Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms

Nicholas Petrick, Heang-Ping Chan, Berkman Sahiner, and Mark A. Helvie
The University of Michigan, Department of Radiology, CGC B2102, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109-0904

As an ongoing effort to develop a computer aid for detection of masses on mammograms, we recently designed an object-based region-growing technique to improve mass segmentation. This segmentation method utilizes the density-weighted contrast enhancement (DWCE) filter as a pre-processing step. The DWCE filter adaptively enhances the contrast between the breast structures and the background. Object-based region growing was then applied to each of the identified structures. The region-growing technique uses gray-scale and gradient information to adjust the initial object borders and to reduce merging between adjacent or overlapping structures. Each object is then classified as a breast mass or normal tissue based on extracted morphological and texture features. In this study we evaluated the sensitivity of this combined segmentation scheme and its ability to reduce false positive (FP) detections on a data set of 253 digitized mammograms, each of which contained a biopsy-proven breast mass. It was found that the segmentation scheme detected 98% of the 253 biopsy-proven breast masses in our data set. After final FP reduction, the detection resulted in 4.2 FP per image at a 90% true positive (TP) fraction and 2.0 FPs per image at an 80% TP fraction. The combined DWCE and object-based region growing technique increased the initial detection sensitivity, reduced merging between neighboring structures, and reduced the number of FP detections in our automated breast mass detection scheme. © *1999 American Association of Physicists in Medicine.* [S0094-2405(99)00808-1]

Key words: computer-aided diagnosis, digital mammography, breast mass detection, density-weight contrast enhancement, region growing

I. INTRODUCTION

Mammographic screening has proven to be an effective method for early detection of breast cancer. Women in a regular mammographic screening program have a statistically significant reduction in breast cancer mortality when compared to women not in such a program.[1] In addition, independent double reading by two radiologists has proven to significantly increase the sensitivity of mammographic screening.[2] Therefore, regular screening and double reading would appear to be a sensible approach for breast cancer detection. While regular screening is emphasized in health care programs, the higher cost and increased workload on the radiologists may make double reading by two radiologists impractical in a general screening situation. Computer-aided diagnosis (CAD) is one alternative that could allow a large number of mammograms to be double read by a single radiologist aided by the computer. This technique may improve the accuracy of both detection and characterization of breast lesions.

Many researchers have been interested in computerized analysis of mammograms[3] and a number of groups have developed algorithms for automated detection of breast masses. The detection of spiculated masses has been of particular importance because of its high likelihood of malignancy. Karssemeijer *et al.*,[4] Kobatake *et al.*,[5] and Kegelmeyer *et al.*[6] have all proposed methods for detecting spiculated masses on digitized mammograms. However, since a number of malignant masses are not spiculated, other groups have tackled the general problem of identifying all types of breast masses on digitized mammograms.[3,7–11]

Our research group has reported on a method for automatically detecting masses on digitized mammograms.[10,12] The method employed multiple stages of density-weighted contrast enhancement (DWCE) segmentation. The DWCE segmentation was first applied to the full mammogram, and then reapplied to local regions within the mammogram to improve object border definition. A final object splitting stage was employed to eliminate merging between neighboring or overlapping breast structures. False positive (FP) reduction based on extracted morphological features was applied after each segmentation step with texture analysis used as a final arbitrator between masses and normal structures. The segmentation was evaluated on 168 digitized mammograms and it achieved a performance of 4.4 FPs per image at a 90% true positive (TP) detection fraction and 2.3 FPs per image at an 80% TP detection fraction.[10]

Our approach to mass detection has been to first identify all significant structures within the breast region using a global segmentation technique and then refine the initial object borders using local processing. Finally, we differentiate between true masses and normal structures using morphological and texture information. Our method is therefore different from other detection algorithms that utilize the object shape information for initial detection. The disadvantage of

our combined global and local detection approach is that a large number of normal structures are identified in the initial stage. This can lead to additional FPs if the classification is suboptimal. However, the advantage of this approach is that it can identify difficult masses since the initial detection is not based on shape information. The shape information is still used in the classification stage to reduce FPs.

In this paper, we present an improved version of our two-stage DWCE segmentation approach. This new scheme was designed to both increase specificity and reduce the overall complexity of the segmentation. A primary motivation is to develop a method for eliminating the merging between neighboring structures in the local DWCE processing step and thus improve local segmentation. We introduce an object-based region-growing technique to perform this task. Improved local segmentation serves a number of purposes. First, it improves the morphological and texture information used for FP reduction as well as eliminates the need for the shape-based splitting step. It also enables us to eliminate two morphological FP reduction steps. This significantly reduces the overall complexity of the detection program and should lead to a more practical implementation in a general clinical setting. In this paper, we summarize the intermediate and overall detection performance of the improved mass segmentation algorithm and describe some of its limitations.

II. METHODS

A. Database

The clinical mammograms used in this study were selected from the files of patients who had undergone biopsy at the University of Michigan Hospital. The mammograms were acquired with American College of Radiology (ACR) accredited mammography systems. Kodak MinR/MRE screen/film systems with extended cycle processing were used as the image recorder. The mammography systems have a 0.3-mm focal spot, a molybdenum anode, 0.03-mm thick molybdenum filter, and a 5:1 reciprocating grid. The selection criterion used by the radiologists was simply that a biopsy-proven mass existed on the mammogram. The data set consisted of 253 mammograms from 102 patients, and it included 128 malignant and 125 benign masses. Sixty-three of the malignant and six of the benign masses were judged to be spiculated by a MQSA approved radiologist. The size of the masses ranged from 5 to 29 mm (mean size=12.5 mm), and their visibility ranged from 1 (obvious) to 5 (subtle) (mean=2.1). Figures 1 and 2 show the histograms of mass size and mass visibility for the data set.[13] These distributions characterize the difficulty and diversity of the cases contained in the data set.

The mammograms were digitized with a LUMISYS DIS-1000 laser film scanner with a pixel size of 100 μm and 12 bit gray level resolution. The gray levels were linearly proportional to optical density in the 0.1 to 2.8 optical density unit (O.D.) range. The slope was 0.001 O.D./pixel value. The slope gradually fell off in the 2.8 to 3.5 O.D. range.[10,13] A large pixel value corresponds to a low optical density with this digitizer.
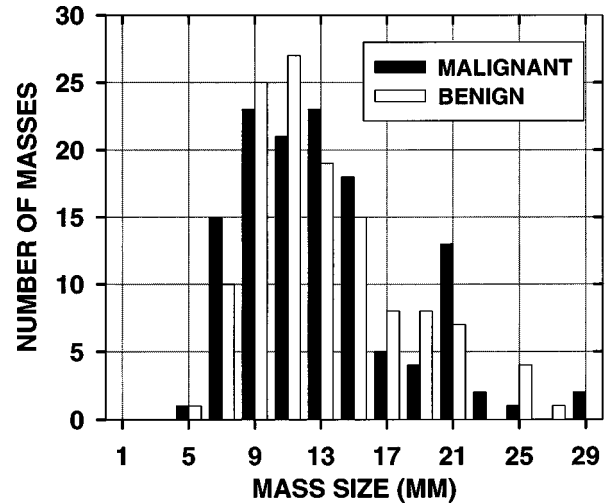

FIG. 1. Histograms of mass size for the 253 masses contained in our data set. Mass sizes were measured as the largest axis of the mass by an experienced breast radiologist.

The location and extent of all the biopsy-proven masses were marked on the original films. The radiologist then identified both the centroid of the lesion and the smallest bounding box containing the entire lesion using an interactive image manipulation tool on a workstation. Both procedures were performed using the original marked film as a guide. The lesion centroid was used to identify TP detections after the morphological FP reduction step. If a segmented object was within 4 mm of the mass centroid, it was considered a TP. All other segmented objects were considered as FPs. The final free-response receiver operating characteristic (FROC) curves following texture-based classification used the more precise mass bounding box for TP identification. A region was considered a TP only when it contained more than 50% of the mass bounding box.
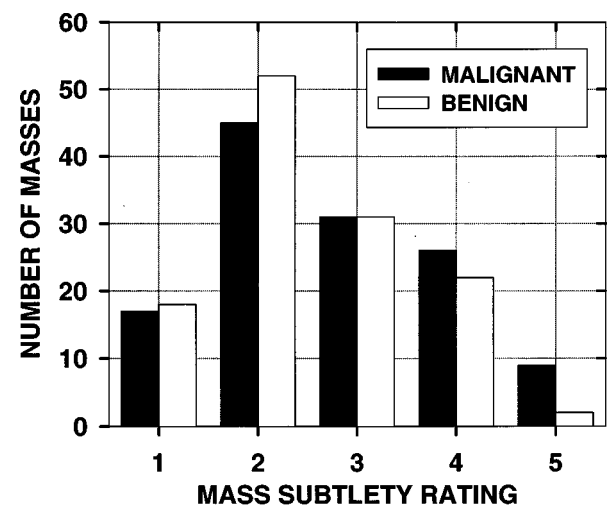

FIG. 2. Histograms of mass subtlety for the 253 masses contained in our data set. Mass subtleties were rated by an experienced breast radiologist from 1 (obvious) to 5 (subtle).
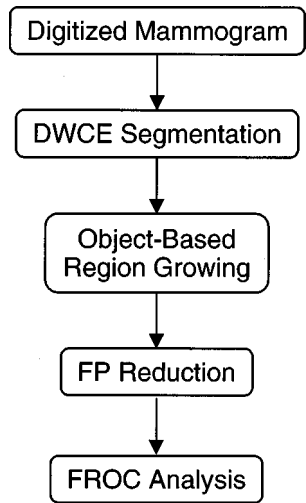
FIG. 3. Block diagram of the breast mass segmentation scheme. A digitized mammogram undergoes DWCE segmentation followed by object-based region growing and then morphological and texture classification. The performance of the segmentation scheme was evaluated by FROC analysis.

B. Density-weighted contrast enhancement segmentation

The block diagram for the proposed detection scheme is shown in Fig. 3. Global DWCE segmentation was used to identify an initial set of breast structures on the digitized mammograms. These objects were then used as seed locations to perform gradient-based region growing. A thorough description of the DWCE technique can be found in the literature.[10,12,14] Briefly, the DWCE technique employs an adaptive filter to enhance the local contrast and thus accentuate mammographic structures in an image. As the term implies, the parameters of the enhancement filter are based on the local density within the image and the filter is applied to the image on a pixel-by-pixel basis. The filter is designed to suppress very low contrast values, to emphasize the low to medium contrast values and to just slightly deemphasize the high contrast values. The effect of suppressing the extremely low contrast values is to reduce bridging between adjacent breast structures. Pixels with low to medium contrast values are enhanced so that more subtle structures can be detected. Finally, the slight deemphasis of the high contrast structures is included to provide a more uniform intensity distribution for detected structures. After contrast enhancement, Laplacian–Gaussian edge detection is applied and all enclosed objects are filled to produce a set of detected structures for the image. The DWCE segmentation is applied to mammograms that have been smoothed and subsampled from their original 100 μm pixel size to an 800 μm pixel resolution.[10] The DWCE stage has been found to be effective in detecting most breast structures including a significant portion of breast masses. However, the DWCE borders usually fall well inside the true borders of an object and a significant number of adjacent structures are merged into single objects. This occurs most frequently when the adjacent breast structures have some tissue overlap.

C. Object-based region-growing segmentation

1. Initial gray-scale region growing

Before gradient-based region growing was applied, an initial set of seed objects was identified. This was accomplished by first identifying all local maxima in the original gray-scale image which occurred within the extent of the DWCE objects. Local maxima were defined using the ultimate erosion technique described by Russ.[15] In simple terms, a pixel was a local maximum if and only if its value was at least as large as all nearest neighbor pixel values. All maxima were identified and grown into larger objects by a simple gray-scale region growing technique as follows. Gaussian smoothing ($\sigma=2.0$) was applied to the gray-scale image, and a maximum and a minimum pixel value threshold were specified to select a range of acceptable pixel values. The thresholds were defined as

$$G_i^{\max_1} = 1.01 G_i^{\mathrm{UEP}} \tag{1}$$

and

$$G_i^{\min_1} = 0.99 G_i^{\mathrm{UEP}}, \tag{2}$$

where G_i^{UEP} was the pixel value of the ith maximum and $G_i^{\max_1}$ and $G_i^{\min_1}$ were the maximum and minimum pixel value thresholds, respectively. All pixels within a radius of 20 pixels from a maximum location and with a pixel value inside the defined range were considered to be part of the object. This was repeated for all maxima within an image. Figures 4(a)–4(d) show an original gray-scale image and corresponding images with the DWCE objects, the local maxima, and the gray-scale region-grown objects highlighted. The expanded objects were used as seeds for the gradient-based region growing, described below.

2. Gradient images

A mammogram at 200 μm resolution was used in the gradient-based region-growing stage. The 200 μm resolution image was obtained by averaging 2×2 pixels from the original image. The reduced resolution image had to be smoothed again before gradient filtering because the mammographic tissue produced gradients not only within individual breast structures but also throughout the background portions of the image. Figure 5(b) shows the gradient magnitude image resulting from vertical and horizontal Sobel filtering applied to the 200 μm gray-scale image shown in Fig. 5(a). It clearly demonstrates the large number of gradients throughout the image and the difficulty in applying object-based region growing without additional smoothing. For our application, the smoothing needed to reduce the spurious gradients was accomplished by frequency-weighted Gaussian (FWG) filtering. Frequency-weighted filtering is a technique in which all pixels within the image are split into a base and a residual term. The residual is either positive or negative. This technique produces three subimages from an original image, F, where

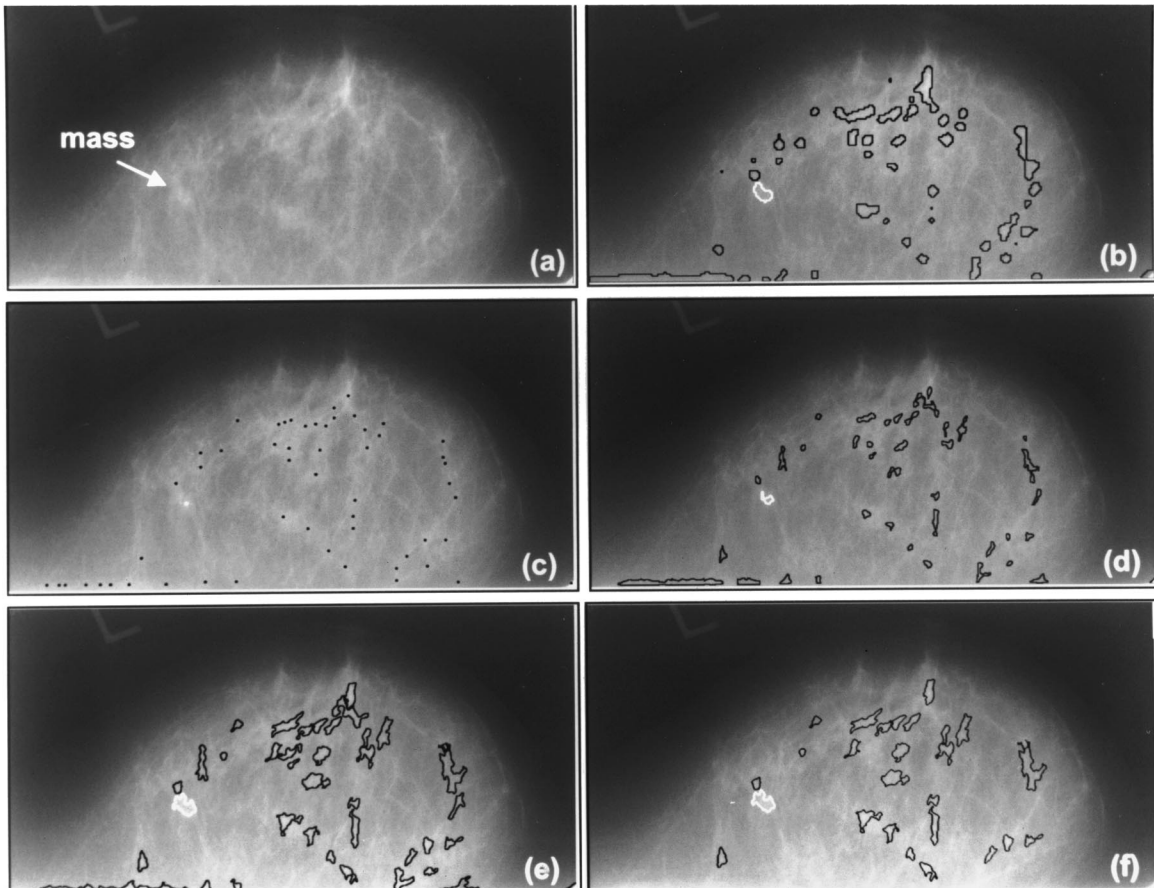$$F = F_F + F_{\mathrm{sub}+} + F_{\mathrm{sub}-}. \tag{3}$$

FIG. 4. Objects produced by each segmentation step for a typical mammogram from our data set: (a) the original mammogram with the mass location identified, (b) the DWCE objects, (c) the local maxima, (d) the objects obtained with gray-scale region growing, (e) the objects obtained with gradient-based region growing, and (f) the objects remaining after morphological FP reduction.

The first filter component, F_F, is a filtered version of the original image. In our case, a Gaussian filter, $G(\mu=0,\sigma=10)$, was used. The second and third images are the positive and negative residual images of $F-F_F$, respectively. The F_{sub+} residual is nonzero where the image intensity is larger than the local background and F_{sub-} is nonzero where the image intensity is smaller than the local background. For a particular image pixel, (x,y), the residual images are defined as

$$F_{sub+}(x,y) \equiv \begin{cases} F(x,y)-F_F(x,y), & F(x,y)>F_F(x,y), \\ 0, & \text{otherwise,} \end{cases}$$
(4)

and

$$F_{sub-}(x,y) \equiv \begin{cases} F(x,y)-F_F(x,y), & F(x,y)<F_F(x,y), \\ 0, & \text{otherwise.} \end{cases}$$
(5)

Two FWG filters were designed for sequentially processing the mammograms. The first FWG filtering step reduced the gradients within the breast structures and produced an intermediate image, F_1, which had the form

$$F_1(F) = \tfrac{3}{4}F_F(F) + \tfrac{1}{4}F_{sub+}(F),$$
(6)

where the F_F and F_{sub+} images were derived from F, the original 200 μm resolution gray-scale image. A second FWG filtering step was used to eliminate gradients in the breast background. It produced image F_2, which had the form

$$F_2(F_1) = F_{sub+}(F_1),$$
(7)

where the F_{sub+} image was derived from image F_1. The result of applying the two FWG filters to the original mammogram in Fig. 5(a) is shown in Fig 5(c). In this image, a significant amount of background has been eliminated and the gradients in the remaining structures have been reduced. Horizontal and vertical Sobel filters[15] were then applied to image F_2 and the magnitude calculated to produce a gradient image as shown in Fig. 5(d). Finally, 5×5 median filtering was used to produce the final gradient image shown in Fig. 5(e). This image was used in the gradient-based region-growing step.

3. Final gradient-based region growing

Each initially grown object (described in Sec. II C 1) was again grown by applying an adaptive technique to the gradient image, F_2, described in Sec. II C 2. The region-growing technique was based on the work of Chang and Li[16] and their adaptive homogeneity test for determining the similarity be-
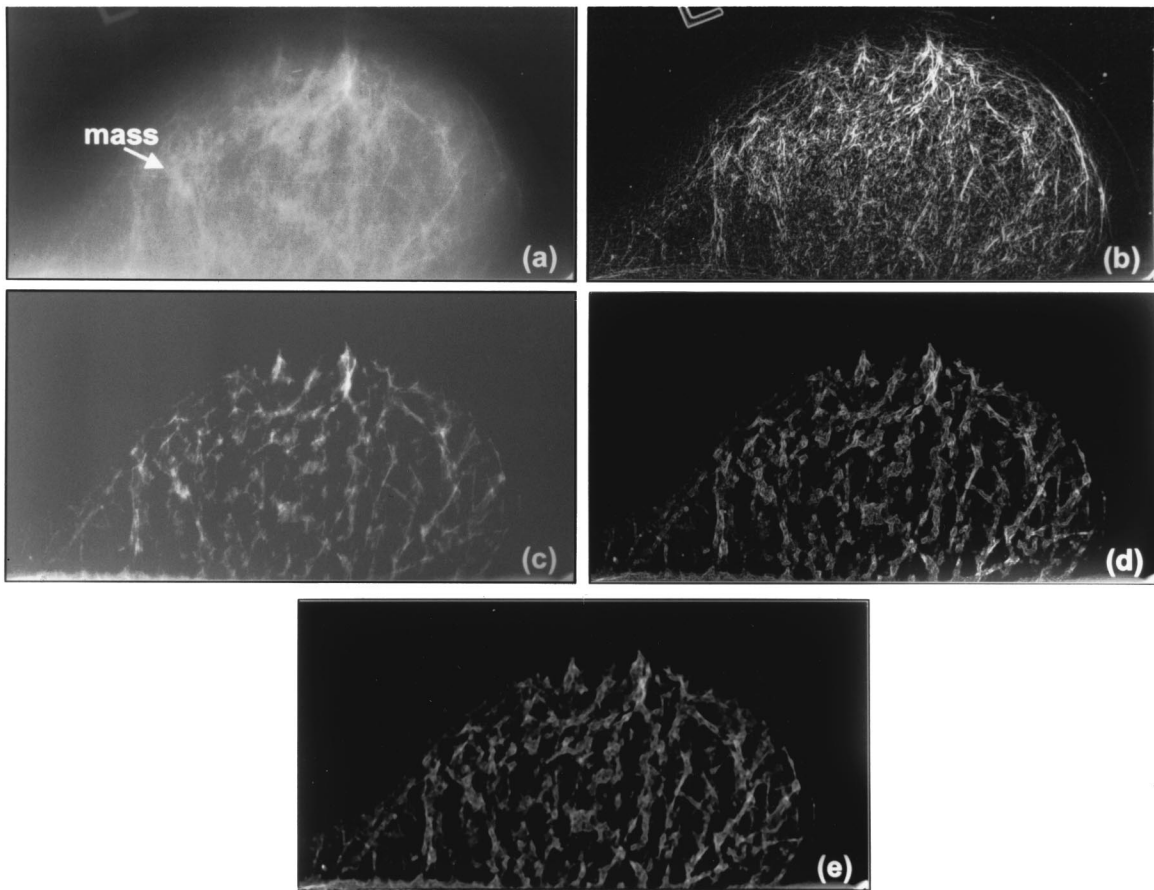
FIG. 5. Processing steps used to define the gradient images: (a) the original mammogram with the mass location identified; (b) the gradient magnitude image obtained from horizontal and vertical Sobel filtering of the original mammogram; (c) the image resulting from FWG filtering of the original mammogram; (d) the gradient magnitude image resulting from horizontal and vertical Sobel filtering of the FWG image; and (e) the image resulting from median filtering of the gradient magnitude image.

tween regions. We have modified this technique to perform object-based region growing. For a mammogram, the corresponding gradient image was smoothed using a Gaussian filter ($\sigma=2.0$). A cumulative distribution function (CDF) of pixel values was then calculated from the smoothed gradient image for each object. For each object, the pixel value thresholds were defined as

$$G_{i,0}^{\max_F} = \{g : \text{CDF}_{i,0}(g) = 1.0\} \tag{8}$$

and

$$G_{i,0}^{\min_F} = \{g : \text{CDF}_{i,0}(g) = 0.0\}, \tag{9}$$

where g was a pixel value and $\text{CDF}_{i,0}(g)$ was the cumulative pixel value distribution within the border of object i and for initial growing iteration 0. The initial growing thresholds simply correspond to the maximum and minimum pixel values within an object. Single-pixel growing was performed on all objects using the thresholds for each individual object to define a range of acceptable pixel values. In this context, single-pixel growing meant growing was limited to only those pixels directly connected to the initial border. Once single-pixel growing was applied to all objects within the image, the thresholds were adjusted and a second iteration of growing was performed. Iterative single-pixel growing was

employed to limit the influence of the order that objects were grown within an image. The thresholds used for the ith object during the jth growing iteration were defined as

$$G_{i,j}^{\max_F} = \{g : \text{CDF}_{i,j}(g) = 1.0\} \tag{10}$$

and

$$G_{i,j}^{\max_F} = \left\{ g : \text{CDF}_{i,j}(g) = \frac{j}{30} \right\}, \tag{11}$$

where $\text{CDF}_{i,j}(g)$ was the cumulative pixel value distribution from the smoothed gradient image within the current borders of object i. Single pixel growing was applied to all objects within the image. This iterative procedure was repeated until no more connected pixels had a value within the appropriately defined range. Note that neighboring objects were not allowed to merge together during this region-growing stage so that growing between adjacent objects stopped with at least a one pixel gap between them. Figures 4(d) and 4(e) show the initial seed objects and the final gradient grown objects for the example shown in Fig. 4(a).
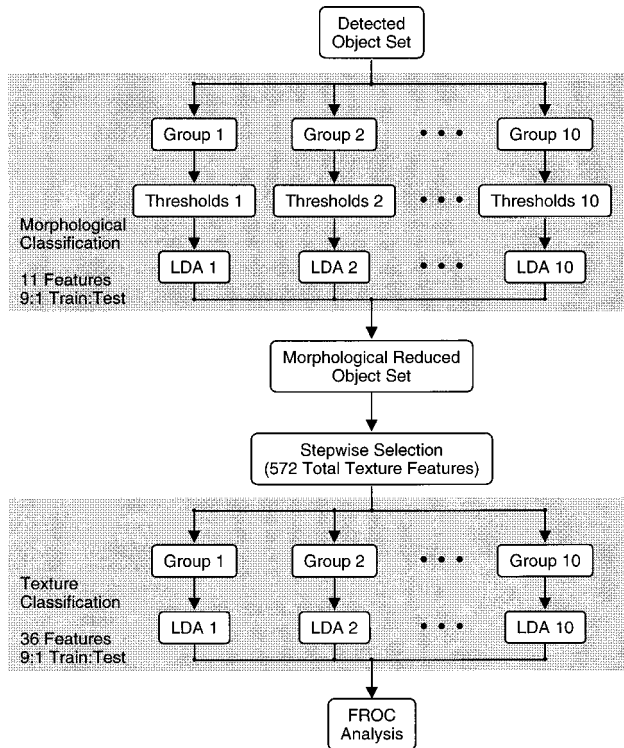
FIG. 6. Flowchart of the FP reduction scheme. The images were separated into ten independent groups. Each group underwent morphological FP reduction with the nine other groups used for classifier training. The reduced objects were recombined and stepwise feature selection was performed. The images were again separated into the ten groups and each group underwent LDA texture classification again using the nine other groups for classifier training. All test scores were then recombined and final FROC analysis was performed.

D. False positive reduction

The DWCE segmentation and region growing do not differentiate masses from normal tissues, therefore, a large number of breast structures were usually detected in each mammogram. Since the shape and texture of mass objects, in general, should be different from those of normal breast structures, a set of features was extracted from each detected object and used to differentiate between the detected structures. The feature set included both morphological and texture features. These features were then used in a sequential classification scheme to reduce the number of FP detections in the mammograms. The sequential application of different classifiers has been found to increase classification accuracy,[17] and it also allows more computationally intensive classifiers to be applied to as few objects as possible. A flow chart depicting the general approach employed for FP reduction is shown in Fig. 6. In this study, morphological classification was initially used to eliminate objects that had shapes significantly different from breast masses. Texture features were then computed for all remaining objects and used with a linear classifier as a final arbiter between masses and normal structures. The following sections describe the major components of the FP reduction scheme.

1. Morphological feature-based FP reduction

The mammograms were partitioned into a number of different groups so that the morphological classifiers could be trained and tested to differentiate masses from normal structures. In this study, the 253 mammograms were randomly partitioned into ten independent groups. Each mammogram was allowed to appear in only one group, and all images from the same patient were grouped together. The goal of the partitioning was to have approximately the same number of images in each group under the given constraints. Classification of the objects within each individual group was performed with a classifier trained using the objects from the nine other image groups. This allowed an approximate 9:1 training-to-test ratio for morphological classification. By rotating the test group through all ten image sets, each mammogram served as a test case once.

Eleven morphological features were used in the initial differentiation of the detected structures. These features included the following object-based measures: number of perimeter pixels, area, perimeter-to-area ratio, circularity, rectangularity, and contrast. In addition, five normalized radial length (NRL) features introduced by Kilday *et al.* were also utilized.[18] They included the NRL mean value, standard deviation, entropy, area ratio, and zero-crossing count. The definition for each morphological feature can be found in the literature.[10] They are also included in Appendix A of this paper.

The morphological features were used as input variables for two different classifiers. A simple threshold classifier was followed by a linear discriminant analysis (LDA) classifier in the morphological FP reduction step. The simple threshold classifier set a maximum and minimum value for each morphological feature based on the maximum and minimum feature values found from the breast masses in the data set. The LDA classification was applied to all objects remaining after threshold classification. The LDA classifier is a linear classifier based on Fisher's discriminant, which is optimal for the two-class, multivariate normal, equal covariance problem.[19,20] The LDA classifier was trained for each training set and applied to the appropriate test set. The LDA classifier produced a single discriminant score for each object in the test set. A threshold was defined as the maximum discriminant score of the masses. This threshold was applied to the test set to further differentiate breast masses for normal structures. The threshold was again based on all masses in the data set to ensure that no mass would be lost during this initial stage. Figure 4(f) shows the results of morphological FP reduction for the example depicted in the figure.

2. Texture feature-based FP reduction

Texture-based classification followed the morphological FP reduction. A large set of multiresolution texture features was extracted for each detected object in the mammogram. Stepwise feature selection was then used to choose the most appropriate set of features for linear classification. The selected features were subsequently used with a LDA classifier to produce a single discriminant score for each detected ob-

TABLE I. The number of detected masses and FPs, the single stage reduction, the mean object area (μ_{Area}), and standard deviation of the object areas (σ_{Area}) for the initial stages in the mass detection scheme. Note texture FP reduction followed the morphological FP reduction stage.

Stage	TPs fraction	FPs/image (initial stages)	Reduction	μ_{Area} (mm^2)	σ_{Area} (mm^2)
DWCE	97%	49.1	⋯	33.6	66.8
Region growing	97%	45.3	0%	52.4	85.1
Morph. FP reduction	97%	35.5	22%	51.9	52.1

ject. The overall performance of the detection scheme was then evaluated with FROC analysis. The texture-based reduction scheme has been documented in the literature; therefore, this paper will only summarize the important components of the texture analysis and point out any differences from the previously described techniques.[10,21,22]

Regions of interest (ROIs) containing each object remaining after morphological FP reduction were extracted from the 100 μm resolution mammograms. The ROIs had a fixed size of 256×256 pixels and the center of each ROI corresponded to the centroid location of a detected object. The only exception was when the object was located near the border of the breast and a complete 256×256 pixel ROI could not be defined. In this case the ROI was shifted until the appropriate edge coincided with the border of the original mammogram.

Global and local multiresolution texture features, based on the spatial gray level dependence (SGLD) matrix,[23,24] were used in texture analysis.[22] An element of the SGLD matrix, $p_{d,\theta}(i,j)$, is defined as the joint probability that gray levels i and j occur at a given interpixel separation d and direction θ. In this study, 13 texture measures were defined for each SGLD matrix. These measures were correlation, energy, entropy, inertia, inverse difference moment, sum average, sum variance, sum entropy, difference average, difference variance, difference entropy, information measure of correlation 1, and information measure of correlation 2. The definition for all texture measures can be found in the literature[22] and are included in Appendix B of this paper.

The wavelet transform with a four-coefficient Daubechies kernel was used to decompose individual ROIs into different scales. For global texture features, four different wavelet scales, 14 different interpixel distances and 2 different angles were used to produce 28 SGLD matrices. This resulted in 364 global multiresolution texture feature for each ROI. To further describe the information specific to the mass and its surrounding normal tissue, a set of local texture features were calculated for each ROI.[10,22,25] Five rectangular subregions were segmented from each ROI; an object subregion defined by the detected object in the center and four peripheral regions at the corners. Eight SGLD (four interpixel distances and two angles) and a total of 208 local features were calculated from the object subregion and the periphery. They included 104 features in the object region and an additional 104 features defined as the difference between the feature values in the object and the periphery.

In order to improve the generalization of the texture clas-

sification, stepwise feature selection was used to select a subset of feature from the pool of 572 global and local features. Feature selection was performed using texture features derived from the ROIs obtained from all 253 images. A total of 40 texture features were selected by stepwise feature selection. Details on the application of stepwise feature selection can be found in our previous publications.[21,26]

At this point in texture classification, the mammograms were again divided into the same ten partitions as described in the morphological FP reduction step. Texture classification was performed on each test group with a trained LDA classifier employing the selected features. The training was based on the texture features derived from the ROIs in the nine other image groups. The test scores within each group were combined with the scores from the other groups to form a complete test set of discriminant scores.

The FROC analysis based on the single set of test scores was used to evaluate the overall performance of the segmentation method.[27,28]

III. RESULTS

The number of TP and FP detections found following the DWCE, region-growing, and morphological FP reduction stages of the segmentation algorithm are summarized in Table I. The DWCE segmentation identified 97% of the breast masses. Table I also includes the reduction percentage, the mean object areas (μ_{Area}) and the standard deviations in the object areas (σ_{Area}) for these initial stages. Table II summarizes the mass type, mass size, mass subtlety, and the

TABLE II. The mass type, mass size, mass subtlety, and mammographic tissue density for the mammograms where the mass was not identified by the initial segmentation. In the table, B dentifies a benign lesion, M identifies a malignant lesion, the subtlety is on a scale of 1 (obvious) to 5 (subtle), and breast density uses the BIRADS density scale of 1 (fatty) to 4 (dense). Both the subtlety and density rankings were performed by an experienced breast radiologist.

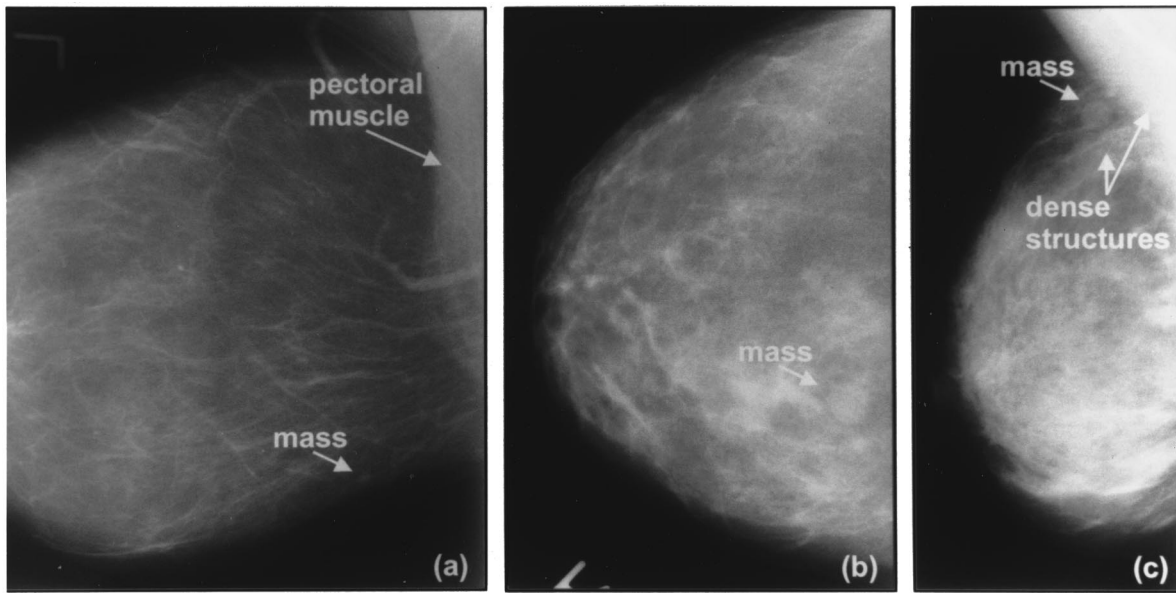| Mass no.	Type	Size (mm)	Subtlety	Breast density
1	M	6	4	1
2	B	10	2	1
3	B	14	2	2
4	B	10	2	3
5	B	10	2	3
6	B	14	2	3
7	B	12	4	4
Average		10.9	2.6	2.4

FIG. 7. Examples of masses missed during the initial DWCE segmentation stage: (a) a mammogram with a dense pectoral muscle, fatty breast tissue, and a subtle malignant mass (mass 1 in Table II); (b) a mammogram containing a low contrast benign mass (mass 3 in Table II); and (c) a mammogram with dense structures next to a lower contrast benign mass (mass 4 in Table II).

overall mammographic tissue density for the seven masses missed during the initial DWCE segmentation stage. Figure 7 shows examples of the cases where the mass was missed during the DWCE stage. Figure 8 shows example images

with corresponding gradient and object images for cases that had problems during the region-growing stage. This figure contains an example where the mass stopped growing before it reach the correct edge, and an example where the mass was
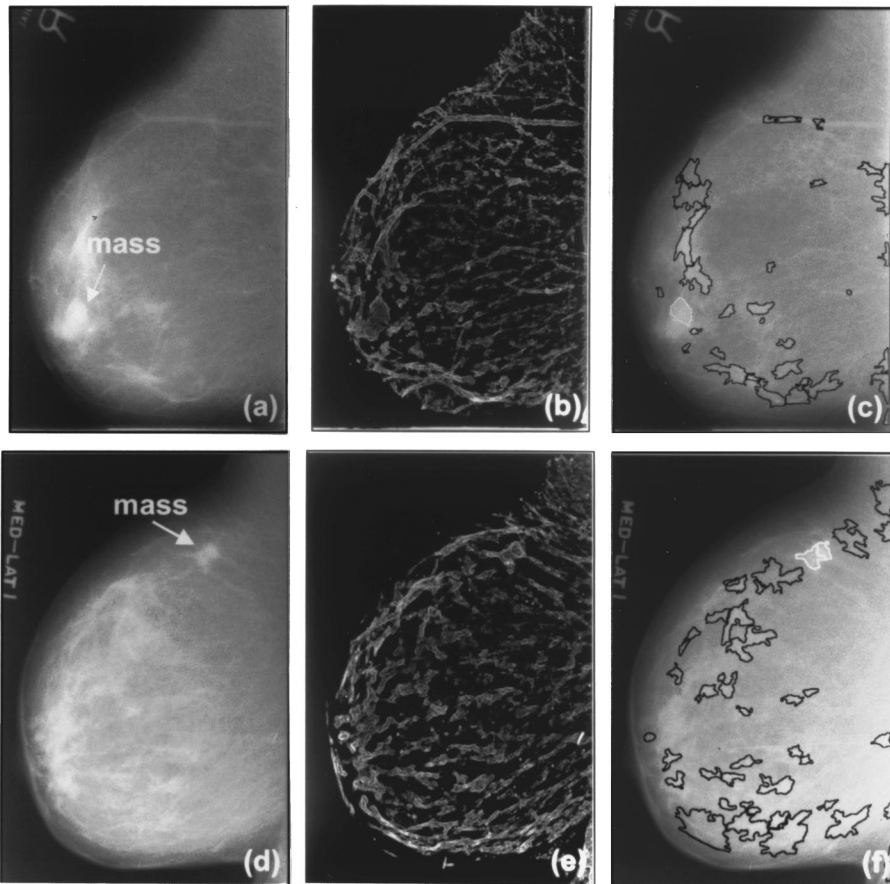

FIG. 8. A mammographic case containing a mass that stopped growing before it reached the correct edge (a)–(c) and a case containing a mass that was split into two pieces during growing (d)–(f). This figure includes (a) and (d) the original mammograms with the mass locations identified, (b) and (e) the corresponding gradient images, and (c) and (f) the final grown objects.
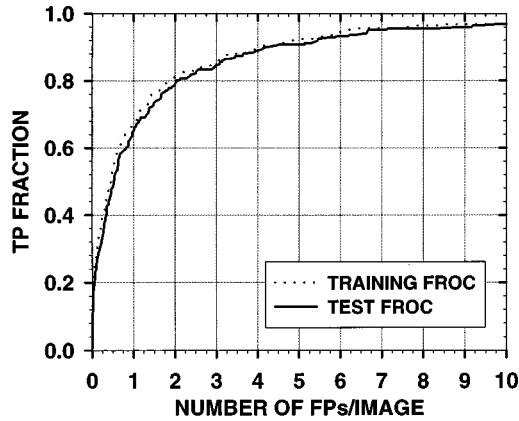
FIG. 9. The training and test FROC curve obtained following LDA classification using 40 selected texture features. The training scores were obtained by averaging the nine training scores from each detected object. The FROC data points were obtained by varying the discriminant decision threshold from the maximum to the minimum value.

split into two pieces during region growing. Finally, Fig. 9 show the FROC training and test performance for the complete segmentation scheme. A summary of the overall performance is given in Tables III and IV for a number of different TP detection fractions. The test performance for the combined DWCE and region-growing segmentation technique at a 90% TP detection level was 4.2 FPs per image and 2.0 FPs per image at an 80% TP level.

IV. DISCUSSION

The purpose of the initial DWCE segmentation stage was to have a method sensitive enough to identify breast masses but which also limited the number of normal structures detected. We have found the DWCE segmentation to be effective in this task. In this study, DWCE segmentation identified 246 of the 253 (97%) masses in the images. Table II summarizes the properties of the masses missed in DWCE segmentation. Masses 1 and 2 were missed because of a dense pectoral muscle visible on the mammogram which overwhelmed all lower-density structures (i.e., both mammograms had BIRADS category 1 breast density). The dense pectoral muscle caused the lower level of the DWCE intensity range to be set so high that lower intensity structures were missed. Figure 7(a) shows the mammogram of the missed malignant mass (mass 1 from Table II). The pectoral muscle is much denser than the mass. This led to the miss. One possible method for eliminating this type of miss may be to identify the pectoral muscle in the mammogram and to apply DWCE segmentation to only the remaining breast region. Mass 3 in Table II was missed because of the small contrast difference between the mass and the background tissue even though the mass was not particularly small or subtle. The mammogram containing this mass is depicted in Fig. 7(b). The remaining masses were missed in mammograms containing denser breast tissue. It was observed that DWCE segmentation had problems detecting masses that were located near much denser normal structures. The dense

structures were detected but the masses were missed. Figure 7(c) shows an example of this type of miss. It shows the mammogram containing mass 4 from Table II. Again the dense pectoral muscle may have also hindered detection of the mass in this case. Other than these problems, the DWCE segmentation performed reasonable well as a first stage in mass segmentation. It could identify the majority of the masses while eliminating many of the lower contrast background structures. However, the DWCE segmentation usually underestimated the actual borders of most structures. It also had a tendency to merge the mass with neighboring structures that may have had some tissue overlap with the breast mass. A total of 48 masses had significant merging between the mass and adjacent tissues after DWCE segmentation. This limited the effectiveness of the morphological FP reduction step and limited the localization of the mass during texture-based classification.

The region-growing stage reduced the effects of object merging and significantly increased the size of the initial DWCE objects. This is clearly shown in Table I where the average size of a structure increases from 33.6 mm^2 with DWCE alone to 52.4 mm^2 following region growing. Likewise, a comparison of objects from Figs. 4(b) and 4(e) shows the improvement in border definition following region growing. A combination of gray-scale and gradient-based region growing was used because of the difficulty in stopping gray-scale region growing at the correct edge and the need for large seed objects in gradient-based region growing. The combination approach performed adequately in our detection task and led to an improvement in both morphological and texture-based FP reduction. However, some problems were observed. One problem was that small and low-contrast structures had a tendency to grow into the background and become large regions even though the actual structures were quite small. This did not occur with masses, but it did occur with other breast structures. Another problem was that structures containing internal gradients did not always grow to the correct border, but ended up containing only a section of the true object. This occurred to some mass objects and led to either inaccurate structural information or a mass being split into multiple pieces. Figure 8 shows an example of both incomplete growing and a mass split into pieces during region growing. While these problems reduced the effectiveness of the morphological FP reduction, we have found that the overall benefit of region growing outweights its drawbacks and leads to an improvement in detection accuracy with our segmentation scheme.

The final step in the segmentation was FP reduction. Morphological feature classification was performed first in our reduction scheme. The morphological classification reduced the number of FPs per image from 45.3 to 35.5 as shown in Table I. Following morphological reduction, the average size of the objects was similar to the average size before reduction, but the standard deviation in object size fell from 85.1 mm^2 before reduction to 52.1 mm^2 after reduction. This indicates that morphological reduction eliminated objects that were either much larger or much smaller than the average object size, but had trouble differentiating between TPs and

TABLE III. Summary of the training FROC result depicted in Fig. 9. The table contains the number of FPs per image for different TP fractions along with the percentage of FPs reduced at each TP level relative to the initial value of 19.4 FPs per image. The first entry in the table is the reduction achieved without missing any additional breast masses.

TP fraction	FPs/image	FP reduction
98%	19.4	0%
95%	6.1	69%
90%	4.0	79%
80%	1.9	90%

FPs of similar sizes. Therefore, a classifier that can better differentiate between these similar shaped objects was still necessary. This was achieved, to a large extent, with texture-based feature classification.

A LDA, classifier based on SGLD texture features extracted from ROIs defined by each detected object has proven to be effective in differentiating between similar shaped objects. The training and test FROC performance curves following final texture classification are shown in Fig. 9. In addition, the number of FPs per image for different TP fractions are given in Tables III and IV for the two curves. As discussed in the Methods section, the mammograms were divided into ten independent groups and a 9:1 training-to-test ratio was employed in the classification. Therefore, the test value for an object was its single testing score, and its training value was the average of the scores obtained for the object during training with the nine different training group combinations. The first point to note in Tables III and IV is that the initial TP detection fraction has increased from 97% in Table I to 98% (i.e., 247 total masses were detected). This is due to the change in the definition of a TP with the texture ROIs. The additional mass was detected because in one of the seven mammograms where no object contained the mass centroid, an object ROI overlapped with at least 50% of the mass. The texture classification was able to reduce the number of FPs per image from an initial value of 35.5 to approximately 19 without the loss of any TPs, achieving a 45% reduction. While the number of FPs is still large, it indicates that the more computationally intensive texture classification performs better than morphological reduction. Additional reduction in FPs can be achieved with lower TP detection thresholds. For example, at a 90% TP fraction the FPs decreased to 4.2 per image and at an 80% TP level the FPs decreased to 2.0 per image. Comparing with our previously

reported two-stage DWCE edge detection segmentation technique[10] (discussed in Sec. I), we obtained improved performance at all TP levels despite the fact that the data set was increased from 168 to 253 mammograms and two fewer FP reduction stages were used with the new segmentation technique.

The results presented in this paper do not reflect results from a completely independent test set because the feature selection and the selection of morphological classification thresholds were based on the entire image set. This was necessary to obtain the best possible mass statistics from our limited data set at the intermediate stages of the algorithm. A database is currently being collected so that completely independent testing can be performed using the proposed method.

V. CONCLUSION

We have reported on an improved version of a breast mass detection scheme. The scheme employs DWCE segmentation and object-based region growing. Its overall performance has achieved a 90% TP detection level with 4.2 FPs per image and an 80% TP detection level with 2.0 FPs per image with a diverse database of 253 mammograms. The addition of region growing improved the borders of the detected objects and reduced merging between adjacent or overlapping structures. This improved the morphological information extracted from the detected breast masses and thus the differentiation between masses and normal tissues. The FP reduction was also simplified to a single stage of morphological feature classification and a single stage of SGLD texture feature classification. It is expected that a simplified FP reduction scheme has the potential to generalize better than a more complicated scheme when CAD is implemented in a clinical setting. This breast mass segmentation scheme provided improved FROC performance compared to our previously reported two-stage DWCE technique. Further investigations are under way to improve the region-growing segmentation by analyzing different growing methods that may improve the border definition of the detected structures, as well as to develop new object features that may further differentiate masses from normal structures. Preclinical testing of this algorithm on a large set of independent mammograms will also be conducted.

TABLE IV. Summary of the test FROC result depicted in Fig. 9. The table contains the number of FPs per image for different TP fractions along with the percentage of FPs reduced at each TP level relative to the initial value of 19.2 FPs per image. The first entry in the table is the reduction achieved without missing any additional breast masses.

TP fraction	FPs/image	FP reduction
98%	19.2	0%
95%	6.7	65%
90%	4.2	78%
80%	2.0	90%

ACKNOWLEDGMENTS

APPENDIX A: MORPHOLOGICAL FEATURE DEFINITIONS

A set of 11 features is used in morphological FP reduction. Ten of these features are based solely on the binary object defined by the segmentation. The other feature utilizes the original gray scale values inside and surrounding the segmented object. An individual object segmented from image $F(x,y)$ is defined as:

$$F_{\mathrm{obj}_i}(x,y)=\begin{cases}1, & (x,y) \text{ is a pixel in object } i, \\ 0, & \text{otherwise.}\end{cases} \quad (A1)$$

In addition, $F_{BB_i}(x,y)$ defines the pixels contained in the smallest bounding box completely containing object i and $F_{\mathrm{Eqv}_i}(x,y)$ defines the pixels of the circle with the same area as F_{obj_i} and centered at its centroid location. The radius of $F_{\mathrm{Eqv}_i}(x,y)$ is given by

$$r_{\mathrm{Eqv}}=\sqrt{\frac{\mathrm{area}\ (F_{\mathrm{obj}_i})}{\pi}}. \quad (A2)$$

Five features are based on the normalized radial length (NRL), defined as the Euclidean distance from an object's centroid to each of its edge pixels and normalized relative to the maximum radial length for the object.[18] This results in a NRL vector for each object i given as

$$\mathbf{R}_i=\{r_{i,j}:0\leqslant j\leqslant N_e-1\}, \quad (A3)$$

where N_e is the number of edge pixels in the object and $r_{i,j}\leqslant1$. The histogram of the normalized radial length is also calculated and is given by

$$\mathbf{P}_i=\{\mathrm{prob}_{i,j}:0\leqslant j\leqslant N_h-1\}, \quad (A4)$$

where N_h is the number of bins used in the histogram. Using these basic definitions, the morphological features are defined as follows. Perimeter:

$$\mathrm{Perim}_i=\sum_{\forall x,\forall y} p_i(x,y), \quad (A5)$$

where

$$p_i(x,y)=\begin{cases}1, & F_{\mathrm{obj}_i}(x,y) \text{ is an edge pixel of object } i, \\ 0, & \text{otherwise.}\end{cases}$$

Area:

$$\mathrm{Area}_i=\sum_{\forall x,\forall y} F_{\mathrm{obj}_i}(x,y). \quad (A6)$$

Perimeter-to-area ratio:

$$\mathrm{PAR}_i=\frac{\mathrm{Perim}_i}{\mathrm{Area}_i}. \quad (A7)$$

Circularity:

$$\mathrm{Circ}_i=\frac{\sum_{\forall x,\forall y} F_{\mathrm{obj}_i}\cap F_{\mathrm{Eqv}_i}}{\mathrm{Area}_i}. \quad (A8)$$

Rectangularity:

$$\mathrm{Rect}_i=\frac{\mathrm{Area}_i}{\sum_{\forall x,\forall y} F_{BB_i}}. \quad (A9)$$

NRL mean:

$$\mu_{\mathrm{NRL}_i}=\frac{1}{N_e}\sum_{j=0}^{N_e-1} r_{i,j}. \quad (A10)$$

NRL standard deviation:

$$\sigma_{\mathrm{NRL}_i}=\sqrt{\frac{1}{N_e}\sum_{j=0}^{N_e-1}(r_{i,j}-\mu_{\mathrm{NRL}_i})^2}. \quad (A11)$$

NRL entropy:

$$E_{\mathrm{NRL}_i}=-\sum_{j=0}^{N_h-1} \mathrm{prob}_{i,j}\cdot\log_2(\mathrm{prob}_{i,j}). \quad (A12)$$

NRL area ratio:

$$\mathrm{AreaR}_i=\left\{\frac{1}{N_e\mu_{\mathrm{NRL}_i}}\sum_{j=0}^{N_e-1}(r_{i,j}-\mu_{\mathrm{NRL}_i}):r_{i,j}>\mu_{\mathrm{NRL}_i}\right\}. \quad (A13)$$

NRL zero-crossing count:

$$\mathrm{ZCC}_i=\sum_{j=0}^{N_e-1} z_{i,j}, \quad (A14)$$

where

$$z_{i,j}=\begin{cases}1, & (r_{i,j-1}>\mu_{\mathrm{NRL}_i})\cap(r_{i,j+1}<\mu_{\mathrm{NRL}_i}), \\ 1, & (r_{i,j-1}<\mu_{\mathrm{NRL}_i})\cap(r_{i,j+1}>\mu_{\mathrm{NRL}_i}), \\ 0, & \text{otherwise.}\end{cases}$$

Contrast:

$$\mathrm{Cont}_i=\frac{g_{\mathrm{in}_i}}{g_{\mathrm{out}_i}}, \quad (A15)$$

where g_{in_i} is the average gray value inside object i and g_{out_i} is the average gray value of the one-pixel wide background surrounding the object.

APPENDIX B: SGLD TEXTURE FEATURE DEFINITIONS

Global and local multiresolution texture features are based on the spatial gray level dependence (SGLD) matrix.[22–24] An element of the SGLD matrix, $p_{d,\theta}(i,j)$, is defined as the joint probability that gray levels i and j occur at a given interpixel separation d and direction θ. In this study, n is defined as the number of gray levels in an image. A total of 13 different texture measures were defined for each SGLD matrix. They were defined as follows.[22]

Energy:

$$E=\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p_{d,\theta}^2(i,j). \quad (B1)$$

Correlation:

$$R = \frac{\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(i-\mu_x)(j-\mu_y)p_{d,\theta}(i,j)}{\sigma_x\sigma_y}, \tag{B2}$$

where

$$\mu_x = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} ip_{d,\theta}(i,j), \tag{B3}$$

$$\mu_y = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} jp_{d,\theta}(i,j), \tag{B4}$$

$$\sigma_x = \sqrt{\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(i-\mu_x)^2 p_{d,\theta}(i,j)}, \tag{B5}$$

and

$$\sigma_y = \sqrt{\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(j-\mu_y)^2 p_{d,\theta}(i,j)}. \tag{B6}$$

Entropy:

$$H = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p_{d,\theta}(i,j)\log_2(p_{d,\theta}(i,j)). \tag{B7}$$

Inertia:

$$In = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} (i-j)^2 p_{d,\theta}(i,j). \tag{B8}$$

Inverse difference moment:

$$IDM = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \frac{1}{1+(i-j)^2} p_{d,\theta}(i,j). \tag{B9}$$

Sum average:

$$\mu_{x+y} = \sum_{k=0}^{2n-2} kp_{x+y}(k), \tag{B10}$$

where

$$p_{x+y}(k) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p_{d,\theta}(i,j),$$

$$i+j=k \quad \text{and} \quad k=0,...,2n-2. \tag{B11}$$

Sum variance:

$$\sigma_{x+y}^2 = \sum_{k=0}^{2n-2} (k-\mu_{x+y})^2 p_{x+y}(k). \tag{B12}$$

Sum entropy:

$$H_{x+y} = -\sum_{k=0}^{2n-2} p_{x+y}(k)\log_2(p_{x+y}(k)). \tag{B13}$$

Difference average:

$$\mu_{x-y} = \sum_{l=0}^{n-1} lp_{x-y}(l), \tag{B14}$$

where

$$p_{x-y}(l) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p_{d,\theta}(i,j), \quad |i-j|=l \quad \text{and} \quad l=0,...,n-1. \tag{B15}$$

Difference variance:

$$\sigma_{x-y}^2 = \sum_{l=0}^{n-1} (l-\mu_{x-y})^2 p_{x-y}(l). \tag{B16}$$

Difference entropy:

$$H_{x-y} = -\sum_{l=0}^{n-1} p_{x-y}(l)\log_2(p_{x-y}(l)). \tag{B17}$$

Information measure of correlation 1:

$$IMC_1 = \frac{H-H_1}{\max\{H_x,H_y\}}. \tag{B18}$$

Information measure of correlation 2:

$$IMC_2 = \sqrt{1-\exp^{-2(H_2-H)}}, \tag{B19}$$

where

$$H_x = -\sum_{i=0}^{n-1} p_x(i)\log_2(p_x(i)), \tag{B20}$$

$$H_y = -\sum_{j=0}^{n-1} p_y(j)\log_2(p_y(j)), \tag{B21}$$

$$H_1 = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p_{d,\theta}(i,j)\log_2(p_x(i)p_y(j)) \tag{B22}$$

and

$$H_2 = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p_x(i)p_y(j)\log_2(p_x(i)p_y(j)). \tag{B23}$$

[1] L. Tabar *et al.*, ''Reduction in mortality from breast cancer after mass screening with mammography,'' Lancet **1**, 829–832 (1985).

[2] E. L. Thurfjell, K. A. Lernevall, and A. A. S. Taube, ''Benefit of independent double reading in a population-based mammography screening program,'' Radiology **191**, 241–244 (1994).

[3] C. J. Vyborny and M. L. Giger, ''Computer vision and artificial intelligence in mammography,'' AJR, Am. J. Roentgenol. **162**, 699–708 (1994).

[4] N. Karssemeijer and G. te Brake, ''Detection of stellate distortions in mammograms,'' IEEE Trans. Med. Imaging **15**, 611–619 (1996).

[5] H. Kobatake and Y. Yoshinaga, ''Detection of spicules on mammogram based on skeleton analysis,'' IEEE Trans. Med. Imaging **15**, 235–245 (1996).

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

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

[8] D. Brzakovic, X. M. Luo, and P. Brzakovic, ''An approach to automated detection of tumors in mammograms,'' IEEE Trans. Med. Imaging **9**, 233–241 (1990).

[9] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark, ''Markov random field for tumor detection in digital ammography,'' IEEE Trans. Med. Imaging **14**, 565–576 (1995).

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

[11] H. Kobatake, H. Ron Jin, Y. Yoshinaga, and S. Nawano, ''Computer diagnosis of breast cancer by mammogram processing,'' Radiologia Diagnostica **35**, 29–33 (1994).

[12] N. Petrick, H. P. Chan, B. Sahiner, and D. Wei, ''An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection,'' IEEE Trans. Med. Imaging **15**, 59–67 (1996).

[13] B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, ''Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis,'' Med. Phys. **25**, 516–526 (1997).

[14] N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, ''Computer-aided breast mass detection: False positive reduction using breast tissue composition,'' in *Digital Mammography*, edited by K. Doi, M. Giger, R. Nishikawa, and R. Schmidt (Elsevier, New York, 1996).

[15] J. C. Russ, *The Image Processing Handbook* (CRC, Boca Rato, FL, 1992).

[16] Y. L. Chang and X. Li, ''Adaptive image region-growing,'' IEEE Trans. Image Process. **3**, 868–872 (1994).

[17] L. Xu, A. Krzyzak, and C. Y. Suen, ''Methods of combining multiple classifiers and their applications to handwriting recognition,'' IEEE Trans. Syst. Man Cybern. **22**, 418–435 (1992).

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

[19] P. A. Lachenbruch, *Discriminant Analysis* (Hafner, New York, 1975).

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

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

[22] D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, ''False-positive reduction for detection of masses on digital mammograms: Global and local multiresolution texture analysis,'' Med. Phys. **24**, 903–914 (1997).

[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] R. W. Conners, ''Towards a set of statistical features which measure visually perceivable qualities of textures,'' in *Proceedings of the IEEE Conference on Pattern Recognition and Image Processing*, pp. 382–390 (1979).

[25] D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, and M. M. Goodsitt, ''Multiresolution texture analysis for classification of mass and normal breast tissue on digital mammograms,'' Proc. SPIE **2434**, 606–611 (1995).

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

[27] D. P. Chakraborty, ''Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data,'' Med. Phys. **16**, 561–568 (1989).

[28] D. P. Chakraborty and L. H. L. Winter, ''Free-response methodology, Alternate analysis and a new observer-performance experiment,'' Radiology **174**, 873–881 (1990).