Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification

Nicholas Petrick, Heang-Ping Chan, Datong Wei, Berkman Sahiner, Mark A. Helvie, and Dorit D. Adler
The University of Michigan, Department of Radiology, UH B1D403, 1500 E. Medical Center Drive, Ann Arbor, Michigan 48109-0030

(Received 5 July 1995; accepted for publication 17 May 1996)

This paper presents segmentation and classification results of an automated algorithm for the detection of breast masses on digitized mammograms. Potential mass regions were first identified using density-weighted contrast enhancement (DWCE) segmentation applied to single-view mammograms. Once the potential mass regions had been identified, multiresolution texture features extracted from wavelet coefficients were calculated, and linear discriminant analysis (LDA) was used to classify the regions as breast masses or normal tissue. In this article the overall detection results for two independent sets of 84 mammograms used alternately for training and test were evaluated by free-response receiver operating characteristics (FROC) analysis. The test results indicate that this new algorithm produced approximately 4.4 false positive per image at a true positive detection rate of 90% and 2.3 false positives per image at a true positive rate of 80%.

Key words: mammography, computer-aided diagnosis, density-weighted contrast enhancement filtering, linear discriminant analysis, texture classification

I. INTRODUCTION

Breast cancer is the most common malignancy affecting women and is second only to lung cancer in tumor related deaths in females. It was estimated that 182 000 new cases of breast cancer would occur in American women and 42 000 women would die from the disease in 1994.[1] This comprises 32% of all new cases of cancer and 18% of cancer deaths in women.[1] Efforts to decrease the mortality are currently aimed at early diagnosis and complete removal of small non-metastatic lesions.[2] In an attempt to reduce cost and increase effectiveness, investigators are developing new techniques to improve detection of early breast cancers.[3] Computer-aided diagnosis (CAD) is one technique that may achieve both goals of lowering cost and increasing effectiveness.[4] CAD is especially well suited for the digital imaging technology which is being developed to produce digital images in full view mammography.

Several research groups have developed computer algorithms for automated detection of mammographic masses. Kegelmeyer has reported promising results for detecting spiculated lesions based on local edge characteristics and Laws texture features.[5,6] Both Lai *et al.*[7] and Qian *et al.*[8] proposed different variations of median filtering to enhance the digitized image prior to object identification. A thresholding method for mass localization and a mass classification algorithm using fuzzy pyramid linking have been developed by Brzakovic *et al.*[9] Other investigators have proposed using the asymmetry between the right and left breast images to determine possible mass locations. Yin *et al.* uses both linear and nonlinear bilateral subtraction[10] while the method by Lau *et al.* relies on "structural asymmetry" between the two breast images.[11] The above methods produced between one and five false detections for a true positive detection rate of approximately 90%. However, it is difficult to compare the effectiveness of these methods because each used a unique set of digitized mammograms, and the results varied between training and test. A general comparison between algorithms is further complicated by the fact that most of these studies were conducted using small data sets. While initial results from the first large scale preclinical study have been encouraging,[12] the performance of detection programs with clinical samples may not match their performance in laboratory tests.

Our preliminary study introduced the density-weighted contrast enhancement (DWCE) segmentation method and found that it was capable of detecting breast masses on 25 digitized mammograms.[13] In this article, a set of 168 digitized mammograms is used to evaluate a modified version of the original DWCE segmentation method in combination with a texture classification scheme. The following procedure was used to evaluate this new detection scheme. The set of digitized mammograms was first segmented into potential breast masses using the DWCE segmentation.[13] This method employed an adaptive filter to enhance structures within the breast region of a mammogram and then identified the structures using a simple edge detection algorithm. Once the digitized images were segmented using the DWCE, regions of interest (ROIs) based on the detected breast structures were extracted from each mammogram, and a set of multiresolution texture features were calculated for each extracted ROI. The feature set was then used by a linear discriminant analysis (LDA) algorithm to reduce the number of false detections. Finally, the performance of the DWCE segmentation and ROI texture classification scheme was evaluated using

 0094-2405/96/23(10)/1685/12/$10.00

free-response receiver operating characteristics (FROC) analysis.

II. MATERIALS AND METHODS

A. Database

The clinical mammograms used in this study were acquired with American College of Radiology 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 mammograms were selected from the files of patients who had undergone biopsy at the University of Michigan in the last five years. The selection criterion used by the radiologists was simply that a biopsy-proven mass existed on the mammogram. This set excluded lesions visible only by architectural distortions (i.e., no defined mass) but included masses accompanied by calcifications. No attempt was made to match the number of malignant and benign mass cases, but we did try to include a cross section of malignant masses. This led to a much larger proportion of malignant lesions than that in the general screening population. To avoid the effect of the repetitive grid pattern on the texture feature calculations, all mammograms with visible grid lines were excluded for the original set. Our final data set for this preliminary study was composed of 168 single-view mammograms. It included 85 malignant and 83 benign masses. The size of the masses ranged from 5 mm to 26 mm with a mean size of 12.2 mm, and their visibility ranged from 1 (obvious) to 10 (subtle) with a mean visibility of 4.51. A more complete discussion of the images selected for this study can be found in Wei *et al.*[14]

The mammograms were digitized with a LUMISYS DIS-1000 laser film scanner with a pixel size of 100 μm and 4096 gray levels. The digitizer logarithmically amplifies the light transmitted through the mammographic film before analog-to-digital conversion so that the gray levels are linearly proportional to optical densities in the range of 0.1 to 2.8 optical density units (O.D.). The O.D. range of the scanner is 0–3.5 with large pixel values in the digitized mammograms corresponding to low O.D. The digitized images used in this study were approximately 2000×2000 pixels in size. To conserve processing time and reduce noise in the initial DWCE segmentation stages, the full resolution mammograms were first smoothed with an 8×8 box filter and subsampled by a factor of 8, resulting in 800-μm images of approximately 256×256 pixels in size. However, the texture features used in the final LDA classification were calculated from the original images with a 100-μm pixel size.

The location and extent of all the biopsy-proven masses were marked on the original films by a radiologist. They were then localized on the digitized images and stored in a "truth" file on the computer by defining both the centroid (approximate center) of the lesion and the smallest bounding box (rectangle) containing the entire lesion. Both of these procedures were performed by hand using the original marked film as a guide. The centroid "truth" was used to analyze the initial DWCE segmentation. If an object segmented by the DWCE contained the centroid of the mass within the object region, it was considered a true positive (TP); otherwise, it was considered a false positive (FP). The centroid provided a fast method for evaluating the DWCE segmentation in its global and local stages. However, the final texture classification results are based on the more precise bounding box "truth." A region was considered a TP detection when at least 50% of the "truth" bounding box was detected. The centroid and bounding box definitions for the mass provided both an efficient mechanism for development of the DWCE and an accurate final analysis for the overall detection scheme.

For evaluation of the DWCE segmentation and subsequent texture classification, the 168 single-view mammograms were randomly divided into two groups of 84 images, groups G1 and G2, with the constraint that all images from a single patient were kept in the same group. A single set of DWCE segmentation parameters was applied to all images (G1 and G2) to extract potential mass regions. The regions extracted from the G1 and G2 images were then alternately used as training and test sets in the texture classification as described below.

B. Density-weighted contrast enhancement segmentation

Edge detection applied to an unenhanced image was not effective in detecting breast masses because of the low signal-to-noise ratio of the edges and the presence of complicated structured background. To overcome these problems, we have developed a new algorithm using DWCE filtering along with Laplacian–Gaussian (LG) edge detection for automatic segmentation of low contrast structures in digital mammograms.[13] The DWCE segmentation method employed adaptive filtering, edge detection, and morphological FP reduction to detect potential breast masses in a two-stage approach. In the first stage, DWCE segmentation was applied globally to the entire breast region of the mammogram to identify ROIs. In the second stage, the segmentation was applied locally to the ROIs identified in the global stage. Figures 1(a) and 1(b) depict the block diagrams for the global and local stages of this algorithm. The DWCE segmentation was originally introduced by Petrick *et al.*[13] but has been slightly modified in this study to improve its overall performance. In the following subsections we will summarize the main components of both the global and local stages, and highlight the differences between the original and current implementations of the DWCE technique.

1. Global stage: Density-weighted contrast enhancement filtering

The DWCE filter was developed to accentuate mammographic structures before edge detection by adaptively enhancing local contrast and is an extension of the local contrast and mean adaptive filter proposed by Peli and Lim.[15] The block diagram of the filter is shown in Fig. 2, while Fig. 3 contains examples of the images produced by each filter

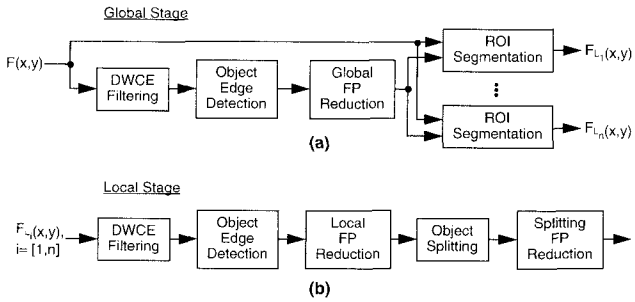

FIG. 1. The block diagram of the two-stage DWCE segmentation method used for the initial breast mass detection. The block diagram for the global stage is depicted in (a) while the local stage is shown in (b). Note, the outputs from the global stage, $F_{L_i}(x,y)$, are individually processed in the local stage.

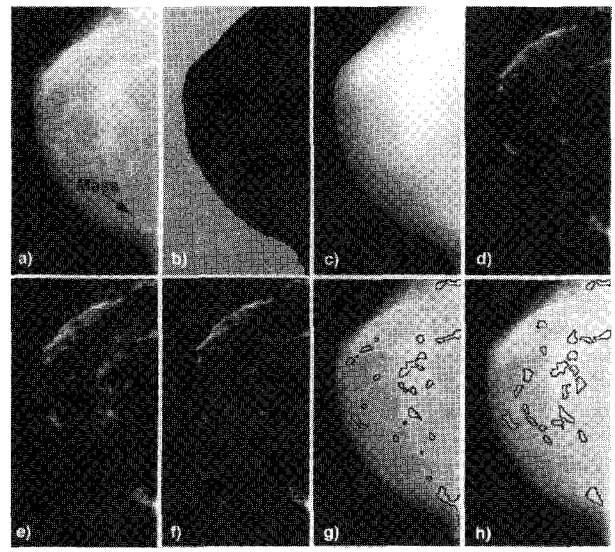

FIG. 3. (a) A typical mammogram from our image database; (b) the corresponding breast map used in the DWCE segmentation; (c) the density image ($F_D(x,y)$); (d) the contrast image ($F_C(x,y)$); (e) the weighted-contrast image ($F_{KC}(x,y)$); (f) the rescaled weighted-contrast image ($F_E(x,y)$); (g) the detected structures remaining after the global FP reduction step; (h) the detected structures remaining after the final splitting FP reduction step.

block for a typical mammogram from our image set. All the DWCE functions introduced in the following discussion correspond to the steps illustrated in Fig. 2.

DWCE filtering was applied to the breast region [i.e., the breast map $F_{\text{Map}}(x,y)$] of each mammogram which had been identified using thresholding and edge detection.[13] Figure 3(a) shows a typical mammogram, $F(x,y)$, at 800-μm resolution while 3(b) shows its breast map. The pixel intensities from $F(x,y)$ within the breast map were next rescaled to be between 0.0 and 1.0 producing a normalized breast image, $F_N(x,y)$. This normalization reduced the gray-level variation due to breast tissue composition and the imaging technique so that a single set of filter parameters could be applied uniformly to all digitized mammograms.

The normalized image was next split into a density and a contrast image, $F_D(x,y)$ and $F_C(x,y)$, respectively. $F_D(x,y)$ was produced by low-pass filtering the normalized input image using $G\{0,\sigma_D\}$, a Gaussian filter with zero mean and standard deviation $\sigma_D=8.0$. Likewise, $F_C(x,y)$ was produced by bandpass or high-pass filtering the normalized image. In the current DWCE implementation, $F_C(x,y)$ is created by subtracting the density image from the normalized input,

$$F_C(x,y)=F_N(x,y)-F_D(x,y), \tag{1}$$

or

$$F_C(x,y)=F_N(x,y)-G\{0,\sigma_D\}*F_N(x,y), \tag{2}$$

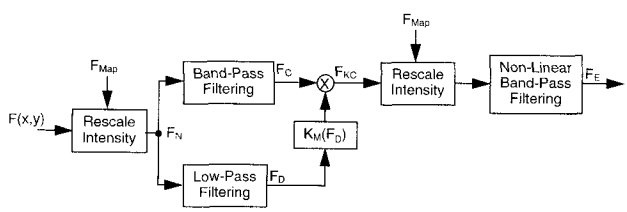

FIG. 2. The block diagram for the DWCE preprocessing filter used for image enhancement.

where * represents two-dimensional convolution. Figures 3(c) and 3(d) show the density and contrast images obtained using this procedure.

The local density value, $F_D(x,y)$, was then used to determine a multiplication factor, $K_M(F_D(x,y))$, for each pixel (x,y) in the image. The multiplication factor was used to either enhance or suppress the local contrast and thereby produced a new weighted contrast image:

$$F_{KC}(x,y)=K_M(F_D(x,y))\times F_C(x,y). \tag{3}$$

This process allowed the DWCE filter to adapt to local background characteristics within the image and was the principle component for our adaptive signal-to-noise ratio (SNR) enhancement. In this case, the signal refers to breast masses or other predominant structures within the breast. The output of the DWCE filter was given as

$$F_E(x,y)=K_{NL}(F_{KC}(x,y))\times F_{KC}(x,y), \tag{4}$$

where each pixel, (x,y), in the weighted contrast image was used to define a second multiplication factor, $K_{NL}(F_{KC}(x,y))$, that nonlinearly scaled the weighted contrast image. This nonlinear scaling was used to further suppress the background and to separate merged structures in the DWCE enhanced image. Figures 3(e) and 3(f) show the weighted contrast and scaled weighted contrast images, respectively, obtained with the DWCE technique.

It can be seen that the two multiplication functions, K_M and K_{NL}, define the enhancement properties of the filter. These functions can be tailored to suit a specific task. Figures 4(a) and 4(b) show the curves selected for K_M and K_{NL}, respectively, in the current filter. The shape of the density-weighted contrast function, K_M, was selected to accentuate $(K_M(z)>1.0,\ z=F_D(x,y))$, the contrast at pixels in the den-

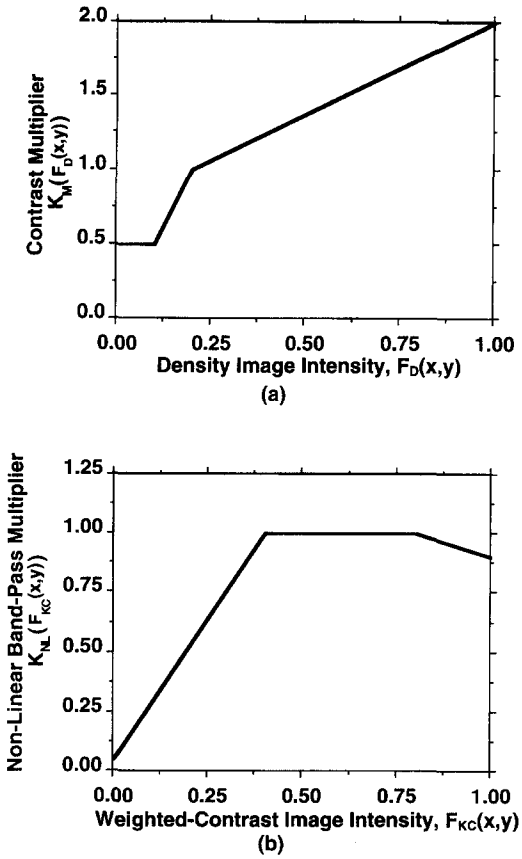

FIG. 4. Plots of (a) the weighted-contrast multiplication function ($K_M(z)$) and (b) the nonlinear rescaling function ($K_{NL}(z)$) used in the DWCE filtering.

sity image with medium to high intensity, while deemphasizing ($K_M(z)<1.0$), the contrast at pixels with low intensity. Thus, this function suppressed small structures mainly surrounded by background tissue and enhanced larger structures which are more likely to be masses. The exact shape of the multiplication function was determined experimentally by observing how detection was affected by variations in K_M. We chose $\{K_M(z)\geq 1.0{:}0.25\leq z\leq 1.0\}$ in the current weighted contrast function so that 75% of the intensity range was enhanced. K_M was found to be effective in reducing the background and enhancing breast structures, but it did not provide adequate separation between the structures. The shape of the nonlinear scaling function ($K_{NL}(z)$, $z=F_{KC}(x,y)$) was selected to provide additional separation between objects. Very low contrast regions were strongly deemphasized, thus eliminating many low-intensity bridges between individual structures. It was also found that a slight suppression of the highest contrast intensities provided a more uniform intensity distribution across detected breast structures. Again, the specific shape of the nonlinear contrast scaling was determined experimentally by observing the effect of different functional forms on the detection and object separation. A complete discussion of the DWCE multiplication functions used in this study can be found in the literature.[13]

2. Global stage: Object edge detection

The DWCE filtering was applied to the original mammogram to facilitate the detection of structures within the image and thus provided an estimate of their physical extent.[16] The DWCE implementation provided significant background reduction, as shown in Fig. 3(f), allowing for the use of a less complex edge detector. In this study, object edges were identified from the DWCE filtered mammogram using a Laplacian–Gaussian (LG) edge detector [Block 2 in Fig. 1(a)]. Edges in the enhanced image, $F_E(x,y)$, were defined as the zero crossing locations of

$$\nabla^2 G\{0,\sigma_E\} * F_E(x,y), \tag{5}$$

where $G\{0,\sigma_E\}$ was a zero mean Gaussian smoothing function with standard deviation $\sigma_E=2.0$.[17] The advantages of this edge detector are that its performance is independent of edge direction and that it tends to produce closed regions.

After the edge detection, all enclosed structures were filled to eliminate any holes that may have formed inside individual objects. The edges from each of the filled objects were tracked and identified. This edge detection is identical to the original DWCE implementation described in the literature.[13]

3. Global stage: False positive reduction

The DWCE filtering and subsequent edge detection do not differentiate between mass and normal tissues, therefore, a large number of potential regions were usually found. Since the shape of breast masses in general are different from those of normal tissue, we extracted morphological features and used a classification algorithm to identify some of these differences [Block 3 in Fig. 1(a)]. The goal here was to reduce the number of FP regions without losing a significant number of true masses, thus allowing the maximum number of TP regions to be passed on to the local processing stage. In this study, six additional morphological features were combined with the original set of five features used in the previous study[13] to improve the differentiation between mass and normal tissue objects. The original features were the number of edge pixels (P), the total object area ($A=\text{area}(F_{obj})$), the object's contrast, circularity, and rectangularity. The new features added in this implementation were the perimeter-to-area ratio (PAR) and a set of five normalized radial length (NRL) features. To define circularity and rectangularity, the minimum sized bounding box completely containing the object, $F_{bb}(x,y)$, and a circle with an area equivalent to the object area, $F_{eq}(x,y)$, were calculated. $F_{eq}(x,y)$ was centered at the object's centroid location and had radius r_{eq} given by

$$r_{eq}=\sqrt{\frac{\text{area}(F_{obj})}{\pi}}=\sqrt{\frac{A}{\pi}}. \tag{6}$$

Circularity and rectangularity were then defined as

$$\text{Circularity}\equiv\frac{\text{area}(F_{obj}\cap F_{eq})}{\text{area}(F_{obj})}, \tag{7}$$

$$\text{Rectangularity} \equiv \frac{\text{area}(F_{\text{obj}})}{\text{area}(F_{\text{bb}})}. \tag{8}$$

The five NRL features were a subset of the features defined by Kilday *et al.*[18] A radial length function was 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. This created an NRL vector given as

$$\mathbf{r}=\{r_k : 0 \leq k \leq N_E - 1\}, \tag{9}$$

where N_E was the number of edge pixels in the object. The histogram of the radial length was also calculated and created the probability vector

$$\mathbf{p}=\{p_j : 0 \leq j \leq N_H - 1\}, \tag{10}$$

where N_H was the number of bins used in the histogram. The NRL features selected in this study were the NRL mean value, standard deviation, entropy, area ratio, and zero crossing count. They are defined as

$$\mu_{\text{NRL}} \equiv \frac{1}{N_E} \sum_{k=0}^{N_E-1} r_k, \tag{11}$$

$$\sigma_{\text{NRL}} \equiv \sqrt{\frac{1}{N_E} \sum_{k=0}^{N_E-1} (r_k - \mu_{\text{NRL}})^2}, \tag{12}$$

$$E_{\text{NRL}} \equiv - \sum_{j=0}^{N_H-1} p_j \log(p_j), \tag{13}$$

$$AR_{\text{NRL}} \equiv \left\{ \frac{1}{N_E \, \mu_{\text{NRL}}} \sum_{k=0}^{N_E-1} (r_k - \mu_{\text{NRL}}) : r_k > \mu_{\text{NRL}} \right\}, \tag{14}$$

$$ZCC_{\text{NRL}} \equiv \text{number of zero crossings of } \{r_k - \mu_{\text{NRL}}\}_{k=0}^{N_E-1}. \tag{15}$$

A complete description of all the NRL features used in this study can be found in the literature.[18]

The extracted morphological features were used in a sequential classification scheme; a simple threshold classifier, followed by an LDA classifier, and finally followed by a backpropagation neural network (BPN). The purpose of each classifier was to reduce the number of FP regions with a minimum number of TP losses. This improved reduction scheme was selected because it has been found that sequential or parallel combinations of the different classifiers often increased the classification accuracy over the individual classifiers.[19,20] This is probably because they extract different information from the feature space. The threshold classifier simply set a maximum and a minimum value for each morphological feature. This provided some initial reduction and prevented the LDA and BPN classifiers from training with nonrepresentative object features. If all the morphological features from a detected object fell within the bounds, it was kept as a potential mass; otherwise, it was considered to be normal tissue and discarded. All DWCE detected objects with features values within the defined limits were saved as potential mass objects and passed on to the LDA classifier. The maximum and minimum feature limits, $f_{\text{Th}\uparrow}$ and $f_{\text{Th}\downarrow}$, respectively, were identical for both the G1 and G2 image groups and were selected as a multiple of the individual mass object bounds:

$$f_{\text{Th}\uparrow_i} = \{k \max_j(f_{i,j}) : j \in [\text{index of all detected mass objects}]\}, \tag{16}$$

$$f_{\text{Th}\downarrow_i} = \{k \min_j(f_{i,j}) : j \in [\text{index of all detected mass objects}]\}, \tag{17}$$

where $f_{i,j}$ is the value of the ith feature ($i \in [1,11]$) for the jth detected object. For this study, the multiplication factor (k) was selected to be 1.0. The second classifier, LDA, formed a linear combination of the morphological features and produced a single discriminant score for all remaining potential mass object. This classification scheme will be described in more detail in Sec. II C. The LDA classifier applied to the G1 objects was trained with the G2 object features and vice versa. This provided independent LDA training for each of the image sets. In order to minimize the probability of losing true masses, a lax discriminant threshold was chosen to retain most of the masses while achieving moderate FP reduction. The reduced sets of G1 and G2 objects with their morphological features were then passed on to a final BPN classification step. BPN formed a nonlinear combination of the morphological features into a single discriminant score. A complete description of the BPN morphological classification can be found in the literature.[13,21,22] In this step, a three input node, four hidden node, single output BPN architecture was utilized. The BPN classifier was trained in a similar fashion as the LDA but only the three most uncorrelated features (area, perimeter-to-area ratio, and contrast) were used as the input features. The individual G1 and G2 image sets were again used to train a pair of BPN classifiers, and the discriminant thresholds were chosen to maximize FP reduction while minimizing the loss of masses. All remaining DWCE detected objects after the application of the three classifiers were considered as potential mass objects and passed on to the ROI segmentation and subsequent local stage of the DWCE segmentation. Figure 3(g) shows the final reduced set of objects detected by the global stage for the original mammogram of Fig. 3(a).

4. Global stage: ROI segmentation

The final step in the global stage was the segmentation of the detected local regions [Blocks $4_{1 \ldots n}$ in Fig. 1(a)]. For each remaining potential mass object, a ROI corresponding to the object's bounding box was defined on the subsampled mammogram. The minimum size for these ROIs was chosen to be 32×32 pixels. A bounding box of an object smaller than this size was uniformly expanded in each direction (horizontal and vertical) until it reached 32×32 pixels. These defined object regions were then used as input ROIs to the local DWCE stage.

5. Local stage: DWCE filtering, edge detection, and local false positive reduction

The local stage of the DWCE segmentation was very similar to the global stage and is again depicted in Fig. 1(b). The main difference was that the processing was performed in local regions within the image. This local processing allowed the DWCE filter to adapt to the intensity distribution within each ROI and thus refined the borders of the detected objects. The input images to this stage, $F_{L_i}(x,y)$, were defined from the detected objects in the global stage. This local stage had five main components. Three of the components had corresponding global stage counterparts, and they included a second DWCE filter, LG edge detector, and local FP reduction step. The local DWCE filter and LG edge detector used identical parameters as their first stage counterparts, while the FP reduction step again used the 11 morphological features and the sequential thresholding, LDA, and BPN classification discussed previously. The only difference in the local FP reduction was that the feature and discriminant thresholds were adjusted to reflect the morphological properties of the locally extracted structures. Again, the goal of this FP reduction step was to reduce the number of potential mass regions before the regions were processed by a final texture classification stage. Therefore, lax decision thresholds were chosen to minimize additional losses of true mass objects.

6. Local stage: Object splitting and splitting FP reduction

The local processing of the mammograms lead to larger objects because of the improved estimate of the local background. However, the larger objects often resulted in region merging, (i.e., different structures within the breast merged into a single detected region). An object splitting step was therefore added to the local stage [Block 4 in Fig. 1(b)]. This splitting step enabled the use of fixed sized ROIs in the final texture classification. The splitting algorithm searched for narrowings in the cross section of an object. The algorithm initially found the cross-section width for each column in the object [$F_X(x)$ with length n]. Using $F_X(x)$, three parameters were calculated for each x. They were the area ratio of the two created objects along with the global and local cross-section width ratios. These ratios were defined as

$$F_{\text{Area}}(x) \equiv \frac{\min(A_R(x), A_L(x))}{\max(A_R(x), A_L(x))}, \tag{18}$$

$$F_{\text{Gbl}}(x) \equiv \left\{ 1.0 - \frac{F_X(x)}{\max(F_X(z))} : \; z \in [0, n-1] \right\}, \tag{19}$$

$$F_{\text{Lcl}}(x) \equiv \left\{ 1.0 - \frac{F_X(x)}{\max(F_X(z))} : \; z \in [x-2, x+2] \right\}, \tag{20}$$

where $A_R(x)$ and $A_L(x)$ were the area of the right and left objects produced by splitting at location x. At each potential neck location, x, a cut value $F_{\text{Cut}}(x)$ was defined as a linear combination of the cross-section ratios and the area ratio

$$F_{\text{Cut}}(x) = 1.5F_{\text{Gbl}}(x) + 2.0F_{\text{Lcl}}(x) + 1.0F_{\text{Area}}(x). \tag{21}$$

After similar cut functions were computed for each row and for the 45° and 135° directions, a maximum cut value was found for the object and compared to a cut threshold. If this maximum cut value exceeded this threshold, the object was split at that point; otherwise, it was left unchanged. If the object was split, the same algorithm was applied to the newly formed objects until no further splitting occurred. The splitting algorithm incorporated area information into the splitting process, thereby giving preference to narrowings closer to the center of the object and minimizing the number of times an object was split. For a complete description of this splitting algorithm refer to Petrick *et al.*[13]

The final FP reduction [Block 5 in Fig. 1(b)] again employed the 11 morphological features and the sequential classification scheme described in Sec. II B 3. The feature and discriminant thresholds were adjusted to reflect the morphological properties of the split objects. Figure 3(h) shows the set of detected objects after the complete two-stage DWCE segmentation for the original mammogram of Fig. 3(a).

C. Texture classification

After the DWCE segmentation identified a set of potential mass objects in the mammograms, ROIs corresponding to the detected object locations were extracted from the original 100-μm images and used as input to a texture classifier. The extracted ROIs had a fixed size of 256×256 pixels and the center of each ROI corresponded to the centroid location of a detected object. When the object was located close to the border of the mammogram and a complete 256×256 pixel ROI could not be defined, the ROI was shifted over until the appropriate edge coincided with the border of the original image. The classification of these fixed sized ROIs was based on a multiresolution texture analysis scheme. The approach has been described in detail in the literature[23] with the essential steps in the classification summarized below.

1. Texture features

The texture features used in the classification were derived from the spatial gray-level dependence (SGLD) matrix.[24,25] An element of the SGLD matrix, $p_{d,\theta}(i,j)$, is the joint probability that the gray levels i and j occur at a given interpixel separation d and direction θ. A set of SGLD matrices can be defined by varying the separation and direction. Thirteen texture features were derived from each SGLD matrix including correlation, energy, entropy, inertia, inverse difference moment, sum average, sum variance, sum entropy, difference average, difference variance, difference entropy, and two measures of correlation information. The mathematical definitions for the SGLD features can be found in the literature.[23–26] These features were selected because they were found to be effective in the classification of ROIs containing masses or normal tissue manually identified by radiologists.[14,23,27,28] Each texture feature was calculated in the θ=0°,45°,90°,135° directions. The features obtained at θ=0°,90° and θ=45°,135° were averaged since no angular bias was seen in the texture of masses, and we did not find any significant difference in classification accuracy between features at separate angles and their averaged values.[28] The

features calculated at adjacent pixels on axis (θ=0°,90°) and those in the diagonal direction (θ=45°,135°) were not averaged because of the significant $\sqrt{2}$ difference in the actual distances.[14]

Before the texture features were calculated, background correction was performed on the individual ROIs using a method described previously.[19,28] An ROI was first low-pass filtered and a pixel in the low-frequency background image was estimated as a weighted sum of the pixel values surrounding the ROI. The difference between the original ROI and the background thus reduced the gray-level variation due to the low-frequency structured background within the ROI.

2. Global multiresolution SGLD features

A wavelet transform with a four coefficient Daubechies kernel was used to decompose the individual ROIs into multiple scales after background correction.[14,29] Multiresolution ROI images were obtained using the original ROI (Scale 1) and the first two low-pass down-sampled approximation wavelets (Scales 2 and 4, respectively). The wavelet coefficients at Scale 8 were obtained by wavelet filtering but without down-sampling so that the minimum image size was maintained at 64×64 pixels. This minimum size was selected in order to reduce the statistical uncertainty when SGLD matrices of large pixel distances were calculated from the Scale 8 wavelet images.

Fourteen SGLD matrices, with effective distances of d={1,2,4,8,12,16,20,24,28,32,36,40,44,48} pixels relative to the original ROI, were calculated in both the on-axis (θ =0°,90°) and diagonal (θ=45°,135°) directions for each ROI using the Scale 1, 2, 4, and 8 wavelet images. Figure 5 contains a graphical representation of how the different wavelet images were related to the different SGLD matrices and the different object features. The SGLD matrices with d={1,2,4} were calculated using a pixel distance of one in the Scale 1, 2, and 4 wavelet images, respectively. The eleven SGLD matrices at d={8,12,16,20,24,28,32,36,40,44,48} were calculated from the Scale 8 wavelet image with pixel distances from 2 to 12 pixels. This process produced a total of 28 different SGLD matrices and 364 global multidistance texture features for each ROI.

3. Local multidistance SGLD features

A set of local texture features was also calculated for each ROI.[23,27] Five rectangular subregions were segmented from each ROI; an object subregion defined by the original DWCE object bounding box located at the center of the ROI, and four peripheral subregions at the corners. For a given pixel distance d and a given direction θ, an SGLD matrix was formed from the object subregion and another SGLD matrix was formed from the pixel pairs in the four peripheral subregions. These local SGLD matrices were calculated for d={1,2,4,8} and θ={0°,90°} and {45°,135°}. The thirteen texture features were calculated for both the object and periphery SGLD matrices. A total of 208 local features were defined for each ROI. They included the 104 features in the object region and 104 additional features defined as the difference between the feature values in the object and the periphery regions.

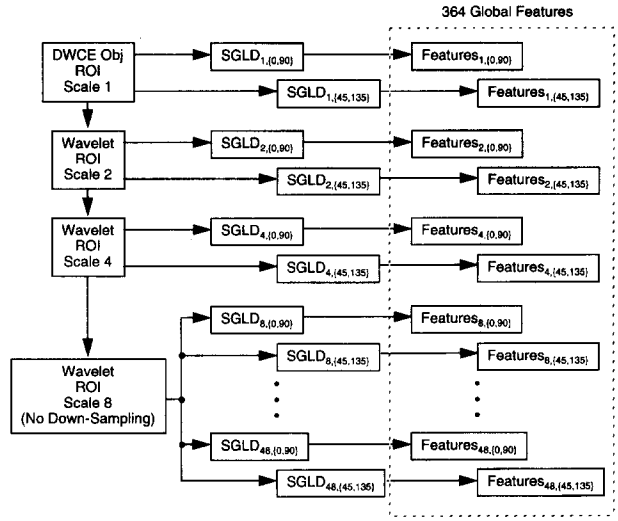

FIG. 5. Graphical representation of the parameters used in extracting features from the multiresolution wavelet images. The effective pixel distance d={1,2,4,8,12,16,20,24,28,32,36,40,44,48} for the SGLD matrices are relative to the original image.

4. Linear discriminant analysis

Linear discriminant analysis (LDA) uses a set of feature variables to classify an individual into one of a set of mutually exclusive classes.[30] We found in our previous studies that the LDA using SGLD texture features can effectively separate masses from normal tissue using ROIs manually selected by radiologists.[14,23] In our two class (mass and normal tissue) problem, the set of 572 global and local texture features was used as a pool of predictor variables in a stepwise selection procedure. This procedure selected a subset of features from the feature space based on the maximization of the Mahalanobis distance.[31] The stepwise selection eliminates irrelevant variables and thus improves the generalization capability of a discriminant function optimized with a finite number of training cases.

With the DWCE segmentation and object splitting algorithm, many of the extracted ROIs overlapped with one another because of the adjacency of the objects. We selected the independent ROIs (i.e., the ROIs that did not overlap with one another) to form a training set in order to avoid biases in the statistical distributions of the feature vectors. Two independent sets, $G1_i$ and $G2_i$, were formed by reducing all pairs of overlapping ROIs to single regions in G1 and G2, respectively. If a true mass ROI overlapped with a normal tissue ROI, the true mass region was saved while the normal region was eliminated. If two normal regions overlapped, one randomly selected region was eliminated. Finally, if two regions containing the full breast mass overlapped, the region defined by the DWCE segmented object which contained the centroid of the true mass was saved while the other was eliminated. These independent $G1_i$ and

TABLE I. The number of detected objects, the single stage reduction, the mean object area (μ_{Area}), and the standard deviation of the object areas (σ_{Area}) for the G1 data set after the global, local, and splitting stage FP reduction steps. The single stage reduction is defined as the reduction achieved by the morphological FP reduction block in each stage.

Stage	TP detections	FP detections per image	Single stage reduction	μ_{Area} (pixels)	σ_{Area} (pixels)
Global	82 of 84	34.6	25%	63.3	109.0
Local	81 of 84	12.4	75%	286.4	351.6
Split	81 of 84	18.9	14%	122.0	122.1

$G2_i$ sets were individually used to train the LDA classifiers while the full G1 and G2 sets were used for classifier evaluation.

To improve the statistical properties of the feature distributions, we used the entire set of segmented ROIs from both the $G1_i$ and $G2_i$ image sets for selection of feature variables. After feature selection, the G1 and G2 groups were used alternately as training and test sets. For example, when the coefficients of the linear discriminant function were optimized by the feature values from the $G1_i$ set, the classification accuracy of the linear discriminant function was tested with the full G2 set. The $G1_i$-trained linear discriminant function was also applied to the full G1 group to evaluate its self-consistency. Therefore, a total of four groups of discriminant scores were obtained: {Train:$G1_i$, Test:G1}, {Train:$G2_i$, Test:G2}, {Train:$G2_i$, Test:G1}, and {Train:$G1_i$, Test:G2}.

In this study, FROC analysis[32] was used to evaluate the performance of the complete segmentation method. The tradeoff between the TP fraction and the number of FP detections per image was determined by varying the decision threshold on the ROI discriminant scores. The raw detection data for both the full group training and test cases are presented, along with the fitted FROC curves obtained using the FROCFIT program.[32]

III. RESULTS

The number of TP and FP objects detected in the global and local stages of the DWCE segmentation are summarized in Tables I and II for the G1 and G2 image sets, respectively. A TP detection for the DWCE segmentation is again simply defined as an object locating the centroid of a breast mass, and a FP is any object other than the true mass (as discussed in Sec. II A). The two-stage DWCE segmentation missed only 8 of the 168 breast masses contained in the entire image set. Using the sets of TP and FP objects, 256×256 pixel ROIs representing each of the detected objects were extracted from the full resolution mammograms. A total of 1690 ROIs were extracted from the set of 84 G1 images and 1874 from the G2 mammograms. The independent $G1_i$ and $G2_i$ sets used for LDA training included 476 and 503 nonoverlapping ROIs, respectively. Stepwise feature selection was then performed on the 572 multidistance texture features using the combined $G1_i$ and $G2_i$ image sets, as described above, and 29 features were selected. These 29 features were used in the LDA texture classification for training and testing both the G1 and G2 image sets. Figures 6 and 7 show the raw and fitted training FROC curves obtained using the LDA texture classifier for the {Train:$G1_i$, Test:G1} and {Train:$G2_i$, Test:G2} combinations. The raw and fitted FROC curves for the test sets, {Train:$G2_i$, Test:G1} and {Train:$G1_i$, Test:G2}, are likewise depicted in Figs. 8 and 9. Finally, Table III contains the raw FROC results at TP detection rates of 90% and 80%, and Table IV contains the FROCFIT program parameters estimated for each of the fitted FROC curves.

TABLE II. The number of detected objects, the single stage reduction, the mean object area (μ_{Area}), and the standard deviation of the object areas (σ_{Area}) for the G2 data set after the global, local, and splitting stage FP reduction steps. The single stage reduction is defined as the reduction achieved by the morphological FP reduction block in each stage.

Stage	TP detections	FP detections per image	Single stage reduction	μ_{Area} (pixels)	σ_{Area} (pixels)
Global	79 of 84	32.9	32%	64.4	112.4
Local	79 of 84	21.4	62%	219.8	289.6
Split	79 of 84	21.6	7%	108.2	91.2

TABLE III. The number of FPs per image of each FROC curve at 90% and 80% TP detection fractions.

Training set	Test set	FPs per image (90% TP fraction)	FPs per image (80% TP fraction)
$G1_i$	G1	3.77	1.88
$G2_i$	G2	4.55	1.47
$G2_i$	G1	3.98	2.50
$G1_i$	G2	4.72	2.08

IV. DISCUSSION

A. DWCE segmentation

The purpose of the global processing stage was to define a set of local regions which contained the true breast masses and as few normal regions as possible. The initial DWCE filtering and subsequent edge detection was able to detect 161 of the 168 true masses in this preliminary study, including 83 of the 85 malignant masses. In addition, five of the seven missed masses, including both malignant masses, were

TABLE IV. Summary of the FROCFIT parameters and goodness of fit values. The headings for the table are; the two estimated FROCFIT parameters (a and b), the standard deviation of the estimated parameters (σ_a and σ_b), the area under the alternative FROC curve (A_{AFROC}), the standard deviation of the area (σ_A), the normalized chi-squared value (χ^2), and the significance probability for the fit (Prob).

Training set	Test set	a	σ_a	b	σ_b	A_{AFROC}	σ_A	χ^2	Prob
$G1_i$	G1	0.19	0.11	0.50	0.05	0.57	0.04	1.39	0.04
$G2_i$	G2	0.28	0.11	0.47	0.05	0.61	0.04	0.92	0.61
$G2_i$	G1	0.11	0.11	0.58	0.05	0.54	0.04	0.96	0.55
$G1_i$	G2	0.11	0.11	0.45	0.04	0.54	0.04	1.03	0.42

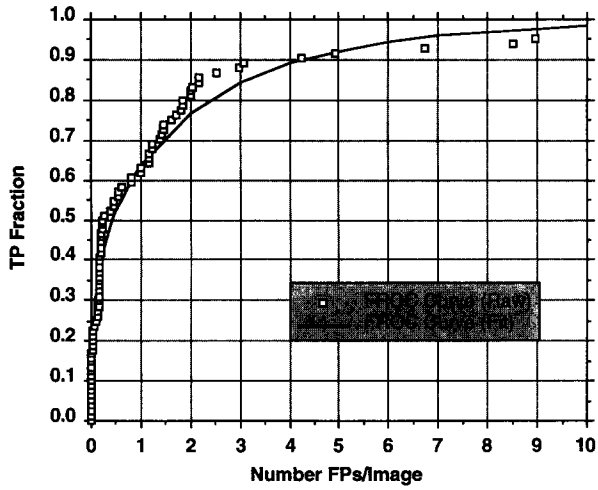

FIG. 6. FROC curves obtained with the image group {Train $G1_i$, Test G1}. The data points are raw data obtained by varying the decision threshold on the discriminant scores. The solid curve is obtained from the FROCFIT program.

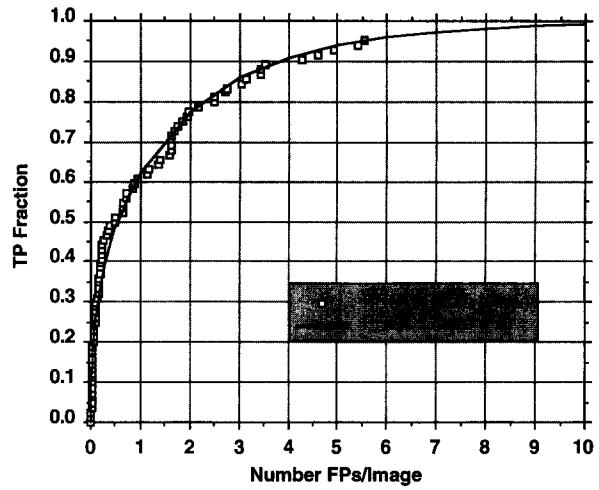

FIG. 8. FROC curves obtained with the image group {Train $G2_i$, Test G1}. The data points are raw data obtained by varying the decision threshold on the discriminant scores. The solid curve is obtained from the FROCFIT program.

detected in another mammogram containing a different view of the same breast. The image set did not include any additional views for the two remaining misses. This indicates that the global stage is effective in the initial detection task. However, the morphological properties of the detected regions proved to be of limited value in differentiating between TP and FP objects in the low-resolution DWCE filtered images. The main problem was that the global detection underestimated the size of the actual structures. This can be clearly seen in Fig. 3(g) where the detected objects are usually much smaller than the actual structures in the image. The average size, after FP reduction, of the global stage objects was 64.4 pixels. This underestimation can be mainly attributed to the large intensity range over which the background suppression was defined. This leads to inaccuracies in the object borders affecting the morphological features and reducing the effectiveness of the FP reduction. The morphological features and sequential classification were still able to achieve a 29% reduction in the initial number of regions, but at the end of the global stage an average of 34 detected regions per image across the G1 and G2 sets still remained. In further analysis of the detected regions, it was observed that fatty breasts had relatively few detected structures while mammograms containing dense tissue had a much larger number of regions.

The limitations of the global stage were partially overcome by repeating the filtering and edge detection in the local regions identified in the global stage. By allowing the DWCE filter to adapt to the background within these much

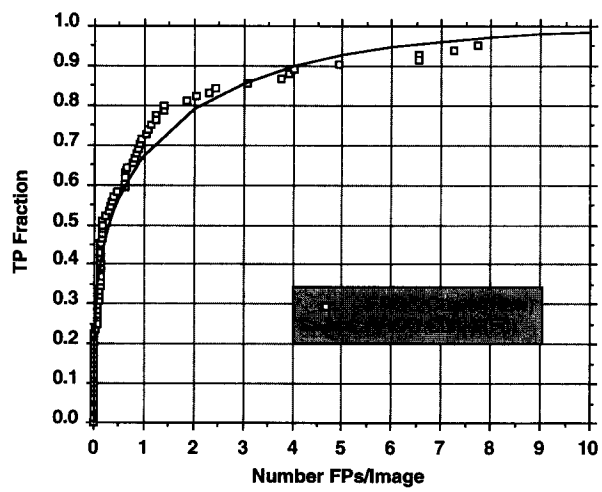

FIG. 7. FROC curves obtained with the image group {Train $G2_i$, Test G2}. The data points are raw data obtained by varying the decision threshold on the discriminant scores. The solid curve is obtained from the FROCFIT program.

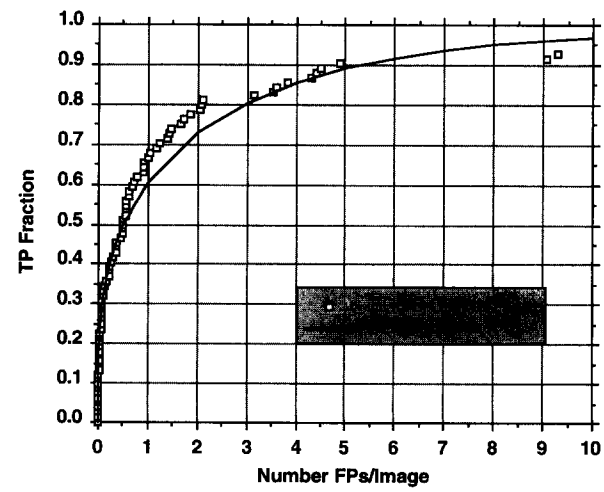

FIG. 9. FROC curves obtained with the image group {Train $G1_i$, Test G2}. The data points are raw data obtained by varying the decision threshold on the discriminant scores. The solid curve is obtained from the FROCFIT program.

smaller regions, better estimates for the true borders of the mammographic structures were achieved without sacrificing true mass detections. The local DWCE stage was able to detect 160 of the 168 true masses in this preliminary study where, again, 83 of the 85 malignant masses were detected. The additional missed mass did not come from the DWCE filtering and edge detection but was instead lost in the local FP reduction. This mass was detected in a different mammogram from our image set which contained a different view of the same breast. It is evident by comparing Figs. 3(g) and 3(h) that the detected objects in the local stage match the true borders better than the global stage objects. The average area of the detected objects following the local FP reduction increased to 253 pixels from the 64.4 pixels following the global stage. The more accurate borders help improve the local FP reduction which provided a 69% average reduction in the initial number of local FP regions and a corresponding 50% reduction in the number of FPs from the output of the global stage. The number of detected regions following the local FP reduction was still quite large, with an average of 16.9 regions detected per image. This large number of regions can be attributed to two factors. First, while improving the object border estimates, the local processing still continued to underestimate their true size [see Fig. 3(h)]. This limited the effectiveness of the morphological FP reduction in distinguishing between the masses and many of the normal structures. In addition, the expanded object area was attributed not only to the more precise edge characterization but also to the merging of neighboring regions into single detected objects. The merged objects caused problems in the final texture analysis stage because the texture information for the mass regions was often averaged with large amounts of normal tissue, thus increasing the likelihood that the true breast masses would be missed. Object splitting partially solved the problem of merged regions by estimating merge points according to geometrical shape. However, some distortion of the morphological features remained. Splitting also inadvertently introduced additional FPs. In this study the number of FPs increased from 16.9 FPs/image after local reduction to 20.3 FPs/image after the splitting reduction step. While the results of this preliminary study indicate that the DWCE segmentation is effective in detecting breast masses, further improvements in the scheme will be necessary to reduce the total number of detected regions.

Closer evaluation of the images where a mass was not detected highlighted a problem in the initial rescaling of some images. As stated previously, the initial rescaling step in the DWCE [refer to Fig. 1(a)] is very important because it allowed a single set of filters to be applied uniformly to all the mammograms. The rescaling should have occurred only within the breast region of the image. We have found that the initial breast map included a strip of pixels belonging to a bright edge outside the breast region of the mammogram in two of the images with missed masses. The pixels in this strip had a higher intensity than any of the other pixels in the breast region of the mammogram, and their inclusion in the rescaling caused many lower-intensity objects to be missed. When this strip was removed from the two images, the masses were detected.

B. Morphological classification

Another important factor that affects the FP reduction is the choice of morphological features. The eleven morphological features used in this study were selected because individually they showed some potential in differentiating between shapes. However, they are probably not the optimal set of morphological features for this task. The best subgroup of the features was found to be the area, perimeter-to-area ratio, and the contrast which provided the best BPN classifier performance. No general conclusions from this preliminary study can be made about the applicability of the individual features because of the small size of the image set and the suboptimal border information provided by the DWCE detection.

The morphological classification is an important component in the overall FP reduction. In this study, we selected the sequential application of a thresholding, an LDA, and a BPN classifier. The order of application was found to be important. The investigation showed that the LDA and especially the BPN classifier were trained faster and performed better when the initial number of FPs in the training set was small, thus leading to the use of the sequential classification scheme. We have not presented the exact values of the fixed thresholds used in this study because of the small size of this preliminary image set. With a larger, more representative training set, the particular threshold values will need to be adjusted. Therefore, we have instead concentrated on describing the general methodology for selecting the individual thresholds, as outlined in Sec. II B.

C. Texture classification

The large number of regions detected in the DWCE segmentation precipitated the need for additional FP reduction. This additional reduction was achieved by classifying with multiresolution texture features extracted from the DWCE detected regions. The LDA classification using SGLD features was selected because it was found to be effective in differentiating breast masses from normal tissue in regions identified by radiologists.[14,23] Again, we have not presented details about the particular feature selected because of the small size of the data set. However, a detailed discussion of the multiresolution texture features and the LDA texture classification method can be found in the literature.[14,23] The texture classification in this final step resulted in an average of 4.4 FPs/image at a 90% TP rate and 2.3 FPs/image at an 80% TP rate (Table III) in the test sets. These results indicate that the overall system (i.e., DWCE segmentation plus LDA texture classification) is capable of automatically detecting breast masses on digitized mammograms. Table III also indicates that the G1 and G2 image sets were reasonably well matched. The G2 set provided slightly better performance at a 90% TP fraction but the G1 set's performance was better for the 80% detection level.

Figures 6–9 contain fitted FROC curves obtained using the FROCFIT program developed by Chakraborty *et al.*[32] Table IV contains the estimated fit parameters and the goodness of fit characteristics obtained with the program. The fitted curves match well visually with the raw FROC results and the normalized χ^2 goodness of fits only varied from 0.92 to 1.39 (optimal value is 1.0). This indicates that the FROCFIT program may be able to fit our raw FROC results. However, it is likely the signals detected by our method do not satisfy the assumptions that the occurrence of an FP follows Poisson statistics and that the FPs are independent. Further studies are therefore needed to investigate if the good fit observed occurs by chance and if the area under the alternative FROC curve (A_{AFROC}) can be used as an indication of the overall performance of the classification system.

D. Future studies

Our results indicate that DWCE segmentation can be used to effectively detect breast structures on a mammogram. The flexible form of the DWCE filter leaves open the possibility that further optimization of the detection parameters may improve overall performance. Evaluation of different DWCE filters (e.g., modifying K_M and K_{NL}) will be pursued in future studies.

One of the difficulties in the DWCE segmentation method is the merging of regions and the subsequent need to split objects. The splitting operation increases the number of false regions and also adversely affects the morphological information by introducing straight edges at the split locations. In future studies, we will investigate alternative methods for separating merged structures. Gray-level information will be used in conjunction with binary shape information to guide the splitting. The sequential change in shape obtained by region growing at different local threshold levels will more precisely define multiple regions within a single DWCE segmented object. This approach should improve the morphological features of the split objects and increase the classification accuracy of masses and normal tissue, thereby reducing the FP detections. Furthermore, a fundamental improvement in the adaptivity of the DWCE segmentation will be needed to reduce the number of objects extracted in the initial stage. One possible improvement may be accomplished by first classifying the breast parenchyma into different types (e.g., fatty, mixed, or dense). The DWCE filter parameters can then be optimized specifically for each tissue type. This would allow better background suppression and more precise object extraction in different types of breast parenchyma. It can be expected that the initial number of FP objects detected in dense breasts will be reduced without impacting the detection on fatty breasts.

Our detection scheme makes use of information on a single mammogram. In mammographic interpretation, it has been found that symmetry information on the left and right mammograms of the same view often improves the detection of subtle abnormal tissue density.[33] The information can also be used to eliminate FP detections when they appear on both mammograms in symmetrical locations.[10] However, the symmetry information should be used with caution because many patient mammograms are not highly symmetrical due to variations in compression and imaging techniques, as well as the natural asymmetry in tissue structures. We will investigate the effectiveness of the symmetry information from paired mammograms in FP reductions in future studies.

V. CONCLUSION

We have developed an image enhancement technique which can adaptively suppress the low-frequency structured background and enhance the contrast of structures on an image. The technique was applied to the segmentation step in a CAD program for detection of breast masses. It was found to be effective in enhancing masses and normal tissue structures on mammograms. To further distinguish between masses and normal tissue, the potential mass regions were classified with an LDA using multiresolution texture features extracted from wavelet coefficients at several scales. Results of FROC analysis indicate that the current algorithm can achieve a TP rate of 90% at 4.4 FPs/image and a TP rate of 80% at 2.3 FPs/image. The consistency in the performance of the algorithm was verified by training and testing two independent data sets. This study demonstrates the feasibility of our approach to computer-assisted detection of masses in mammographic interpretation. Further investigations are under way to improve the detection accuracy and test its performance in large data sets.

ACKNOWLEDGMENTS

This work is supported by a USPHS Grant No. CA 48129 and U.S. Army Grant No. DAMD 17-93-J-007 (through subgrant No. GU RX 4300-803UM from Georgetown University). The content of this publication does not necessarily reflect the position of Georgetown University or the government, and no official endorsement of any equipment or product of any company mentioned in the publication should be inferred. The authors thank Dev Chakraborty, Ph.D. for providing the FROCFIT program.

[1]C. C. Boring, T. S. Squires, T. Tong, and S. Montgomery, "Cancer statistics, 1994," CA: Cancer J. Clin. **44**, 7–26 (1994).
[2]B. A. Porter, V. Taylor, J. P. Smith, and V. Tsao, "Contrast-enhanced magnetic resonance mammography," Acad. Radiol. **1S1**, S36–S46 (1994).
[3]F. Shtern, C. Stelling, B. Goldberg, and R. Hawkins, "Novel technologies in breast imaging: National cancer institute perspective," in *2nd Post-Graduate Course Syllabus* (Society of Breast Imaging, Orlando, FL, 1995), pp. 153–156.
[4]C. J. Vyborny and M. L. Giger, "Computer vision and artificial intelligence in mammography," Am. J. Roentgenol. **162**, 699–708 (1994).
[5]W. P. Kegelmeyer, Jr., "Computer detection of stellate lesions in mammograms," Proc. SPIE Biomed. Image Process. **1660**, 446–454 (1992).
[6]W. P. Kegelmeyer, Jr., 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]S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," IEEE Trans. Med. Imag. **8**, 377–386 (1989).
[8]W. Qian, L. P. Clarke, M. Kallergi, and R. A. Clark, "Tree-structured nonlinear filters in digital mammography," IEEE Trans. Med. Imag. **13**, 25–36 (1994).

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

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

[11]T.-K. Lau and W. F. Bischof, "Automated detection of breast tumors using the asymmetry approach," Comput. Biomed. Res. **24**, 273–295 (1991).

[12]R. M. Nishikawa, R. C. Haldemann, J. Papaioannou, M. L. Giger, P. Lu, D. E. Woverton, R. A. Schmidt, U. Bick, K. J. Munn, and K. Doi, "Initial experience with a prototype clinical intelligent mammography workstation for computer-aided diagnosis," Proc. SPIE Med. Imag. Image Process. **2434**, 65–71 (1995).

[13]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. Imag. **15**, 59–67 (1996).

[14]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: Multi-resolution texture analysis," Med. Phys. **22**, 1501–1513 (1995).

[15]T. Peli and J. S. Lim, "Adaptive filtering for image enhancement," Opt. Eng. **21**, 108–112 (1982).

[16]W. K. Pratt, *Digital Image Processing* (Wiley, New York, 1978).

[17]D. Marr and E. Hildreth, "Theory of edge detection," Proc. R. Soc. London Ser. B Biolog. Sci. **207**, 187–217 (1980).

[18]J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computerized image analysis," IEEE Trans. Med. Imag. **12**, 664–669 (1993).

[19]B. Sahiner, H.-P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograms: A convolution neural network classifier with spatial domain and texture images," IEEE Trans. Med. Imag. (in press).

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

[21]B. Sahiner, H.-P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Image classification using artificial neural networks," Proc. SPIE Med. Imag. Image Process. **2434**, 838–845 (1995).

[22]J. A. Freeman and D. M. Skapura, *Neural Networks: Algorithms, Applications and Programming Techniques* (Addison-Wesley, Reading, MA, 1991).

[23]D. Wei, H.-P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction techniques for the detection of masses on digital mammograms: Global and local multi-resolution texture analysis," Med. Phys. (submitted).

[24]A. Petrosian, H.-P. Chan, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computer-aided diagnosis in mammography: Classification of mass and normal tissue by texture analysis," Phys. Med. Biol. **39**, 2273–2288 (1994).

[25]R. W. Conner, "Toward a set of statistical features which measure visually perceivable qualities of textures," in Proc. IEEE Conf. Pattern Recogn. Image Process. 382–390 (1979).

[26]R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," IEEE Trans. Sys. Man Cybern. **3**, 610–621 (1973).

[27]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 Med. Imag. Image Process. **2434**, 606–611 (1995).

[28]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).

[29]I. Daubechies, "The wavelet transform, time-frequency localization, and signal analysis," IEEE Trans. Inf. Theory **36**, 961–1005 (1990).

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

[31]M. J. Norušis, *SPSS Professional Statistics 6.1* (SPSS Inc., Chicago, IL, 1993).

[32]D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristics (froc) data," Med. Phys. **16**, 561–568 (1989).

[33]L. Tabár and P. B. Dean, *Teaching Atlas of Mammography*, 2nd ed. (Georg Thieme Verlag, New York, 1985).