sprob}(p)$ was obtained as

$$\text{sprob}(p) = \text{blm}(p) \times \text{prob}(p). \quad (2)$$

Otsu thresholding method¹⁹ was used for discrete quantization of $\text{sprob}(p)$ to obtain the potential mass pixels with $\text{sprob}(p) > V_{\text{threshold}}$. Connected component analysis was then employed to obtain several disjointed regions (DRs) $C_1\{p\}, C_2\{p\}, \dots, C_n\{p\}$ on the binarized image. For each DR, we compute a fitness score to determine its likelihood of being a mass as

FIG. 4. Results from the pixel-level classification. (a) The original ROI, (b) the mass probability map produced by the two-phase pixel-level classifiers, (c) the segmentation ground truth generated from multiple radiologists (with majority of five radiologists' consensus), and (d) the ground truth contour superimposed on the original ROI.

FIG. 5. Example of spiculation detection on a synthetic image. (a) Synthetic linear structures superimposed on a fatty mammographic background, (b) the ridge detector response, (c) the extracted backbone of the spiculation, and (d) the superimposed backbone on the original image. Reprinted with permission from R. Zwiggelaar, S. M. Astley, C. R. M. Boggis, and C. J. Taylor, IEEE Trans. Med. Imaging 23, 1077–1086 (2004). Copyright: ©2004 IEEE.

$$F_i = \sum_{p \in c_i\{p\}} G(p|\mu, \Sigma) \times \text{sprob}(p), \quad i = 1, 2, \dots, n, \quad (3)$$

where $G(p|\mu, \Sigma)$ is a 2D multivariate normal distribution representing a spatial prior that the mass is near the center of ROI. The DR possessing the maximum fitness score was selected as the mass candidate. The remaining DRs were regarded as clutters.

II.B.3. Spiculation detection

It is known that malignant masses in mammograms are often characterized by a radial pattern of linear structures (i.e., spicules) and irregular boundaries.^{20,21} Detecting spiculation as a part of mass region is thus essential for further computer analysis. In this task, a steerable ridge detection approach²² was employed and it was further generalized into a multiscale analysis framework to detect the presence of spiculations.

In this study, an M th order ($M=4$) template composed of Gaussian kernels and their derivatives was employed to be the ridge detector as

$$h(x, y) = \sum_{k=1}^M \sum_{i=0}^k \alpha_{k,i} \underbrace{\left(\frac{\partial^{k-i}}{\partial x^{k-i}} \frac{\partial^i}{\partial y^i} g(x, y) \right)}_{g_{k,i}(x, y)} \quad (4)$$

where $g(x, y)$ is a 2D Gaussian function and $\alpha_{k,i}$ represents the weight coefficient for the kernel of $g_{k,i}(x, y)$. The ridge detection procedure is formulated as a rotated template matching. It involves the computation of inner products with the shifted and rotated versions of the 2D template $h(x, y)$ at every pixel in the image of $I(x): x=(x, y)$. A high magnitude of the inner product indicates the presence of the feature and the angle of the corresponding template gives the orientation. Mathematically, the estimation algorithm is formulated as

$$\theta^*(x) = \arg \max_{\theta} (I(x) \cdot h_{\theta}(x)), \quad (5)$$

$$r^*(x) = I(x) \cdot h_{\theta^*}(x), \quad (6)$$

where $r^*(x)$ is the magnitude of the feature, $\theta^*(x)$ is its orientation at the pixel position x , $h_{\theta}(x)$ is the rotated template with a degree of θ , and $\cdot$ is the inner product operator. Due to the property of *steerable filters* defined in Eq. (4), we could cut down on the computational load in Eqs. (5) and (6). Specifically, the inner product of a signal $I(x)$ with an $h_{\theta}(x)$ can be expressed as

$$I(x) \cdot h_{\theta}(x) = \sum_{k=1}^M \sum_{i=0}^k b_{k,i}(\theta) I_{k,i}(x), \quad (7)$$

where $b_{k,i}(\theta)$ are orientation-dependent weights computed using trigonometric polynomials of θ (see the literature²² for details) and the functions $I_{k,i}(x)$ are the inner products of the signal $I(x)$ with unrotated kernels of $g_{k,i}(x)$

$$I_{k,i}(x) = I(x) \cdot g_{k,i}(x). \quad (8)$$

To determine the optimal detector $h(x, y)$ in Eq. (4), which was equivalent to searching for the optimal weight combinations of $\alpha_{k,i}$, a functional optimization method following the Canny-like criteria was used.²²

Figure 5 shows an example of the ridge detection result on a synthetic image. In order to obtain spicules at different widths, the scale of a ridge detector was progressively increased. An estimate of the ridge scale at each pixel was obtained by normalizing the line strength obtained at each scale and choosing the scale that gives the largest response. The line strength and orientation along with the chosen scale were taken as a representative of the pixel in the image. In this study, three scales ($\sigma=3, 5, 7$ for the standard deviations of Gaussian kernels) of detectors were applied on the original image. Based on the line strength and orientation images, nonmaximal-suppression, i.e., thresholding with hysteresis, was used to extract the backbone of each structure, followed by thinning to obtain ridges in the image.

To reduce nonspicule false positives generated from other linear structures (e.g., ligaments, ducts, etc.), several postfil-

FIG. 6. Results of multiscale spiculation detection, (a) The original ROI, (b) the multiscale line-strength image, (c) the detected spicules, (d) the detected spicules superimposed on the ROI, and (e) the final spiculation map.

FIG. 7. Example outputs of the multiphase pixel-level classification. (a) The original ROI, (b) the PM generated by the first phase classifier, (c) the PM generated by the two-phase classifiers, and (d) the ground truth provided by the radiologists. Note that the noisy responses on the upper left corner of (b) and (c) come from the pectoral muscles, which have been clearly suppressed by the two-phase classifier with low classification values produced.

tering rules based on geometric relationships (e.g., position and direction) between the spicule candidates and the extracted mass candidate were applied. The spiculation region was then obtained by applying region growing, using the filtered spicules as seed on the multiscale line-strength image. The final output of the spiculation detection module is a binary mask (named as “spiculation map”), where the foreground object(s) represents the detected potential spicule pixels. Figure 6 shows an example of the detected spicules and the spiculation map.

At the last, graph cuts¹² was employed to integrate all the object-level findings, including mass candidate, clutters, and the spiculation map, along with the pixel-level PM, to produce the final segmentation mask. A very interesting property of graph cuts is that it can easily incorporate topological constraints by setting appropriate weights on edges in the defined graph model of an image.²⁴ These constraints indicate some image pixels *a priori* known to be a part of the foreground (mass candidate and spiculation) or background (normal breast tissues and clutters). The optimal segmentation mask could be obtained by finding the minimum cost cut on the graph model. Note that a simple classification logic to determine a mass as spiculated or nonspiculated was also implemented using the area ratio between the spiculation map and the mass candidate. If the mass was determined to be nonspiculated, detected spiculations would be ignored.

III. EXPERIMENTS AND RESULTS

III.A. Pixel-scale classification results

For the pixel-level classification, a two-phase classification scheme was adopted for the experiment. A threefold cross-validation experiment was used to test the performance of the method. Specifically, the image data set was randomly but evenly divided into three parts. For each round of training/testing procedure, we combined two parts out of three as the training image set for ISLM and the remaining part as an *unseen* testing set. An accuracy of 87.81% was obtained for the pixel-scale classification task. The effect of the multiphase classification approach is illustrated by the enhancement results shown in Fig. 7.

III.B. Segmentation results

In this work, we compared the MLS approach to a region growing segmentation method based on maximum likelihood function analysis described by Kinnard *et al.*³ (named as “MLFA”) and an active contour method using level set described by Shi *et al.*⁷ (named as “LS”). To measure the level

of agreement between radiologists’ delineation and the segmented masses, three measurements including the area overlapping ratio (AOR), the average minimum distance (AMINDIST), and the Hausdorff distance (HSDIST) were adopted. Detailed mathematical definitions of these measurements could be found in the literature.⁶ The summary statistics of the segmentation results are shown in Table I. The box and whisker plots of the corresponding distributions of these segmentation measurements are shown in Fig. 8.

The ROIs shown in Fig. 9 demonstrate the segmentation results of various methods. It is seen that the segmented contours by the MLS approach are capable of closely delineating mass body contours and they include a sufficient amount of mass margin portion. The approach is also technically robust in segmenting masses in the presence of ill-defined texture patterns and unsmooth intensity changes inside masses. The MLFA method has the notorious “leaking” problem for region growing approaches, where the segmented contours leak into nonmass tissue areas, e.g., dense glandular tissues and pectoral muscles, as shown in the second and third examples in Fig. 9. When comparing the MLS and LS methods, it is seen that the MLS approach performed better in capturing the irregular shapes of these masses and including substantial amount of spiculations. The disadvantages of the LS method may be due to, first, it has the constraint for boundary curvature smoothness and second, it is dependent on the gradient information, which is ill-defined on the margin areas of these masses. Hence, the LS method seems unsuitable for segmenting mass boundary with ill-defined margin or irregular shape.

These segmentation results show that the MLS produces the best segmentation contours that match to the radiologists’ manual segmentation. The difference of segmentation results between MLS and other methods were found to be statistically significant, according to the two-sided Wilcoxon test with confidence interval (CI) at 0.95, for all measurements.

TABLE I. Measurements of segmentation accuracy with the three proposed methods. Note that AMINDIST and HSDIST are distance based measurements. The high distance value means low similarity between computer segmentation and radiologists’ delineation. The reciprocal of the distance value is used as the segmentation similarity measurement for testing the multiobserver agreement in Sec. III C.

	MLS	MLFA	LS
AOR	0.689 (± 0.160)	0.544 (± 0.196)	0.588 (± 0.128)
AMINDIST	13.94 (± 11.05)	23.30 (± 17.79)	20.02 (± 11.38)
HSDIST	72.13 (± 33.13)	87.61 (± 42.04)	81.88 (± 32.67)

FIG. 8. The box and whisker plots of the distribution of the segmentation measurements. (a) Box plots of AOR, (b) box plots of AMINDIST, and (c) box plots of HSDIST. The vertical lines of the boxes correspond to the lower, median, and upper quartile. Each cross represents an outlier. The vertical dashed line in (a) corresponds to the threshold for a good segmentation with AOR=0.7, which is adopted in the literature (Ref. 8).

III.C. Multiobserver agreement

As a part of evaluation, Williams index (WI)^{25,26} was employed to measure the *difference* between the consistency of the segmentation results with the radiologists and the consistency within radiologists themselves. WI is a statistical measure of agreement between multiple raters (a total of n) defined as follows:

$$WI_i = \left(\frac{n-2}{2} \right) \frac{\sum_{j=1, i \neq j}^n s_{ij}}{\sum_{j=1, j \neq i}^{n-1} \sum_{k>j}^n s_{jk}}, \quad (9)$$

where s_{ij} is a measure of similarity or agreement between raters i and j . We used AOR, the reciprocal of AMINDIST, and the reciprocal of HSDIST as the similarity measures. As studied in the literature,²⁶ if the upper limit of the CI of WI is greater than the value 1, it can be concluded that the measurement data are consistent with the hypothesis that the individual observer agrees with the group at least as well as the group members agree with each other (i.e., the individual observer is a reliable member of the group). The CI is estimated using a jack-knife scheme. The WIs with CI at 0.95 for the three segmentation measurements are shown in Fig.

10, with AOR of 1.002 (± 0.010), AMINDIST of 0.975 (± 0.047), and HSDIST of 0.995 (± 0.029).

III.D. Margin segmentation results

The performance of the proposed algorithm on segmenting only the margin portion was also evaluated. Here the margin is defined as the remaining foreground pixels by subtracting the mass core region from the complete ground truth. The mass core region was obtained through boundary smoothing via a rotation structure element algorithm²⁷ followed by morphological erosion. The margin area overlapping ratio (MAOR) of the segmented masses with the ground truth margin was then computed. MAOR for various methods are summarized in Table II. The box and whisker plots of the distribution of the segmentation results are shown in Fig. 11(a). We also measured the WI of MAOR between the algorithm and multiple radiologists. The result is shown in Fig. 11(b), with a value of 0.9815 (± 0.021). It is seen that the proposed approach well agreed with multiple radiologists in segmenting mass margin portion.

IV. DISCUSSION

It is seen that the computer segmentation and the ground truth show significant disagreement in some challenging

FIG. 9. Segmentation results. From left to right, they are the original ROI, segmentation result of MLFA, LS, MLS, manual segmentation, and the contours superimposed on the original image of the MLS approach (white contour) and radiologists' manual segmentation (black contour). The BI-RADS descriptors of a mass in margin and shape are also shown under each original image in the first column.

cases. Two example cases are illustrated in Fig. 12. Figure 12(a) shows that computer segmentation achieves less than 0.50 AOR comparing to the ground truth on a low contrast mammogram. This is due to that there is no clear margin between the mass and high density background. Even in this case, the segmentation may be acceptable for some applications, such as the detection of masses. Figure 12(b) shows a mass region overlapping with ducts. It is seen that ducts are extracted as part of segmentation by the spiculation detection module. It is expected that by using an advanced classification based method,²³ high intensity ducts would be excluded and thus enhance the overall segmentation performance for these cases.

Taking $\text{AOR} \geq 0.7$ as the criteria for “good” match of segmentation, which was a threshold used in the literature,⁸ the median of AOR of results generated by the MLS approach was 0.735. The thresholding result indicated 32 out of 54 masses achieving good segmentation. It should be mentioned that this study specifically focused on segmentation of ill-defined and spiculated masses. Segmentation of circum-

scribed masses has been studied by many investigators in the past. One can establish an automatic segmentation scheme using a high-level decision module to determine the applicability of specific algorithms to cope with various types of masses.

V. CONCLUSIONS

A multilevel learning-based segmentation (i.e., MLS) approach is proposed for the mammographic mass segmentation. This framework consists of pixelwise ISLM, object-level mass and clutters extraction, and multiscale spiculation detection. The approach was validated using a data set containing masses with ill-defined margins and irregular shape or spiculations. WI (with CI at 0.95) of 1.002 ± 0.010 for AOR and 0.9815 ± 0.021 for MAOR were obtained for segmenting the whole mass and the margin portion, respectively. This indicated that the results of the MLS approach well agreed with the radiologists' delineation.

FIG. 10. The WIs with CI at 0.95 for the three segmentation measurements of the algorithm and the radiologists. (a) WI of AOR, (b) WI of AMINDIST, and (c) WI of HSDIST.

In the field of mammographic mass segmentation, investigators have been successful in applying image processing techniques for delineations of mass body. Differing from previous works, the proposed MLS approach specifically addresses the technical issue in effective inclusion of the ill-

defined margin and spiculations as a part of mass segmentation. It copes with this important issue with the following advantages:

- By ISLM, the algorithm could enhance the mass structure to be segmented while substantially suppressing the influences from clutters and background structures. Thus, the algorithm may help segmenting masses in the presence of overlapping or surrounding dense glandular tissues and pectoral muscles, when their intensity or patterns are differentiable from the mass region through

TABLE II. MAOR for various methods.

	MLS	MLFA	LS
MAOR	0.540 (± 0.164)	0.333 (± 0.200)	0.307 (± 0.148)

FIG. 11. Margin segmentation performance. (a) The box and whisker plots of the distribution of MAOR, (b) WI of MAOR for the algorithm and the radiologists. The vertical dashed line in (a) corresponds to the threshold for a good margin segmentation with MAOR=0.5, which is determined in this study by considering the difficulty in segmenting the margin portion.

FIG. 12. Illustration of cases with considerable disagreement between the computer segmentation and the ground truth. (a) A segmentation of AOR = 0.361 is obtained on a low contrast mammogram. (b) A segmentation case with falsely included ducts as spiculations. The contours superimposed on the original image are the MLS approach (white contour) and manual segmentation (black contour).

the learning step. Traditional region growing based methods³ may easily “flood” into these unwanted areas, if they are designed to include more mass margin portion.

- In the PM generated by ISLM, image values for the mass are more uniform than the original image patterns (e.g., intensity and texture). In other words, the ill-defined margin and mass appearance variations are normalized through ISLM. Therefore, more mass margin could be included relatively easily in comparison with methods directly using the image intensity and gradient information.^{7,8}
- By integrating spiculation detection, the MLS approach would become a more effective segmentation method for its ability in delineating ill-defined margins, irregular shape and spiculations. This would benefit the mass characterization module in many mammographic computer aided detection and diagnosis systems.

We also learned that the ISLM intermediate output (i.e., PM) is a desired byproduct of the approach, which could be used as the image content descriptors for analyzing the characteristics of mammographic masses. These features may be useful for automatic analysis of malignancy of breast lesions and to retrieve mammograms with similar image patterns in a content-based image retrieval system.²⁸

ACKNOWLEDGMENTS

This project was supported in part by an NIH/NCI grant (Grant No. R33CAI02960). The authors would like to express their gratitude to Catherine Chow, M.D., Zahide Erkmen, M.D., Lisa Johnson, M.D., Sara Petrillo, M.D., and Cullen Ruff, M.D., for their clinical discussions in the study.

^{a)}Electronic mail: ymtao@vt.edu

¹S. Pohlman, K. A. Powell, N. A. Obuchowski, W. A. Chilcote, and S. Grundfest-Broniatowski, “Quantitative classification of breast tumors in digitized mammograms,” *Med. Phys.* **23**, 1337–1345 (1996).

²M. A. Kupinski and M. L. Giger, “Automated seeded lesion segmentation on digital mammograms,” *IEEE Trans. Med. Imaging* **17**, 510–517 (1998).

³L. Kinnard, S.-C. B. Lo, E. Makariou, T. Osicka, P. Wang, M. F. Chouikha, and M. T. Freedman, “Steepest changes of a probability-based cost function for delineation of mammographic masses: A validation study,” *Med. Phys.* **31**, 2796–2810 (2004).

⁴G. M. te Brake and N. Karssemeijer, “Segmentation of suspicious densities in digital mammograms,” *Med. Phys.* **28**, 259–266 (2001).

⁵H. Li, Y. Wang, K. J. R. Liu, S.-C. B. Lo, and M. T. Freedman, “Computerized radiographic mass detection—Part I: Lesion site selection by morphological enhancement and contextual segmentation,” *IEEE Trans. Med. Imaging* **20**, 289–301 (2001).

⁶B. Sahiner, H.-P. Chan, N. Petrick, M. A. Helvie, and L. M. Hadjiiski, “Improvement of mammographic mass characterization using spiculation measures and morphological features,” *Med. Phys.* **28**, 1455–1465 (2001).

⁷J. Shi, B. Sahiner, H.-P. Chan, J. Ge, L. M. Hadjiiski, M. A. Helvie, Y.-T. Wu, A. Nees, J. Wei, C. Zhou, Y. Zhang, and J. Cui, “Characterization of mammographic masses based on level set segmentation with new image features and patient information,” *Med. Phys.* **35**, 280–290 (2008).

⁸A. R. Domínguez and A. K. Nandi, “Improved dynamic-programming-based algorithms for segmentation of masses in mammograms,” *Med. Phys.* **34**, 4256–4269 (2007).

⁹M. Elter and A. Horsch, “CADx of mammographic masses and clustered microcalcifications: A review,” *Med. Phys.* **36**, 2052–2068 (2009).

¹⁰J. E. Martin, *Atlas of Mammography: Histologic and Mammographic Correlations*, 2nd ed. (Williams and Wilkins, Baltimore, 1988).

¹¹J. R. Harris, M. E. Lippman, M. Morrow, and S. Hellman, *Diseases of the Breast* (Lippincott-Raven, Philadelphia, 1996).

¹²Y. Boykov and V. Kolmogorov, “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision,” *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 1124–1137 (2004).

¹³American College of Radiology (ACR), *Breast Imaging Reporting and Data System (BI-RADS) Atlas*, 4th ed. (American College of Radiology, Reston, 2003).

¹⁴R. Haralick, “Statistical and structural approaches to texture,” *Proc. IEEE* **67**, 786–804 (1979).

¹⁵T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 971–987 (2002).

¹⁶P. Campadelli, E. Casiraghi, and D. Artioli, “A fully automated method for lung nodule detection from postero-anterior chest radiographs,” *IEEE Trans. Med. Imaging* **25**, 1588–1603 (2006).

¹⁷Z. Tu, “Auto-context and its application to high-level vision tasks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

¹⁸R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. (Wiley-Interscience, New York, 2000).

¹⁹N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).

²⁰C. J. Vyborny, T. Doi, K. F. O’Shaughnessy, H. M. Romsdahl, A. C. Schneider, and A. A. Stein, “Breast cancer: Importance of spiculation in computer-aided detection,” *Radiology* **215**, 703–707 (2000).

²¹S. J. Caulkin, S. M. Astley, A. Mills, and C. R. M. Boggis, “Generating realistic spiculated lesions in digital mammograms,” in *Proceedings of the 5th International Workshop on Digital Mammography*, 2000, pp. 713–720.

²²M. Jacob and M. Unser, “Design of steerable filters for feature detection using canny-like criteria,” *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 1007–1019 (2004).

²³R. Zwiggelaar, S. M. Astley, C. R. M. Boggis, and C. J. Taylor, “Linear structures in mammographic images: Detection and classification,” *IEEE Trans. Med. Imaging* **23**, 1077–1086 (2004).

²⁴Y. Tao, S.-C. B. Lo, M. T. Freedman, and J. Xuan, “Joint segmentation and spiculation detection for ill-defined and spiculated mammographic masses,” *Proc. SPIE* **7624**, 762407 (2010).

²⁵G. W. Williams, “Comparing the joint agreement of several raters with another rater,” *Biometrics* **32**, 619–627 (1976).

²⁶V. Chalana and Y. Kim, “A methodology for evaluation of boundary detection algorithms on medical images,” *IEEE Trans. Med. Imaging* **16**, 642–652 (1997).

²⁷B. D. Thackray and A. C. Nelson, “Semi-automatic segmentation of vascular network images using a rotating structuring element (ROSE) with mathematical morphology and dual feature thresholding,” *IEEE Trans. Med. Imaging* **12**, 385–392 (1993).

²⁸Y. Tao, S.-C. B. Lo, M. T. Freedman, and J. Xuan, “A preliminary study of content-based mammographic masses retrieval,” *Proc. SPIE* **6514**, 65141Z (2007).