Computerized characterization of breast masses on three-dimensional ultrasound volumes

Berkman Sahiner,[a] Heang-Ping Chan, Marilyn A. Roubidoux, Mark A. Helvie, Lubomir M. Hadjiiski, Aditya Ramachandran, Chintana Paramagul, Gerald L. LeCarpentier, Alexis Nees, and Caroline Blane
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109-0904

(Received 5 August 2003; revised 17 December 2003; accepted for publication 23 December 2003; published 11 March 2004)

We are developing computer vision techniques for the characterization of breast masses as malignant or benign on radiologic examinations. In this study, we investigated the computerized characterization of breast masses on three-dimensional (3-D) ultrasound (US) volumetric images. We developed 2-D and 3-D active contour models for automated segmentation of the mass volumes. The effect of the initialization method of the active contour on the robustness of the iterative segmentation method was studied by varying the contour used for its initialization. For a given segmentation, texture and morphological features were automatically extracted from the segmented masses and their margins. Stepwise discriminant analysis with the leave-one-out method was used to select effective features for the classification task and to combine these features into a malignancy score. The classification accuracy was evaluated using the area A_z under the receiver operating characteristic (ROC) curve, as well as the partial area index $A_z^{(0.9)}$, defined as the relative area under the ROC curve above a sensitivity threshold of 0.9. For the purpose of comparison with the computer classifier, four experienced breast radiologists provided malignancy ratings for the 3-D US masses. Our dataset consisted of 3-D US volumes of 102 biopsied masses (46 benign, 56 malignant). The classifiers based on 2-D and 3-D segmentation methods achieved test A_z values of 0.87 ± 0.03 and 0.92 ± 0.03, respectively. The difference in the A_z values of the two computer classifiers did not achieve statistical significance. The A_z values of the four radiologists ranged between 0.84 and 0.92. The difference between the computer's A_z value and that of any of the four radiologists did not achieve statistical significance either. However, the computer's $A_z^{(0.9)}$ value was significantly higher than that of three of the four radiologists. Our results indicate that an automated and effective computer classifier can be designed for differentiating malignant and benign breast masses on 3-D US volumes. The accuracy of the classifier designed in this study was similar to that of experienced breast radiologists. [DOI: 10.1118/1.1649531]

Key words: computer-aided diagnosis, 3-D ultrasound, breast mass characterization, segmentation

I. INTRODUCTION

The importance of early breast cancer detection requires a vigorous approach to the characterization of breast lesions. At present, the positive biopsy rate for nonpalpable breast lesions as well as for nonpalpable breast masses is between 15%–30%.[1–4] This means that 70%–85% of breast biopsies are performed for benign lesions. In order to reduce patient anxiety and morbidity, as well as to decrease health care costs, it is desirable to reduce the number of benign biopsies without missing malignancies. Computer-aided diagnosis (CAD) can provide a consistent and reproducible second opinion to the radiologists, and has a potential to assist them in reducing benign biopsies. Recent studies on the computerized classification of breast masses based on mammographic image features suggest that the radiologists' performance may be significantly improved if they are aided by a well-trained CAD system.[5–7] Breast ultrasound (US) is an important imaging modality for the characterization of breast masses as malignant and benign. An objective and reproducible second opinion from a computer classifier for the classification of breast masses based on US image features may be an important addition to CAD tools being developed for mammographic image analysis.

Breast US is widely accepted as a highly accurate modality for the differentiation of cystic and noncystic masses. As a result of technological improvements and more sophisticated utilization by radiologists, US has been gaining popularity for the characterization of noncystic, or solid, breast masses. By combining several ultrasonic characteristics, Stavros *et al.*[8] achieved a specificity of 98.4% and a sensitivity of 68.7% on a dataset of 750 solid breast masses. Using strict criteria for a benign diagnosis, Skaane *et al.*[9] achieved a positive predictive value of 66% and a negative predictive value of 98% for the differentiation of fibroadenoma and invasive ductal carcinoma on sonograms. Recently, Taylor *et al.* investigated whether the complementary use of US imaging could decrease the biopsy of benign, noncystic masses. On a dataset of 761 biopsied masses, they

found that the addition of US evaluation to mammography alone could increase the specificity from 51.4% to 63.8% while slightly increasing the sensitivity from 97.1% to 97.9%.[10] In our study we aim at developing techniques for the computerized characterization of solid breast masses, which may eventually improve the radiologists' accuracy in this difficult and important task.

A number of researchers have recently investigated the application of CAD to breast US images.[11–14] Chen *et al.*[12] extracted autocorrelation features from rectangular regions of interest (ROIs) containing solid breast masses. Using a neural network classifier, they obtained an area A_z under the receiver operating characteristic (ROC) curve of 0.956 for classification of a dataset of 140 biopsy-proven masses as malignant or benign. Horsch *et al.*[13] developed an automated segmentation method for delineating the mass boundaries, and compared its characterization accuracy on different subsets with that obtained from manual segmentation. Using manual and automated segmentation methods, they obtained A_z values of 0.91 and 0.87, respectively, in the task of differentiating all malignant and benign lesions in their dataset, and 0.88 and 0.82, respectively, in the task of differentiating the subset of malignant and benign solid lesions. Chen *et al.*[14] used morphological features extracted from manually segmented mass boundaries for classification. Using a neural network classifier, they obtained an A_z of 0.959 for classification of a dataset of 271 biopsy-proven masses as malignant or benign.

A 3-D US is rapidly gaining popularity as it moves out of the research environment and into the clinical setting.[15] A computerized analysis of 3-D US images may be useful for two reasons. First, 3-D or volumetric US data may be more time consuming for a radiologist to interpret, thus making CAD more desirable. Second, 3-D or volumetric US provides more data and better statistics, which should improve statistical image analysis.

In clinical practice, breast US may be performed in different ways. In many breast imaging clinics, the US examination is performed by a US technologist. Once the technologist locates the mass, and determines the appropriate settings for optimal image quality, representative static US images of the mass are printed on hardcopy film. The radiologist only reads the images chosen by the technologist. A second possibility is that the US scan is videotaped by the technologist and the radiologist reads the examination on a video display. In a third method, a radiologist will perform the US examination interactively and optimize the image quality by changing the probe angle, direction, and US machine settings. Since the US image quality is operator dependent, the way in which the examination is performed may have an impact on the diagnostic accuracy. At our institution, the third method is employed. As described in Sec. II, the data acquisition system in this study did not permit interactive modification during 3-D image acquisition. As a result, the data that was used by the computer and the radiologists for mass characterization in this study may not be as informative as the data that the radiologists could have obtained by examining the patient interactively. However, since the mass is entirely imaged in the 3-D dataset, our data should be at least comparable to that obtained by using the first method described above.

In this study, we investigated the computerized characterization of noncystic breast masses as malignant and benign in 3-D US images. We developed a 3-D segmentation method to delineate the masses. Morphological and texture features were extracted from the mass and its margins for classification. A linear classifier was used to merge the features into a malignancy score. The classification accuracy was evaluated by ROC methodology. The ROC curves of the computer and four experienced breast radiologists were compared. To our knowledge, this is the first study on 3-D US images that investigates a computer segmentation method followed by a computer classifier for breast cancer characterization.

II. METHODS

A. Dataset

Institutional review board approval was obtained prior to the commencement of this investigation. The images used in this study were acquired between 1998 and 2002. Our study group was 102 women (average age: 51 years) who had a solid mass deemed suspicious or highly suggestive of malignancy. All patients underwent biopsy or fine needle aspiration. Fifty-six masses were malignant and 46 were benign. Forty-three of the malignancies were invasive ductal carcinoma, five were invasive lobular carcinoma, one was medullary carcinoma, three were ductal carcinoma *in-situ*, and four were other invasive carcinoma. Of the benign masses, the majority were fibroadenoma ($N=18$) and fibrocystic disease ($N=11$). The mean equivalent lesion diameter was 1.28 cm (standard deviation=0.78 cm).

The 3-D US data were acquired using an experimental system that was previously developed and tested at our institution.[16,17] The 3-D system consisted of a commercially available US scanner (GE Logiq 700 with an M12 linear array transducer), a mechanical transducer guiding system, and a computer workstation. The linear array transducer was operated at 11 MHz. The technologist was free to set the focal distance and the overall gain adjustment to obtain the best possible image. Before 3-D image acquisition, the technologist used clinical US and mammogram images to identify the suspicious mass. During 3-D image acquisition, the technologist manually translated the transducer linearly in the cross-plane, or the z direction, while the image acquisition system recorded 2-D B-mode images in the image scan plane (x-y plane). The 2-D images were obtained at approximately 0.5 mm incremental translations, which were measured and recorded using a translation sensor. The number of 2-D slices was typically around 90, and varied depending on the lesion size. The maximum distance between two 2-D slices was 0.5 mm, and some of the distances were slightly less than 0.5 mm. The scanned breast region measured typically 4.5 cm long by 4.0 cm wide by 4.0 cm deep. The typical pixel size in a slice was approximately 0.11 mm.

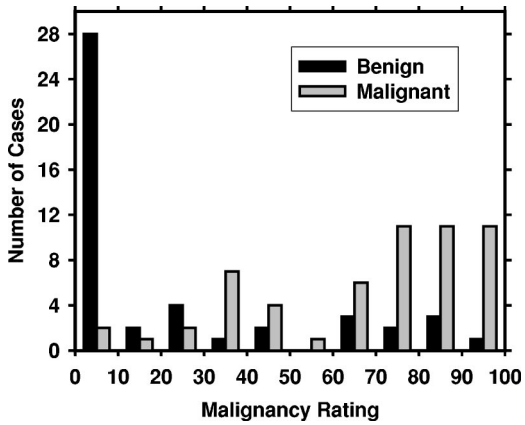

FIG. 1. The distribution of the malignancy rating of the masses in our dataset based on the appearance on US images, by an experienced radiologist. 1: Very likely benign; 100: very likely malignant.

The B-mode images were recorded into a buffer in the US scanner. After data acquisition, the images and the position data were transferred digitally to a workstation, where individual planes were cropped and stacked to form a 3-D volume. The biopsied mass in each volume was identified by a MQSA (Mammography Quality Standards Act) qualified radiologist (RAD1) using clinical US and mammographic images to confirm that the 3-D images contained the suspicious mass. The likelihood of malignancy for each mass, based on the 3-D US image alone, was rated by the same radiologist on a scale of 1 to 100, where a higher number corresponded to a higher likelihood of malignancy. The distribution of the ratings for the malignant and benign masses is shown in Fig. 1. The radiologist was also asked to fit a 3-D ellipsoid to the mass. The 3-D ellipsoid was used to initialize the computerized mass segmentation described in the next section. The best fit was obtained by scaling, rotating, and translating an ellipsoid superimposed on the 3-D dataset using a dynamic object manipulation tool developed for this purpose.

B. Mass segmentation

We investigated the use of 2-D and 3-D active contour models for the segmentation of mass boundaries.[18] An active contour model is a high-level segmentation method that uses energy terms derived from the image gray-level information as well as the *a-priori* knowledge about the object to be segmented for accurate segmentation. The segmentation problem is defined as an energy minimization problem. In order for the model to lock onto the contours in the image, the image-based energy terms, also referred to as the external energy terms, are usually defined in terms of the image gray levels and the image gradient magnitude. The *a-priori* knowledge of the object shape is used to define internal energy terms related to features such as the continuity and the smoothness of the contour to constrain the segmentation problem. These terms can compensate for noise or apparent gaps in the image gradients, which often mislead segmentation methods that do not use *a-priori* information.

In a 2-D segmentation problem, the contour of the object can be represented by V vertices, (i_ν, j_ν), $\nu = 1,...,V$, where i and j represent the two dimensions of the image. In the discrete formulation of the active contour model, the total energy to be minimized is defined as

$$E = \sum_{\nu=1}^{V} E(\nu), \tag{1}$$

where $E(\nu)$ is the energy at vertex (i_ν, j_ν). $E(\nu)$ is defined as the sum of the internal and external energy terms,

$$E(\nu) = \sum_{m=1}^{M} w_m E_m(\nu), \tag{2}$$

where $E_m(\nu)$ is the mth energy term at vertex ν, and w_m is the weight of the mth energy term. In our 2-D active contour model, we used four internal and external energy terms ($M = 4$). The energy terms E_1, E_2, E_3, and E_4 were determined by the gradient magnitude of the image and the continuity, smoothness, and balloon energy of the contour, respectively.

To obtain the image gradient magnitude, the image $A(i,j)$ was first filtered using a Gaussian smoothing filter,

$$H(i,j) = e^{-(i^2+j^2)/2\sigma^2}, \tag{3}$$

where $\sigma^2 = 6$. The resulting filtered image $B(i,j)$ was further processed using Sobel filters $S_x(i,j)$ and $S_y(i,j)$, defined as

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \tag{4}$$

which calculated the x- and y-direction gradients, $G_x(i,j)$ and $G_y(i,j)$, respectively. The image gradient magnitude at vertex $\nu = (i_\nu, j_\nu)$ was computed as

$$E_1(\nu) = \sqrt{G_x(i_\nu, j_\nu) + G_y(i_\nu, j_\nu)}. \tag{5}$$

The weight of the gradient energy was defined to be a negative number; thus, minimizing $w_1 E_1$ attracted the contour to image edges.

To find the continuity energy term, we first computed the average line segment length $\overline{d}$ as

$$\overline{d} = \frac{\sum_{\nu=1}^{V} d(\nu)}{V}, \tag{6}$$

where

$$d(\nu) = \begin{cases} \sqrt{(i_\nu - i_{\nu+1})^2 + (j_\nu - j_{\nu+1})^2}, & \nu = 1,2,...,V-1, \\ \sqrt{(i_\nu - i_0)^2 + (j_\nu - j_0)^2}, & \nu = V. \end{cases} \tag{7}$$

The continuity energy term was defined as

$$E_2(\nu) = |d(\nu) - \overline{d}|. \tag{8}$$

Minimizing the continuity energy helped the vertices maintain regular spacing along the contour.

The curvature term, $E_3(\nu)$, was approximated by the second derivative of the contour,

$$E_3(\nu)=\sqrt{(i_{\nu-1}-2i_\nu+i_{\nu+1})^2+(j_{\nu-1}-2j_\nu+j_{\nu+1})^2}. \quad (9)$$

When the vertices were spaced regularly along the contour, this term would be large when the angle at vertex ν was small.[19] By discouraging small angles at vertices, this term attempted to smooth the contour.

The balloon energy $E_4(\nu)$ pushed the contour outward or pulled it inward, depending on whether w_4 was positive or negative, respectively, along a path normal to the contour. This energy term helped the active contour traverse spurious, isolated, or weak image edges, and countered its tendency to shrink. The resulting model was reported to be more robust to the initial position and image noise.[20]

To solve the energy minimization problem, we have chosen the iterative method proposed by Williams and Shah.[19] The contour is first initialized by defining V vertices (i_ν,j_ν), $\nu=1,...,V$. At a given iteration, the method visits each vertex (i_ν,j_ν). Let $\mathbf{D}(\nu)$ represent the set of pixels (i',j') in a $(2M+1)x(2M+1)$ neighborhood centered around (i_ν,j_ν). For each pixel in $\mathbf{D}(\nu)$, the sum $\Sigma_m w_m E_m$ is computed, and the vertex (i_ν,j_ν) is moved to the (i'^*,j'^*) location that minimizes this sum. The definitions of the energy terms E_1, E_2, and E_3 are given above. The balloon energy E_4 was defined as $E_4=\cos\theta$, where θ represents the angle between the normal vector to the curve at vertex ν and the vector $(i'-i_\nu,j'-j_\nu)$. After the minimization is performed locally at vertex (i_ν,j_ν), the algorithm moves to the vertex $(i_{\nu+1},j_{\nu+1})$. The method converges when no vertex changes location at a given iteration. In practical implementation, iterations may be stopped when a large, predetermined percentage of vertices stop moving. The cross section of the radiologist-defined ellipsoid with each image slice was used for initializing the contour.

When the 2-D active contour model described above is applied to a 3-D dataset, segmentation is performed independently on each slice of the 3-D volume. However, this kind of segmentation ignores the continuity of the object across slices. When the slice spacing is small compared to the rate of change of the object shape, it is expected that the shape of the object is unlikely to change drastically from one slice to the next. Our 3-D active contour model is aimed at using the shape information across the 3-D slices to improve upon the 2-D active contour model. Our 3-D active contour model was defined by including in the curvature energy term, an additional component related to the smoothness of the mass in the z direction. Let $(i_{\nu,k},j_{\nu,k})$ denote the νth vertex in image slice k. The curvature energy in our 3-D active contour model was defined as

$$\begin{aligned}E_3(\nu)&=\sqrt{(i_{\nu-1,k}-2i_{\nu,k}+i_{\nu+1,k})^2+(j_{\nu-1,k}-2j_{\nu,k}+j_{\nu+1,k})^2}\\&\quad+\alpha\sqrt{(i_{\nu,k-1}-2i_{\nu,k}+i_{\nu,k+1})^2+(j_{\nu,k-1}-2j_{\nu,k}+j_{\nu,k+1})^2},\end{aligned} \quad (10)$$

where α was the weight of the out-of-plane component of the curvature relative to the in-plane component. The out-of-plane component forced the contour to be smooth in the z direction. Our implementation of the 3-D active contour model started by optimizing the contour in the first slice of the 3-D dataset $(k=1)$. Since slice $k=0$ did not exist, we assumed that $(i_\nu,j_\nu,0)=(i_\nu,j_\nu,1)$ for all ν. The contour optimization in slice $k=1$ followed the steps described above for 2-D active contours, except that the curvature energy was replaced by Eq. (10). After the contour was optimized for slice $k=1$, the optimization was performed for slice $k=2$, and so on, until the contours were optimized for all slices. This constituted one 3-D iteration. The 3-D model repeated the 3-D iterations until there was no movement of the vertices for the 3-D contour, or when a predetermined percentage of vertices stopped moving. Similar to our 2-D active contour, the 3-D active contour was initialized using the radiologist-defined ellipsoid.

We did not employ an optimization method for determining the active contour weights because automatic optimization required the comparison of the automated contour with a gold standard such as the radiologist's manual segmentation for training. The ''true'' borders of many masses on US images were not well defined, even to experienced radiologists. Furthermore, the features that we designed did not require a border that followed the detailed boundary of an ill-defined or a spiculated mass. We therefore used more subjective judgment on the ''goodness of segmentation'' for the mass boundary based on our experience with the need of the features. To determine the weights for the 2-D model, we started with weights we had previously used for the segmentation of masses on mammograms.[21] We experimentally modified the weights and observed the effect on the segmentation quality for the first 15 volumes in our dataset. We found that the combination $w_1=-1.5$, $w_2=1$, $w_3=2.6$, and $w_4=0.2$ provided a good balance between the smoothness of the contour and its the attraction to the mass borders. These weights were then used for the 2-D segmentation of the entire dataset. For the 3-D active contour model, we maintained the weights at the values that we determined for the 2-D active contour model, and selected $\alpha=0.5$. The choice of α was again based on a qualitative assessment of segmentation on the first 15 cases.

C. Feature extraction

We have evaluated a number of morphological and texture features for characterization of the masses as malignant or benign. Each of the features described below was extracted from every slice where the mass was segmented using either the 2-D or the 3-D automated segmentation algorithm. The features extracted from different slices of the same mass were then combined to define the feature measures (such as mean or maximum) for that mass.

1. Extraction of morphological features

The taller-than-wide shape of a sonographic mass is a good indication of malignancy.[8] This characteristic was defined by the ratio of the widest cross section (W) of the

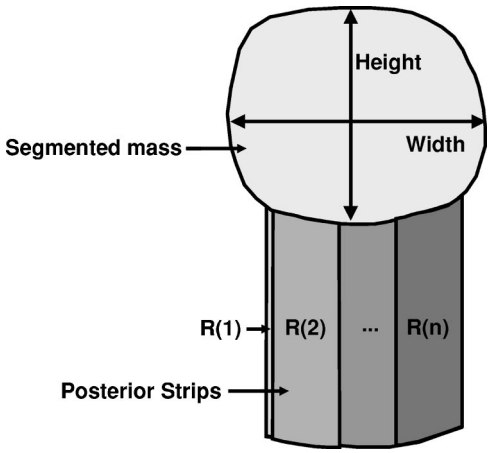

FIG. 2. The definition of the width-to-height and PSF features. The width-to-height feature was defined as the ratio of the widest cross section of the segmented mass shape in the image plane to the tallest cross section. The PSF feature was defined by first finding the average gray value in the posterior strips $\overline{R(i)}$, $i=1,...,n$, then finding the minimum of $\overline{R(i)}$ among the n strips, and finally by normalizing this value by the average gray value within the segmented mass.

automatically segmented lesion shape to the tallest cross section (T) in a slice (Fig. 2). Another feature that has been reported to be useful for differentiation of malignant and benign masses is posterior shadowing. In order to define a posterior shadowing feature (PSF), we first calculated the mean pixel value $\overline{R(i)}$ in overlapping vertical strips $R(i)$, $i=1,...,n$ posterior to the mass, as shown in Fig. 2. The width W_R of a strip was equal to one-fourth of the width of the mass ($W/4$), and the height of the strip was equal to the height of the mass (T). The left and right edges of strips $R(i)$ and $R(i+1)$ differed by one pixel. In other words, the strip $R(i+1)$ was obtained by moving the strip $R(i)$ to the right by one pixel, while, of course, the strip remained posterior to the mass and its height remained as T. In order to exclude the bilateral posterior shadowing artifacts that are sometimes associated with fibroadenomas, the strips were defined only posterior to the central $3W/4$ portion of the mass (Fig. 2). The minimum value of these averages, $\min\{\overline{R(i)}, i=1,...,n\}$, was the darkest posterior strip. The PSF was defined as the normalized average gray-level difference between the interior of the segmented mass and the darkest posterior strip,

$$\mathrm{PSF}=\frac{\overline{M}-\min\{\overline{R(i)}, i=1,...,n\}}{\overline{M}}, \tag{11}$$

where $\overline{M}$ denotes the mean gray level value inside the segmented mass.

2. Extraction of texture features

The features used in this study were extracted from spatial gray-level dependence (SGLD) matrices, or co-occurrence matrices, derived from 2-D slices of the 3-D dataset. The (i,j)th element of the co-occurrence matrix is the relative frequency with which two pixels: one with gray level i and the other with gray level j, separated by a pixel pair distance d in a direction θ occur in the image. Features extracted from SGLD matrices of US images have been shown to be useful in the classification of malignant and benign breast masses on mammograms in previous studies.[22] In this study, six texture feature measures that are invariant under linear, invertible gray scale transformations were extracted. These features were information measures of correlations 1 and 2 (IMC1 and IMC2), difference entropy (DFE), entropy (ENT), energy (ENE), and sum entropy (SME). The mathematical definitions of these features can be found in the literature.[23] Although many gray scale transformations may not be invertible due to pixel saturation or roundoff, these features are largely independent of the gray-level gain adjustments.

It is known that the margin characteristics of a mass are very important for its characterization, and previous studies have indicated that texture features extracted from the mass margins are effective for classification.[24] For this reason, the texture features in this study were extracted from two disk-shaped regions containing the boundary of each mass, as well as presumably mass and normal tissue adjacent to the boundary of the mass. These regions followed the contour determined by the active contour model, as shown in Fig. 3. The areas for the upper and lower disk-shaped regions were chosen to be equal, and their sum was equal to the area of the segmented mass. The pixel pair distances used for SGLD matrix computation were chosen to be $d=2$, 4, and 6. Two pixel pair angles, $\theta=0°$ and $\theta=90°$, were evaluated for each d in both regions. The number of SGLD matrices computed for a disk-shaped region was therefore 6, and the number of features extracted from an image containing the segmented mass was 72 (6 features, extracted from 6 SGLD matrices in the upper disk-shaped region and the lower disk-shaped region).

D. Classification

The features extracted from different slices of the same mass were combined to define the feature measures for that mass. For the width-to-height feature and the PSF, we computed the mean, variance, minimum, and maximum of the extracted value from each slice containing the mass. Therefore eight morphological feature measures were defined for each mass. For texture features, we only computed the mean, hence 72 texture feature measures were defined for each mass.

Fisher's linear discriminant analysis (LDA)[25] was used for combining the features into a discriminant score. Since the number of available features in the feature space was relatively high compared with the number of available cases, stepwise feature selection[26] was used in order to reduce the number of the features and to obtain the best feature subset to design an effective classifier. For partitioning the dataset into trainers and testers, we used the leave-one-case-out resampling method. Feature selection is performed as part of the classifier design such that both the feature selection and the classifier coefficient estimation procedures were repeated 102 times, as each case was left out once as the test sample. The test discriminant scores were analyzed using ROC

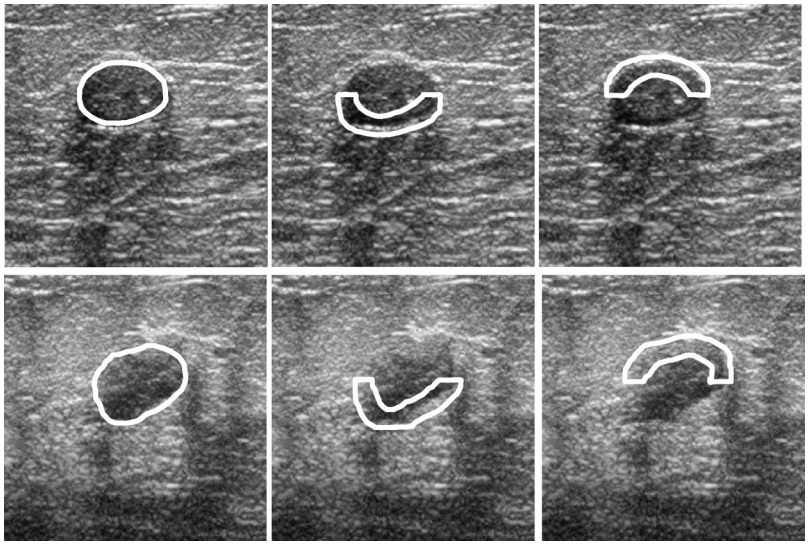

FIG. 3. Left column: The segmented object for a malignant mass (upper row) and a benign mass (lower row). Middle and right columns: The lower and upper disk-shaped regions from which texture features were extracted.

methodology.[27] The classification accuracy was evaluated using the area under the ROC curve, A_z, as well as the partial area index, $A_z^{(0.9)}$. $A_z^{(0.9)}$ is defined as the area under the ROC curve above a sensitivity threshold of 0.9 ($TPF_0 = 0.9$) normalized to the total area above TPF_0, which is equal to $(1 - TPF_0)$.[28]

E. Malignancy ranking by radiologists

Although all the cases in our dataset were suspicious enough to warrant biopsy or fine needle aspiration, the degree of difficulty of our cases can best be measured by investigating the accuracy of the radiologists in classifying the cases in our dataset as malignant or benign. As described in Sec. II B, one radiologist (RAD1) who was familiar with the clinically obtained images had initially provided a malignancy rating. To compare with the computer's accuracy, we are interested in measuring the accuracy of other radiologists, who would not be biased by memory or familiarity with the cases. For this purpose, we have developed an interactive graphical user interface with which the radiologists could navigate through 3-D volumes, adjust the window and level of the displayed images, and enter a malignancy rating between 1 and 100 (a higher rating indicating a higher likelihood of malignancy) when they finish examining a case. Three additional radiologists (RAD2–RAD4) participated in the malignancy rating study. The radiologists RAD1–RAD4

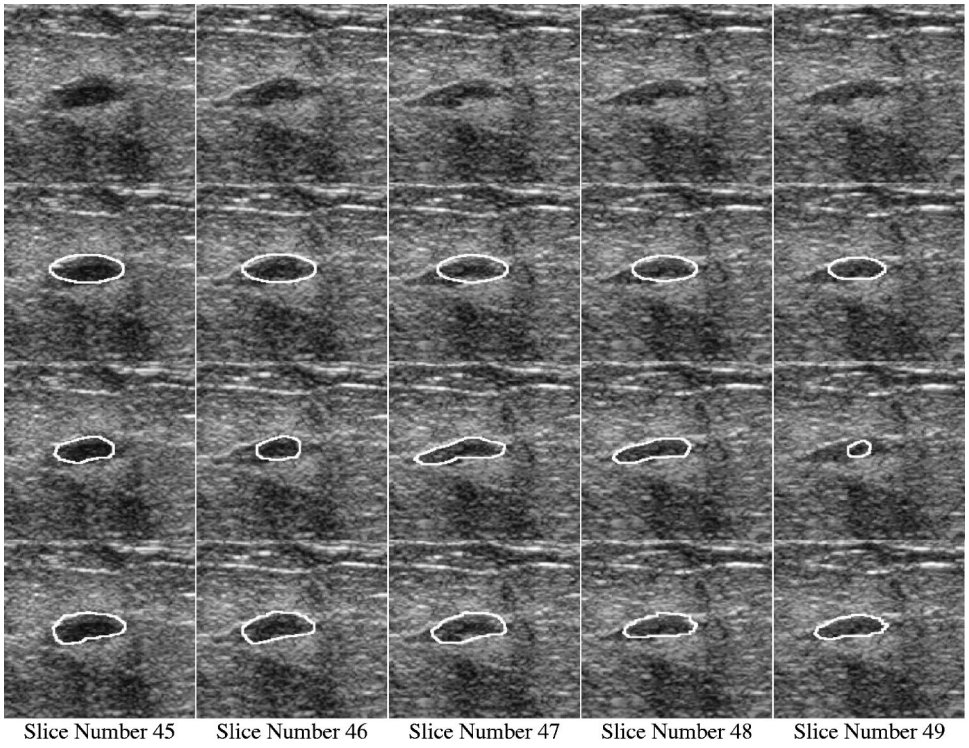

FIG. 4. Row 1: Five original slices of a breast mass that was visible on a total of ten US slices; row 2: The cross section of the initial 3-D ellipsoid at each slice; row 3: The result of the 2-D active contour segmentation method; row 4: The result of the 3-D active contour segmentation method. Note that the 2-D segmentation method missed part of the mass on slice 46. The 3-D segmentation method, apparently using the information from slices 45 and 47, was able to provide better segmentation on slice 46.

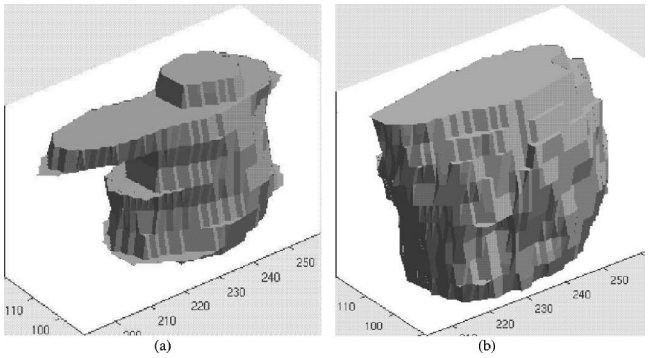

FIG. 5. 3-D rendering of the segmented object for the mass shown in Fig. 4. (a) 2-D active contour segmentation; (b) 3-D active contour segmentation.

were either fellowship trained in breast imaging or had over 25 years of experience in breast imaging. All four radiologists were MQSA qualified and their experience in mammographic and US interpretation ranged from 2 to 25 years (mean, 11.3 years). The location of the mass center, as determined by RAD1, was displayed on each slice, so that all the radiologists would rank the same mass if more than one mass existed in the volume. There was no time limitation for the radiologists to read a case. The case reading order was randomized for each radiologist. The malignancy rating was entered by means of a slide bar. Before participating in the study, the radiologists were trained on five cases that were not part of the test dataset described in Sec. II A. The malignancy rating study was intended to measure the difficulty of the dataset, and was not intended to measure how the radiologists' interpretation would be affected by CAD. Therefore, the computer classification results were not displayed to the radiologists in this study.

III. RESULTS

We evaluated the accuracy of characterization based on both 2-D and 3-D active contour segmentation methods. Rows 1 to 4 of Fig. 4 show the original images, radiologist-defined ellipsoid, 2-D active contour results, and 3-D active contour results for five consecutive slices of a mass that was visible on a total of 10 slices. Figure 5 shows a 3-D rendering of the segmented object using the 2-D and 3-D active contour models. It is seen from Fig. 5 that the shape of the object segmented by the 3-D active contour model is smoother in the z direction.

Table I shows the range (minimum and maximum) of the A_z values provided by each texture feature alone, extracted from the upper and lower disk-shaped regions determined by the 2-D and 3-D active contour models. The ranges in this table are for different pixel pair distances and directions used in extracting the same feature (e.g., IMC1). Table II shows the range of A_z values provided by each morphological feature alone, using the 2-D and 3-D active contour models. The ranges in Table II are for different methods of combining the features extracted from individual slices, i.e., mean, variance, minimum, and maximum. The most discriminatory feature in this study was the IMC1 feature ($d=6$, $\theta=0°$, extracted from the upper disk-shaped region segmented by the 3-D method) with an A_z value of 0.76.

When stepwise LDA was used to combine the features into a discriminant score in the 102 leave-one-case-out training subsets, an average of 6.09 and 7.98 features were selected with the 2-D and 3-D segmentation methods, respectively. For the 2-D segmentation method, the most frequently selected features were two IMC1 features, two IMC2 features, one DFE feature, and one width-to-height feature. For the 3-D segmentation method, the most frequently selected features were two IMC1 features, two IMC2 features, one DFE feature, one ENT feature, one PSF feature, and one

TABLE I. The range of A_z values for different texture features extracted from the lower and upper disk-shaped regions using the 3-D and 2-D segmentation methods. For each particular texture feature (e.g., IMC1 feature at pixel-pair distance $d=2$, and direction $\theta=0°$), the feature values from all the slices containing the segmented mass were averaged before computing the A_z value. The range indicates the minimum–maximum A_z values for a particular feature among the parameters $d=2$, 4, 6 and $\theta=0°$, 90°.

	3-D segmentation		2-D segmentation	
Texture feature	Upper	Lower	Upper	Lower
IMC1	0.66–0.76	0.58–0.67	0.65–0.72	0.59–0.66
IMC2	0.65–0.75	0.58–0.65	0.65–0.73	0.61–0.67
DFE	0.58–0.68	0.61–0.67	0.56–0.68	0.62–0.70
ENT	0.59–0.64	0.55–0.60	0.62–0.69	0.58–0.62
ENE	0.57–0.63	0.53–0.60	0.53–0.60	0.50–0.54
SME	0.52–0.58	0.51–0.56	0.57–0.64	0.52–0.57

TABLE II. The range of A_z values for the width-to-height feature and posterior shadowing feature (PSF) extracted using the 3-D and 2-D segmentation methods. The range indicates the minimum–maximum A_z values among the mean, variance, minimum, and maximum of each feature extracted from each slice containing the segmented mass.

Morphological feature	3-D segmentation	2-D segmentation
Width-to-height	0.58–0.73	0.54–0.69
PSF	0.53–0.66	0.53–0.59

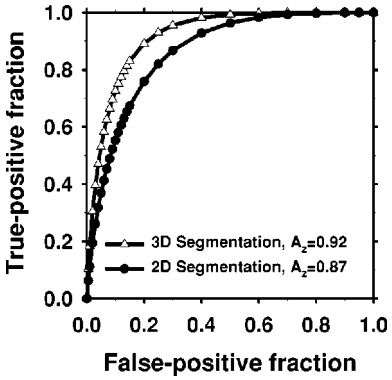

FIG. 6. The test ROC curves obtained by the classifiers that were based on features extracted from the 2-D ($A_z=0.87$) and 3-D ($A_z=0.92$) active contour models. The difference between the two A_z values did not achieve statistical significance ($p=0.07$).

TABLE III. The dependence of the computer classification accuracy on the variation of the initial contour. The effects of three transformation parameters, namely, scaling, translation, and rotation of the initial ellipsoid, was investigated by moving the initial ellipsoid using one of these three parameters at a time. A translation by ±10 pixels in the image plane corresponded to approximately ±1 mm.

Scale	Rotation (degrees)	*x*-translation (pixels)	*y*-translation (pixels)	A_z
1	0	0	0	0.92±0.03
1.3	0	0	0	0.89±0.03
0.8	0	0	0	0.89±0.03
1	0	10	10	0.90±0.03
1	0	10	−10	0.87±0.04
1	0	−10	10	0.87±0.04
1	0	−10	−10	0.88±0.03
1	15	0	0	0.93±0.02

width-to-height feature. Figure 6 shows the test ROC curves obtained by the LDA using leave-one-case-out resampling for the 2-D and 3-D segmentation methods. The test A_z values for the 2-D and 3-D methods were 0.87±0.03 and 0.92 ±0.03, respectively, and the $A_z^{(0.9)}$ values were 0.51±0.08 and 0.67±0.08, respectively. The difference between the two test A_z values did not achieve statistical significance ($p = 0.07$). Figure 7 shows the distribution of the discriminant scores obtained from the 3-D method for the malignant and benign cases.

In order to investigate the dependence of the classification accuracy on the initialization of the 3-D active contour model, we scaled, rotated, and translated the initial 3-D ellipsoid and repeated the steps of active contour segmentation, feature extraction, and classification for these modified initial ellipsoids. The classification accuracies for these experiments are presented in Table III. None of the differences between the A_z values on Table III achieved statistical significance.

The ROC curves for the radiologists' malignancy ratings are shown in Fig. 8. The computer and radiologist A_z values and $A_z^{(0.9)}$ values are compared in Table IV. The area A_z under the ROC curve for radiologists RAD1–RAD4 varied between 0.84±0.04 and 0.92±0.03, which are lower than or equal to that of the 3-D computer classifier. The average A_z value, obtained by averaging the slope and intercept parameters (*a* and *b* in a ROC analysis) of the individual ROC curves was 0.87. The difference between the A_z values of the individual radiologists and the computer classifiers (2-D and 3-D methods) did not reach statistical significance ($p > 0.05$). The $A_z^{(0.9)}$ values of the computer classifiers based on 2-D and 3-D segmentation were consistently higher than those of all four radiologists. The difference between the $A_z^{(0.9)}$ values of only one of the radiologists (RAD4) and the classifier based on 2-D segmentation achieved statistical significance ($p = 0.05$). The differences between the $A_z^{(0.9)}$ values of three of the four radiologists and that of the classifier based on 3-D segmentation were statistically significant ($p = 0.03$, 0.02, and 0.001 for RAD1, RAD2, and RAD4, respectively).

TABLE IV. The area under the ROC curve (A_z), and the area under the ROC curve above a sensitivity threshold of 0.9 ($A_z^{(0.9)}$) for the computer classifier using the 2-D and 3-D active contour segmentation results, and the four radiologists. The radiologists' results that are significantly ($p < 0.05$) different from the 3-D computer results are noted with an asterisk.

	A_z	$A_z^{(0.9)}$
Computer classifier, 2-D segmentation	0.87±0.03	0.51±0.09
Computer classifier, 3-D segmentation	0.92±0.03	0.67±0.08
RAD1	0.85±0.04	0.47±0.10*
RAD2	0.87±0.03	0.38±0.11*
RAD3	0.92±0.03	0.45±0.15
RAD4	0.84±0.04	0.28±0.11*

IV. DISCUSSION

The computer classifier designed in this study to characterize breast masses on US volumes was able to discriminate between malignant and benign masses that were suspicious enough to warrant a biopsy. From Fig. 7, it is observed that if an appropriate decision threshold was chosen for the discriminant scores of the classifier based on 3-D segmentation, more than 43% (20/46) of biopsied benign masses could be correctly identified while no malignant masses were misclassified (at 100% sensitivity). Based on 2-D segmentation, the corresponding percentage of correctly identified benign masses was 35% (16/46).

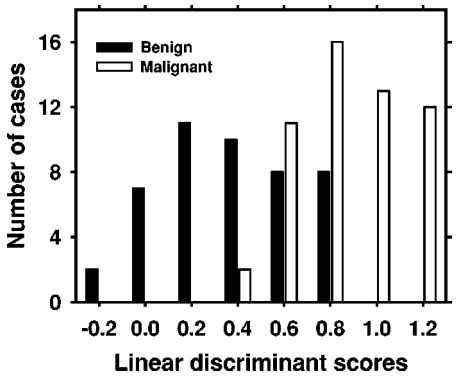

FIG. 7. The distribution of the test discriminant scores for the classifier that was based on 3-D active contour segmentation. By choosing an appropriate decision threshold on these scores (e.g., decision threshold=0.3) more than 43% (20/46) of biopsied benign masses could be correctly identified while no malignant masses would be misclassified.

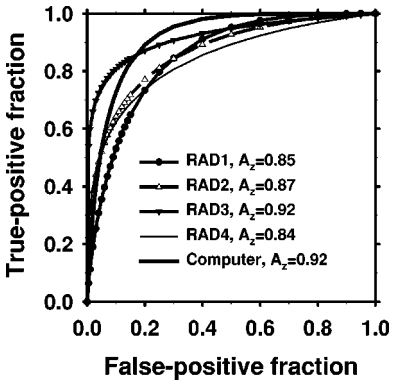

FIG. 8. ROC curves for the computer and for the four radiologists who participated in the malignancy rating experiment. The difference between the computer's A_z value and that of any of the four radiologists did not achieve statistical significance. However, the computer classifier had significantly higher ($p<0.05$) partial area index, $A_z^{(0.9)}$, than three of the four radiologists at high sensitivity (TPF$>$0.9).

Lesion segmentation is an important task in computerized lesion characterization. The segmentation of US images can be challenging because boundaries are not always conspicuous, due to the noise and contrast characteristics, and the speckled nature of US images. For breast US, an additional source of difficulty is the presence of posterior shadowing artifacts, a major source of which is the US attenuation due to the fibrous stroma caused by the tumor.[29] Previous research on the segmentation of breast masses on US images includes work by Horsch *et al.*,[30] Xiao *et al.*,[31] and Madabhushi *et al.*[32] Their segmentation methods were applied to 2-D US images. In our study, we compared the classification accuracy when 2-D and 3-D active contour models were used for segmentation. The 2-D model provided reasonable segmentation results for many of the masses. However, the 2-D model does not take advantage of the image information in adjacent slices when a particular slice is being segmented. If the 2-D active contour is misled on one slice, there is no interaction from adjacent slices to improve the segmentation. This is illustrated in Fig. 4, row 3. It can be observed that the 2-D segmentation results on slices #45 and #47 are reasonable; however, part of the lesion is missed by the 2-D active contour model on slice #46. Our 3-D active contour model uses the smoothness of the segmented shape in the out-of-plane direction as an interaction term between adjacent slices. The 3-D segmentation results, shown in row 4, are more consistent across slices. Figure 5 compares the segmented object using the 2-D and 3-D methods for the entire lesion, which was visible on a total of ten slices. It is again observed that the lesion shape in the out-of-plane direction is smoother for the 3-D method. Although our classification accuracy using the 3-D method was satisfactory, further improvement may be required for applications such as accurate lesion volume measurement. More sophisticated and inherently 3-D methods, such as deformable surfaces[33] and level set methods, may be good candidates for further improvement.

The texture features in this study were extracted from disk-shaped regions at the upper and lower margins of the mass on each slice. The total area of the two disk-shaped regions was equal to the area of the segmented mass. From Table I, it is observed that a texture feature extracted from the upper disk-shaped region tended to be more discriminatory than the same feature extracted from the lower disk-shaped region. The maximum of the range of A_z values (the second number in each cell) was larger for the upper region in 11 of the 12 comparisons that can be made (6 texture features and 2 segmentation methods). The lower boundaries of many masses were difficult to perceive and hence difficult to automatically segment because of posterior shadowing. This may have contributed to the difference of discrimination ability between the features extracted from the upper and lower regions. Another possible factor may be the changes in the spatial and gray level resolutions in different regions of the US image as the distance from the US probe increases. Further work is underway to investigate the reasons for the apparent lower discrimination ability of the features extracted from the lower disk-shaped regions.

Although the disk-shaped region depends on mass segmentation, there can be a large overlap between the regions from the 2-D and 3-D segmentation results if the objects segmented by the two methods are not very different. From Table I, it can be observed that the ranges of A_z values for 2-D and 3-D segmentation for each texture measure have a large overlap. As mentioned in Sec. III, when the stepwise feature selection method was used for classifier design from 2-D segmentation results, an average of 6.09 features were selected, where the average was computed over the 102 cycles of the leave-one-out partitioning of the dataset. Out of the six most frequently selected features, five were texture features and one was a morphological feature. The IMC1 feature was selected twice (at $d=2$, $\theta=0°$ and $d=6$, $\theta=90°$), the IMC2 feature was selected twice (at $d=2$, $\theta=0°$ and $d=6$, $\theta=0°$), and the DFE feature was selected once (at $d=6$, $\theta=0°$). For 3-D segmentation, out of the eight most frequently selected features, six were texture features, and two were morphological features. The IMC1 feature was selected twice (at $d=2$, $\theta=90°$ and $d=4$, $\theta=0°$), the IMC2 feature was selected twice (at $d=2$, $\theta=0°$ and $d=6$, $\theta=0°$), and the DFE feature was selected once (at $d=6$, $\theta=0°$). Thus, out of 11 most frequently selected texture features (5 for 2-D and 6 for 3-D segmentation), 10 were IMC1, IMC2, or DFE features. The classification accuracy with the stepwise LDA for the 3-D segmentation ($A_z=0.92$) was better than that for 2-D segmentation ($A_z=0.87$). However, the difference did not achieve statistical significance (a two-tailed p value$=0.07$).

The active contour method requires an initial boundary to start iterating toward the optimal contour. In this study, the initial boundary was defined by a 3-D ellipsoid that approximated the mass shape. The ellipsoid was placed in the volume by one of the radiologists (RAD1) using an interactive graphical user interface (GUI). The radiologist thus had to shift and scale a single object to define the initial contour. Although the error between the true and approximated shapes can be large when a single object is used for approxi-

mating the mass, this method was faster than other possible methods that would require initialization on each slice separately, and was therefore preferred. The robustness of the 3-D segmentation method to active contour initialization was studied by translating, rotating, and scaling the 3-D ellipsoid. There are many possibilities as to how these three operations (moving, rotating, and scaling) can be combined to modify the initial ellipsoid. In Table III, the classification results are presented when these three operations are performed one at a time. Row 1 shows the A_z value when the original ellipsoid is used. The ellipsoid was scaled in rows 2–3, translated in rows 4–6, and rotated in row 7. For the magnitudes of scaling, translation, and rotation studied in Table III, the variation of the A_z value was within two standard deviations of the A_z value provided by the LABROC program.[27] In a step toward automating the initialization of the contour, we are currently investigating methods for automatically determining an initial contour from a rectangular box containing the mass.

The comparison of the ROC curves by the radiologists and the computer indicated that the computer can be as effective as the radiologists in differentiating malignant and benign breast masses in this dataset. In fact, the accuracy of the computer classifier using 3-D segmentation was greater than three and equal to one of the radiologists, although the difference between the computer and the individual radiologists in terms of A_z did not achieve statistical significance. Furthermore, from Fig. 8, it is observed that the computer has a tendency to be better at high sensitivity. This was also confirmed by the statistically significant difference between the computer classifier (3-D segmentation method) and three out of the four radiologists when the comparison was based on the $A_z^{(0.9)}$ values. It should be noted that the purpose of our study was not to evaluate our US mass characterization method in a clinical setting. As noted in Secs. I and II, the semiautomated 3-D data acquisition system used in this study is still under investigation and is different from that in current clinical practice. The first difference is that, in our department, radiologists interactively perform handheld US examination themselves, which may yield better image quality and may result in higher characterization accuracy. The second difference is that our study concentrated only on mass characterization of lesions already detected, whereas the actual detection of suspicious masses by US is a very important step in a clinical examination. These other aspects of comparing 3-D US images to US images acquired with current clinical methods are subjects of future investigations.

In this study, the features were extracted from individual US slices and then combined into object-based features, as explained in Sec. II D. Although this method is found to provide effective features in this study, it may not have fully utilized the information available in the 3-D dataset. The potential improvement in classification accuracy by using truly 3-D features, for example, texture features extracted from 3-D SGLD matrices, needs to be investigated. Furthermore, in clinical practice, the decision about whether the mass is malignant or benign is made using both mammographic and US image information, as well as other pertinent patient information. A study is currently underway in our laboratory to design a classifier that combines computer-extracted features or scores from these two imaging modalities.

V. CONCLUSION

A computer segmentation and classification method has been developed for the task of the characterization of breast masses on 3-D US images. On a dataset of 102 biopsy-proven masses the classifier achieved an A_z value of 0.92. The average A_z value of four experienced radiologists on the same data set was 0.87. The computer classifier was more accurate than three and equal to one of the four radiologists participated in the study. However, the difference between the A_z values of the computer and the individual radiologists did not achieve statistical significance for this dataset. At high sensitivity, the computer classifier was consistently more accurate than all four radiologists and achieved statistical significance ($p<0.05$) for the difference in $A_z^{(0.9)}$ from three of the four radiologists. The robustness of the iterative segmentation algorithm in terms of the initial contour provided to the algorithm was studied. The classification accuracy was found to depend on the initialization; however, the A_z value did not significantly deteriorate when the initial contour was scaled, rotated, or translated by a moderate amount. Future work includes verifying the results of this study by applying it to a larger and independent dataset, expanding the feature space by designing truly 3-D features, and combining the developed US characterization method with mammographic characterization methods. The observer performance study will also be performed to evaluate the effects of CAD on the characterization of breast masses by radiologists.

ACKNOWLEDGMENTS

This work was supported by USAMRMC Grant No. DAMD 17-01-1-0328 and by USPHS Grants No. CA095153 and No. CA91713. The content of this publication does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned in the publication should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for providing the LABROC program.

[a]Author to whom correspondence should be addressed. Berkman Sahiner, Ph.D., Department of Radiology, University of Michigan, 1500 E. Medical Center Drive, CGC B2102, Ann Arbor, Michigan 48109-0904. Telephone: (734) 647-7429; fax: (734) 615-5513; electronic mail: berki@umich.edu

[1] L. W. Bassett, T. H. Liu, A. I. Giuliano, and R. H. Gold, "The prevalence of carcinoma in palpable vs impalpable, mammographically detected lesions," AJR, Am. J. Roentgenol. **158**, 688–689 (1992).

[2] H. Opie, N. C. Estes, W. R. Jewell, C. H. Chang, J. A. Thomas, and M. A. Estes, "Breast biopsy for nonpalpable lesions: a worthwhile endeavor?," Am. Surg. **59**, 490–493 (1993).

[3] G. Hermann, C. Janus, I. S. Schwartz, B. Krivisky, S. Bier, and J. G. Rabinowitz, "Nonpalpable breast lesions: Accuracy of prebiopsy mammographic diagnosis," Radiology **165**, 323–326 (1987).

[4] F. M. Hall, J. M. Storella, D. Z. Silverstone, and G. Wyshak, ''Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography,'' Radiology **167**, 353–358 (1988).
[5] J. A. Baker, P. J. Kornguth, J. Y. Lo, and C. E. Floyd, ''Artificial neural network: Improving the quality of breast biopsy recommendations,'' Radiology **198**, 131–135 (1996).
[6] Z. M. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, ''Automated computerized classification of malignant and benign masses on digitized mammograms,'' Acad. Radiol. **5**, 155–168 (1998).
[7] H.-P. Chan, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. S. Gopal, ''Improvement of radiologists' characterization of mammographic masses by computer-aided diagnosis: an ROC study,'' Radiology **212**, 817–827 (1999).
[8] 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 malignant and benign lesions,'' Radiology **196**, 123–134 (1995).
[9] P. Skaane and K. Engedal, ''Analysis of sonographic features in differentiation of fibroadenoma and invasive ductal carcinoma,'' AJR, Am. J. Roentgenol. **170**, 109–114 (1998).
[10] K. J. W. Taylor *et al.*, ''Ultrasound as a complement to mammography and breast examination to characterize breast masses,'' Ultrasound Med. Biol. **28**, 19–26 (2002).
[11] B. Sahiner, G. L. LeCarpentier, H. P. Chan, M. A. Roubidoux, N. Petrick, M. M. Goodsitt, S. S. Gopal, and P. L. Carson, ''Computerized characterization of breast masses using three-dimensional ultrasound images,'' Proc. SPIE **3338**, 301–312 (1998).
[12] D. R. Chen, R. F. Chang, and Y. L. Huang, ''Computer-aided diagnosis applied to US of solid breast nodules by using neural networks,'' Radiology **213**, 407–412 (1999).
[13] 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).
[14] C. M. Chen, Y. H. Chou, K. C. Han, G. S. Hung, C. M. Tiu, H. J. Chiou, and S. Y. Chiou, ''Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks,'' Radiology **226**, 504–514 (2003).
[15] D. B. Downey, A. Fenster, and J. C. Williams, ''Clinical utility of three-dimensional US,'' Radiographics **20**, 559–571 (2000).
[16] G. L. LeCarpentier, P. B. Tridandapani, J. B. Fowlkes, M. A. Roubidoux, A. P. Moskalik, and P. L. Carson, ''Utility of 3-D ultrasound in discriminating and detection of breast cancer,'' RSNA EJ http://ej.rsna.org/ej3/0103-99.fin/titlepage.html, 1999.
[17] P. T. Bhatti, G. L. LeCarpentier, M. A. Roubidoux, J. B. Fowlkes, M. A. Helvie, and P. L. Carson, ''Discrimination of sonographically detected breast masses using frequency shift color Doppler imaging in combination with age and gray scale criteria,'' J. Ultrasound Med. **20**, 343–350 (2001).
[18] M. Kass, A. Witkin, and D. Terzopoulos, ''Snakes: active contour models,'' Int. J. Comput. Vis. **1**, 321–331 (1987).
[19] D. J. Williams and M. Shah, ''A fast algorithm for active contours and curvature estimation,'' CVGIP: Image Understand. **55**, 14–26 (1992).
[20] L. D. Cohen, ''On active contour models and balloons,'' CVGIP: Image Understand. **53**, 211–218 (1991).
[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 characterization,'' IEEE Trans. Med. Imaging **20**, 1275–1284 (2001).
[22] B. S. Garra, B. H. Krasner, S. C. Horri, S. Ascher, S. K. Mun, and R. K. Zeman, ''Improving the distinction between benign and malignant breast lesions: The value of sonographic texture analysis,'' Ultrason. Imaging **15**, 267–285 (1993).
[23] R. M. Haralick, K. Shanmugam, and I. Dinstein, ''Texture features for image classification,'' IEEE Trans. Syst. Man Cybern. **SMC-3**, 610–621 (1973).
[24] Y. Zheng, J. F. Greenleaf, and J. J. Gisvold, ''Reduction of breast biopsies with a modified self-organizing map,'' IEEE Trans. Neural Netw. **8**, 1386–1396 (1997).
[25] P. A. Lachenbruch, *Discriminant Analysis* (Hafner, New York, 1975).
[26] N. R. Draper, *Applied Regression Analysis* (Wiley, New York, 1998).
[27] C. E. Metz, B. A. Herman, and J. H. Shen, ''Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data,'' Stat. Med. **17**, 1033–1053 (1998).
[28] Y. Jiang, C. E. Metz, and R. M. Nishikawa, ''A receiver operating characteristic partial area index for highly sensitive diagnostic tests,'' Radiology **201**, 745–750 (1996).
[29] B. Mesurolle, M. Ariche-Cohen, F. Mignon, J. Guinebretiere, A. Tardivon, and P. Goumot, ''Small focal areas of acoustic shadowing in the breast,'' J. Clin. Ultrasound **30**, 88–97 (2002).
[30] K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, ''Automatic segmentation of breast lesions on ultrasound,'' Med. Phys. **28**, 1652–1659 (2001).
[31] G. Xiao, M. Brady, J. Noble, and Z. Yongyue, ''Segmentation of ultrasound B-mode images with intensity inhomogeneity correction,'' IEEE Trans. Med. Imaging **21**, 48–57 (2002).
[32] A. Madabhushi and D. N. Metaxas, ''Combining low, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions,'' IEEE Trans. Med. Imaging **22**, 155–169 (2003).
[33] I. Cohen, L. D. Cohen, and N. Ayache, ''Using deformable surfaces to segment 3-D images and infer differential structures,'' CVGIP: Image Understand. **56**, 242–263 (1992).