Classification of mass and normal breast tissue on digital mammograms: Multiresolution texture analysis

Datong Wei, Heang-Ping Chan,[a] Mark A. Helvie, Berkman Sahiner, Nicholas Petrick, Dorit D. Adler, and Mitchell M. Goodsitt
Department of Radiology, University of Michigan, Ann Arbor, Michigan

(Received 16 August 1994; accepted for publication 31 May 1995)

We investigated the feasibility of using multiresolution texture analysis for differentiation of masses from normal breast tissue on mammograms. The wavelet transform was used to decompose regions of interest (ROIs) on digitized mammograms into several scales. Multiresolution texture features were calculated from the spatial gray level dependence matrices of (1) the original images at variable distances between the pixel pairs, (2) the wavelet coefficients at different scales, and (3) the wavelet coefficients up to certain scale and then at variable distances between the pixel pairs. In this study, 168 ROIs containing biopsy-proven masses and 504 ROIs containing normal parenchyma were used as the data set. The mass ROIs were randomly and equally divided into training and test groups along with corresponding normal ROIs from the same film. Stepwise linear discriminant analysis was used to select optimal features from the multiresolution texture feature space to maximize the separation of mass and normal tissue for all ROIs. We found that texture features at large pixel distances are important for the classification task. The wavelet transform can effectively condense the image information into its coefficients. With texture features based on the wavelet coefficients and variable distances, the area A_z under the receiver operating characteristic curve reached 0.89 and 0.86 for the training and test groups, respectively. The results demonstrate that a linear discriminant classifier using the multiresolution texture features can effectively classify masses from normal tissue on mammograms.

Key words: mammography, computer-aided diagnosis, mass, wavelet transform, multiresolution texture analysis, linear discriminant classifier

I. INTRODUCTION

Mammography is considered the most reliable method for the early detection of breast cancers.[1] However, it has been reported that radiologists do not detect all breast cancers that are visible on mammograms in retrospective studies.[2–5] Previous studies indicate that computer-aided diagnosis (CAD) can provide a second opinion to the radiologists and potentially decrease the missed detection rate.[6,7] Computerized classification of the malignant or benign features of an abnormality may also be expected to reduce the number of negative biopsies. Improvement in the accuracy of mammography will increase its efficacy for screening and diagnosis of breast cancer.

Computer vision and artificial intelligence techniques have been developed to detect or characterize abnormalities on digital mammograms.[8] Image processing is usually a first step in computer vision to enhance the signal-to-noise characteristics of the objects being detected. Features are then extracted for classification between the signal and the background. Microcalcifications are ideal targets for computer detection due to their clinical relevance, their potential subtlety, and the lack of coexisting normal structures that have the same appearance.[8] The detection and classification of microcalcifications have received a lot of attention and demonstrated significant progress. Breast masses are more difficult to detect and classify than microcalcifications because masses can be simulated or obscured by normal breast parenchyma.[9,10] Fourier analysis of the masses does not show consistent and distinctive high-frequency components. Most of the signal (mass) energy is in the low-frequency region and overlaps with the frequency components of the normal tissue. The gray level changes at the mass boundary are usually gradual and not as abrupt as those at the boundary of microcalcifications. Moreover, the shape, size, and the gray level profile of the masses vary from case to case. These cause difficulties in the application of conventional image processing methods to the detection and feature characterization of masses.

Morphological features have been used to decrease the number of false-positive detections.[11] Spiculated masses were the focus of detection in the analysis of edge orientation in Kegelmeyer's work.[7,12] Breast cancers can also manifest as circumscribed masses.[13,14] Selective median filtering and template matching techniques were proposed to detect suspicious circumscribed masses.[14] For both types of masses, texture features were extracted from regions of interest (ROIs) in digital mammograms and were used in a decision tree to classify the masses from normal tissue with some success.[15]

The discovery of cortical neurons which respond specifically to stimuli within certain orientations and spatial frequencies suggests that multiorientation and multiresolution are part of the biological mechanism of the human visual system.[16,17] Interest in multiresolution image analysis has been growing rapidly in the field of computer vision. A multiresolution representation provides a simple hierarchical framework for analyzing image information. The compression of images by wavelet transforms can achieve a high compression ratio without significant loss of image details,[18] indicating that important image features are condensed in the wavelet coefficients. Texture analysis in the wavelet trans-

 0094-2405/95/22(9)/1501/13/$6.00

form domain was used to distinguish different texture patterns (e.g., French canvas, beach sand, and oriental straw cloth) with some success.[19] Wavelet transform has been applied to mammographic image processing, especially to the enhancement and detection of microcalcifications. Laine *et al.*[17,20] proposed adaptive multiscale processing with wavelet decomposition and reconstruction for feature analysis and contrast enhancement. Richardson[21] discussed the use of wavelet packets that can be superior to wavelets for certain classes of mammographic signals. Qian *et al.*[22] proposed a tree-structured nonlinear adaptive filter and the wavelet transform for the detection and segmentation of microcalcifications on mammograms.

In this paper, we discuss the application of multiresolution texture analysis to digitized mammograms to distinguish mass from normal tissue. Multiresolution texture features were extracted from the spatial gray level dependence (SGLD) matrices (1) of the original image at variable distances, (2) of the wavelet coefficients at different scales, and (3) of the wavelet coefficients up to certain scales and then at variable distances, forming three feature vectors for each ROI. We used stepwise linear discriminant analysis to select features from each of these three texture spaces to maximize the separation of masses and normal tissue. The ability of the three feature vectors for classifying mammographic masses and normal tissue was compared. Receiver operating characteristic analysis was used to evaluate the classification accuracy of the texture features from the different feature spaces.

II. METHODS

A. Database selection

The mammograms used in this study were randomly selected from the files of patients who had undergone biopsies in the Department of Radiology at the University of Michigan. The mammograms were acquired with dedicated mammographic systems with a 0.3-mm focal spot, a molybdenum anode, 0.03-mm-thick molybdenum filter, and a 5:1 reciprocating grid or a stationary grid. The image receptor was a Kodak MinR/MRE screen/film system with extended cycle processing. Our selection criterion was that a biopsy-proven mass could be seen on the mammogram. Initially, more than 300 mammograms were acquired. To avoid the effect of the repetitive grid pattern on the texture feature calculation and the classification, all mammograms with grid lines were excluded. Our final data set was composed of 168 mammograms.

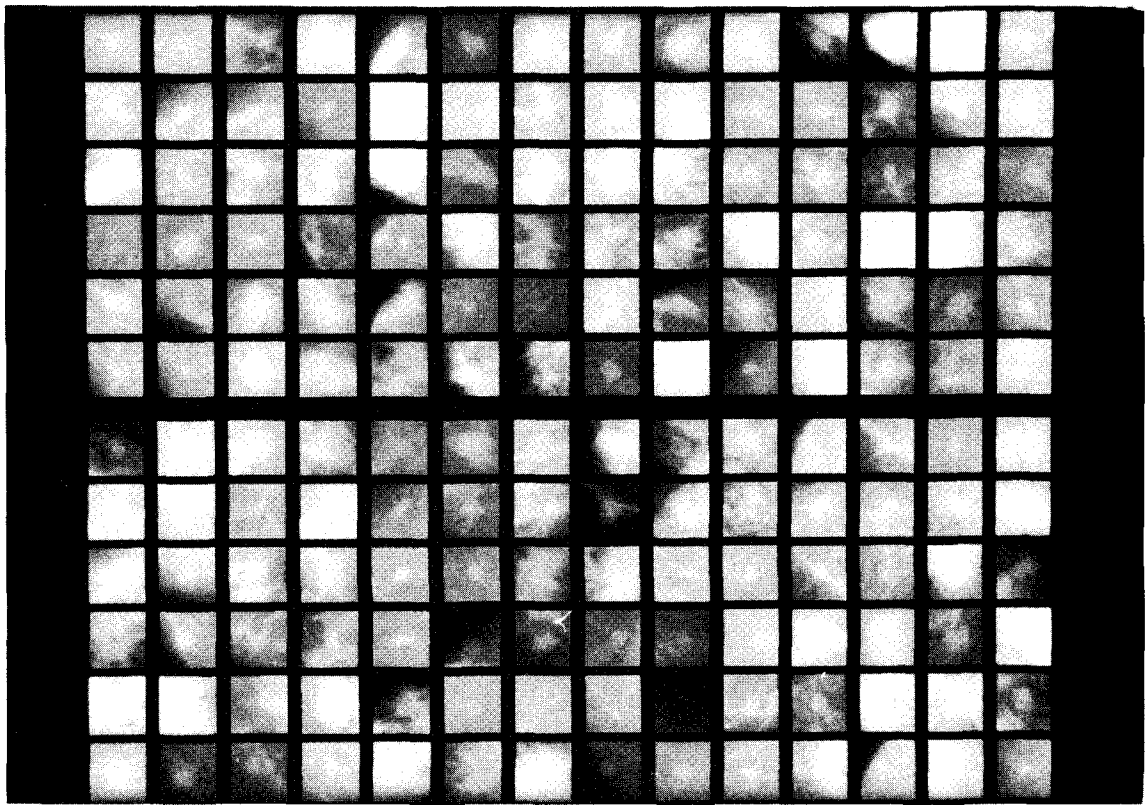
The mammograms were digitized with a laser film scanner (LUMISYS DIS-1000) at a pixel size of 0.1 mm×0.1 mm and 4096 gray levels. The light transmitted through the mammographic films was amplified logarithmically before digitization. After the calibration, the pixel values were linearly proportional to the optical density in the range of 0.1–2.8 optical density units. The slope of the calibration curve decreases gradually outside this range.

Before an automated computer segmentation procedure was developed, we used manual ROI extraction to study the feasibility of using texture features for the classification of mass and normal tissue in all types of breast parenchyma. Four different ROIs, each with 256×256 pixels, were selected manually from each mammogram. One ROI contained a true mass which was identified by an experienced mammographer. A second contained normal parenchyma including the densest tissue on that mammogram, a third, mixed dense/fatty tissue, and a fourth, fatty tissue. Figure 1 shows the 672 ROIs from the 168 mammograms in reduced spatial resolution. The 168 case samples in the data set contained a mixture of benign ($n=83$) and malignant ($n=85$) masses. Forty-five of the malignant masses and six of the benign masses were spiculated. The visibility of the masses was ranked by experienced radiologists on a scale of 1–10 (1 =most obvious, 10=most subtle), which corresponded to the range of masses seen on clinical mammograms. The length of the long axis (size) of the masses was also measured by the radiologists. The distributions of the visibility scores and the sizes are shown in Fig. 2. It can be seen from Figs. 1 and 2 that the masses with different shapes and visibility found in clinical practice were fairly well represented in the data set.

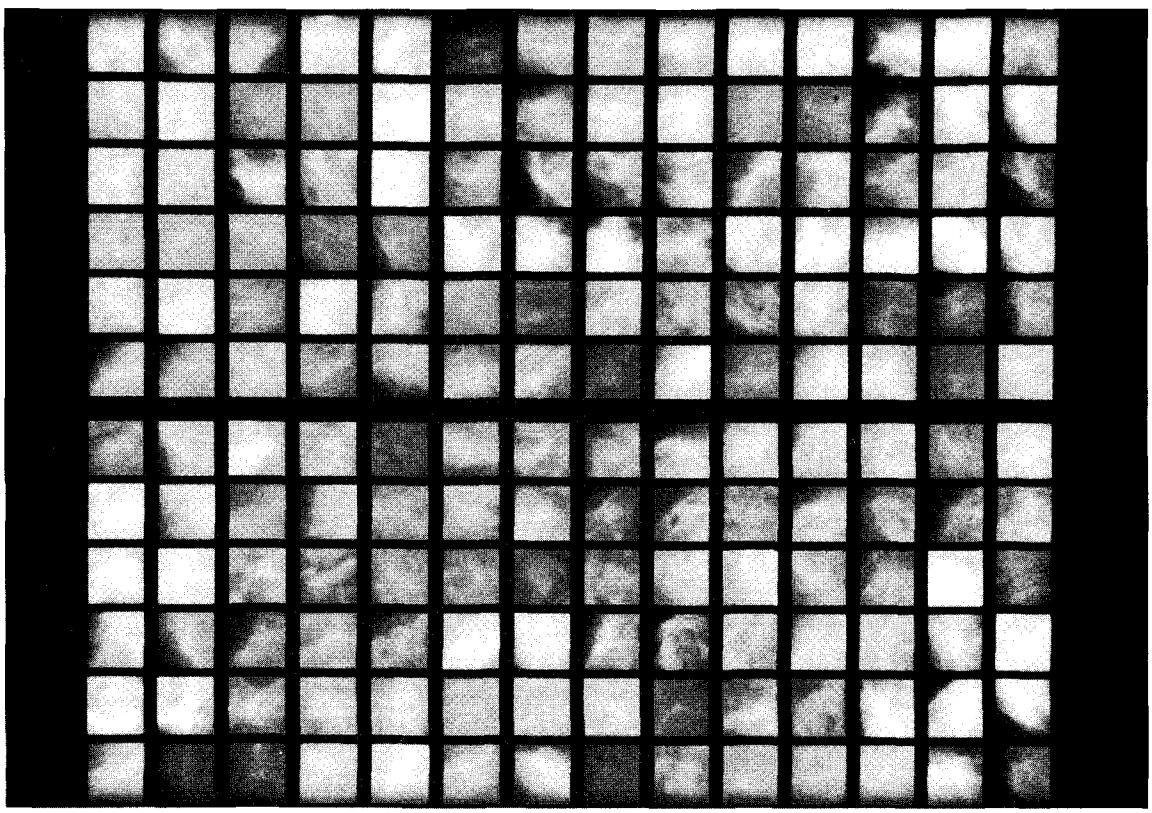
B. Texture features

The input images were digitized to 12 bits of resolution. The average gray level of each ROI of the images was removed and replaced by a constant for all the ROIs before the texture analysis and wavelet transform were performed in order to reduce the variability of the texture features caused by exposure conditions. The texture features were calculated based on the SGLD matrix, also known as the concurrence or co-occurrence matrix.[23,24] The (i,j)-th element of the SGLD matrix, $p_{d,\theta}(i,j)$, is the joint probability that the gray levels i and j occur in a direction θ at a distance of d pixels apart (d is the distance in terms of number of pixels and is referred to as pixel distance in the following discussion) over the entire ROI. The joint probability describes the frequency that a pair of gray level values occurs between pixel pairs with a defined, relative spatial relationship. The SGLD matrix is a two-dimensional histogram. The matrix size depends on the gray level resolution (i.e., the bit depth) of the digitized image and the bin width used in determining the histogram. If the gray level resolution is n bits and the bin width is b gray levels, then the size of the SGLD matrix will be $a\times a$, where $a=2^n/b$. For example, for a 12-bit image, the matrix size of an SGLD matrix constructed with a bin width of 1 gray level is 4096×4096. The matrix size is reduced to 256×256 if a bin width of 16 gray levels is used. The increased bin width is equivalent to reducing the gray level resolution of the 12-bit image to 8 bits by eliminating the 4 least significant bits and using a bin width of 1 gray level in determining the SGLD matrix. Based on the findings of our previous study,[25] 8-bit gray level resolution provided the best classification accuracy when texture features calculated at a fixed pixel distance d were used. Therefore, 8-bit gray level resolution was chosen for the formulation of the SGLD matrices in this study.

Eight texture features were examined: correlation, energy, entropy, inertia, inverse difference moment, sum average, sum entropy, and difference entropy. Some of the texture features can be used to describe some visual properties of the images while others may be more abstract. For example, cor-

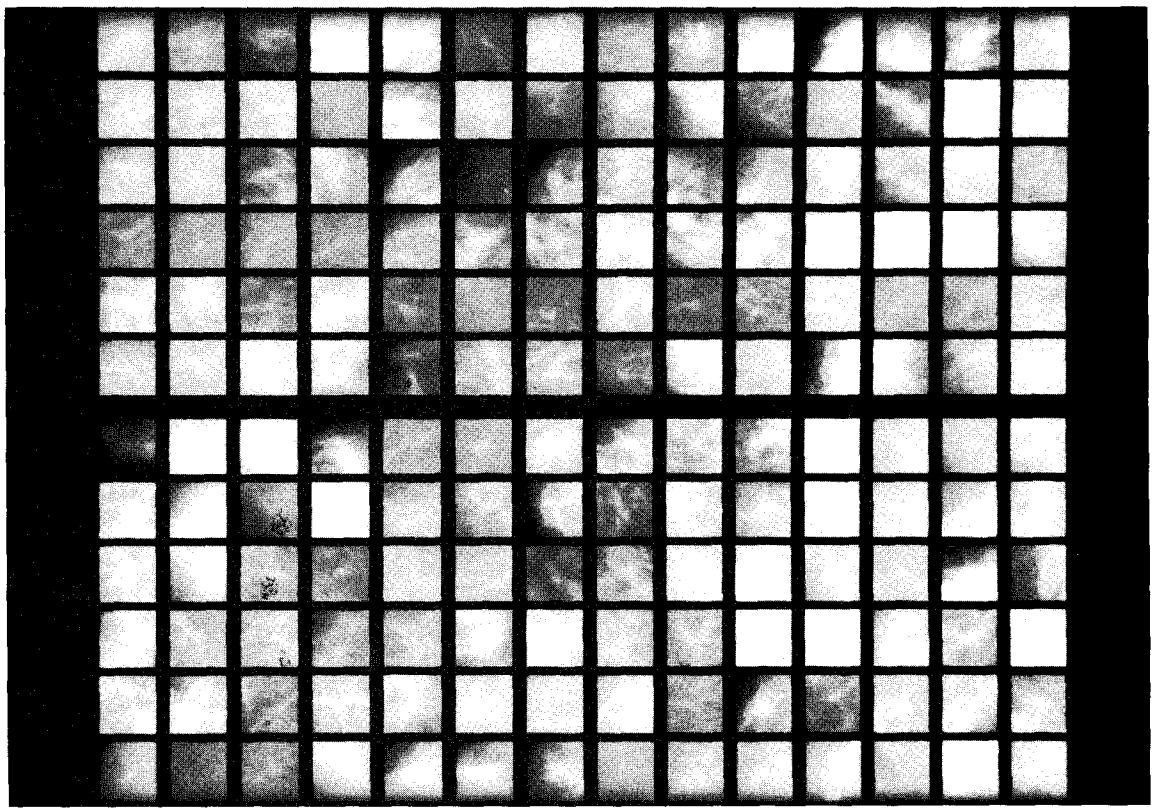


(a)

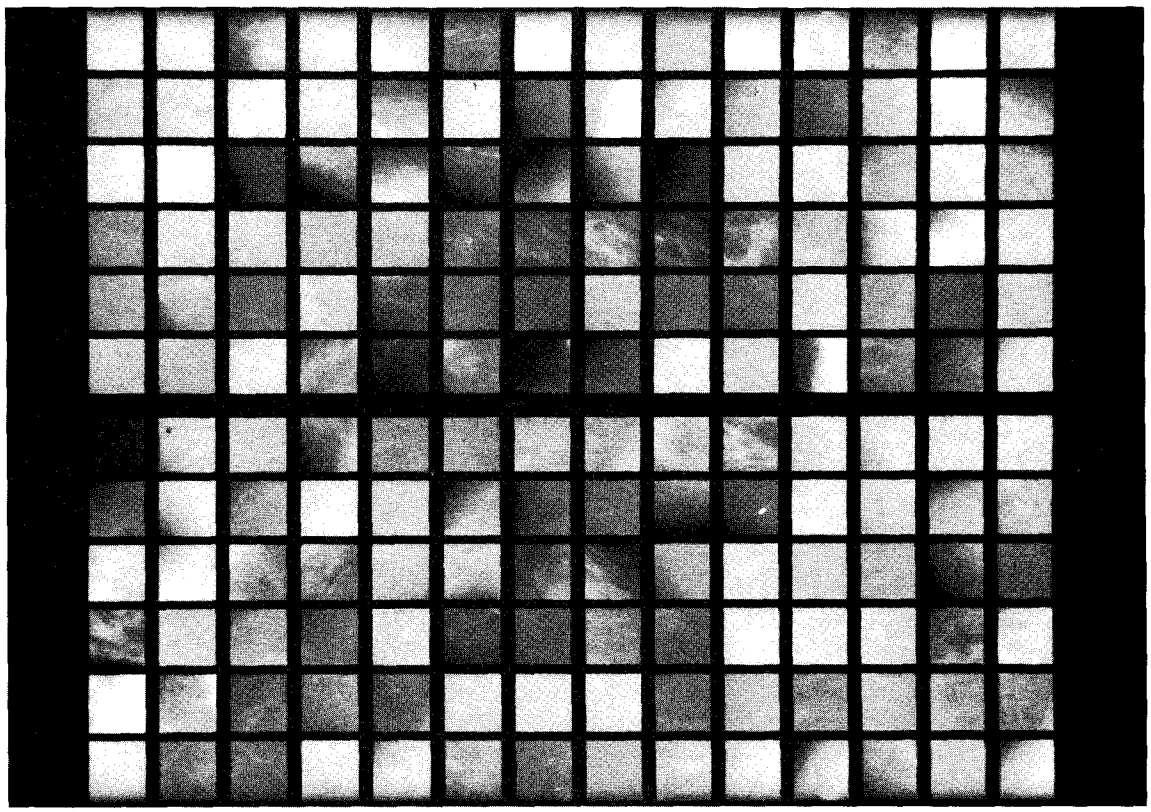


(b)

FIG. 1. The 168 case samples used in this study with ROIs containing (a) biopsy-proven masses, (b) dense breast tissue, (c) mixed dense/fatty breast tissue, and (d) fatty breast tissue. The upper halves are the G_1 cases and the lower halves the G_2 cases.



(c)



(d)

FIG. 1 (*Continued.*)

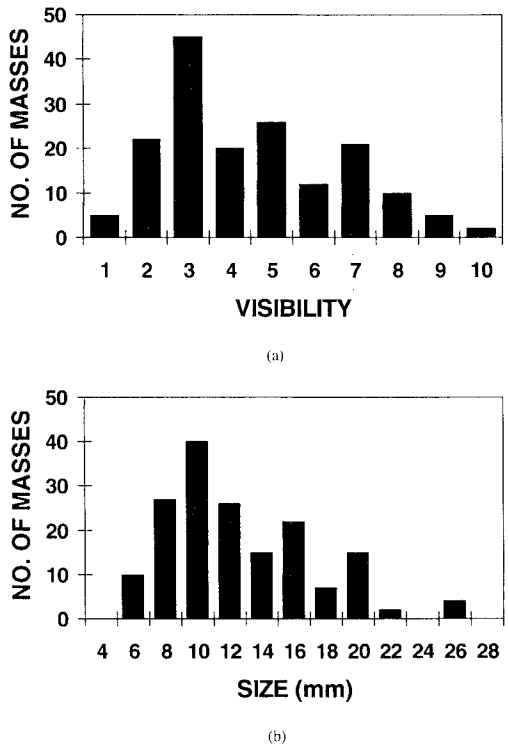


FIG. 2. The distribution of (a) the visibility score and (b) the size of the 168 masses.

relation is a measure of gray level dependency. Energy (or angular second moment) and entropy are measures of pixel homogeneity. Inertia (or contrast) represents the amount of intensity variation. It is difficult, however, to relate specific image characteristics to each of these features. The mathematical definitions of the features can be found in the literature[15,23,24] and are given in Appendix B.

Each texture feature was calculated at $\theta=0°$, 45°, 90°, and 135° for specified distances and/or scales. Since it is expected that the shape and the texture of masses in the ROIs do not have angular preferences, we averaged the features at $\theta=0°$, 90°, and at $\theta=45°$, 135°, and referred to these averaged features as features at 0° and 45°, respectively, in the following discussion. For a given pixel distance, the actual distance between the pixels on the image at 45° was equal to $\sqrt{2}$ times the actual distance at 0°. When the pixel distance increased, the differences in the actual distances between these angles become more significant. Because the texture features depended on the actual distance between the pixel pairs, the features at the two angles were treated separately in our multiresolution texture analysis.

C. Wavelet transform

The wavelet transform produces a multiscale representation of an image in which the geometric structures of the image are preserved within each sub-band or level. In Appendix A, we present a brief introduction to the wavelet transform. More details of the theory and applications can be found in the literature.[26–29]

Mallat presented a multiresolution framework with the discrete wavelet transform inherently embedded.[29] In this framework, the original image (Y) that has the highest resolution is referred to as level 0 ($i=0$) or scale 1 ($s=2^i|_{i=0}$). At scale 2, the original image is decomposed in the wavelet transform domain (similar to the spatial frequency domain in Fourier transform) into a low-pass sub-band image $Y_{2^i}^{LL}|_{i=1}$ (referred to as approximation image at level 1 or scale 2, low-pass low-pass quadrant) and three bandpass sub-band images Y_2^{LH}, Y_2^{HL}, Y_2^{HH} (referred to as detail images in the low-pass high-pass, high-pass low-pass, and high-pass high-pass quadrants). At the next scale (scale 4), the approximation image at scale 2, Y_2^{LL}, is decomposed further into a low-pass sub-band approximation image Y_4^{LL} and three more bandpass sub-band images Y_4^{LH}, Y_4^{HL}, Y_4^{HH}. The decomposition can be stopped at some desired (lower) resolution or (larger) scale. Figures 3(a) and 3(b) illustrate the wavelet decomposition to level 4 or scale 16 of the ROIs containing a mass and normal parenchyma, respectively. The reconstruction of an image from the wavelet coefficients in the transform domain starts from the lowest resolution (largest scale) sub-band images.

In this study, Daubechies' filter with four coefficients[27]

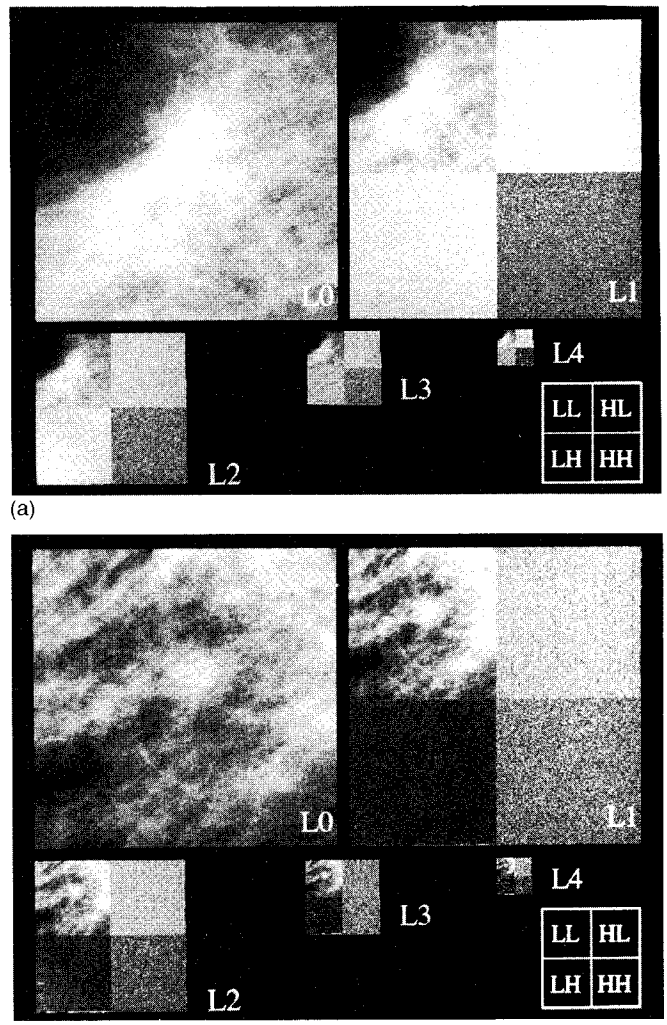


FIG. 3. Wavelet decomposition from level 0 ($L0$ or scale 1) to level 4 ($L4$ or scale 16) of (a) an ROI with a mass and (b) an ROI with normal breast tissue.

was used as the wavelet filter for image decomposition (see Appendix A). A filter with a small number of wavelet coefficients was chosen because the width of the uncertainty band at the image boundary caused by convolution would be narrower. This allowed the decomposition to be performed to larger scales while still providing a sufficient number of usable pixels in the approximation image for the construction of an SGLD matrix. The chosen filter was also separable so that the fast wavelet transform algorithm could be employed in two-dimensional image analysis.

D. Multiresolution texture feature space

We used the original image Y (scale 1) and the low-pass sub-band approximation image $Y_{2^i}^{LL}$ (scale 2^i, $i=1,...,4$) to formulate SGLD matrices at multiple scales. The distance of the pixel pairs used at each scale was one pixel. The decomposition stopped at scale 16 so that the approximation image in the transform domain had 16×16 pixels. Effectively, the pixel distances of SGLD matrices formulated in this way at scales of 1, 2, 4, 8, and 16 corresponded to pixel distances of 1, 2, 4, 8, and 16 in the original image. A total of 80 features were calculated from each ROI (8 features×2 angles×5 levels) in this feature space. These 80-dimensional feature vectors based on the wavelet transform were denoted as $\mathbf{F}_{WT}$.

As the scale in the wavelet transform increased, the statistical fluctuations in the SGLD matrices based on the smaller and smaller images could not be neglected due to the random sample errors. To reduce the statistical error in the SGLD matrices, we decomposed the original ROIs by wavelet transform to scale 4 so that the smallest image size was 64×64 pixels. Then the wavelet filter was applied once more without downward sampling. The resulting wavelet coefficients were obtained at scale 8 and were overcomplete and redundant.[29,30] However, this allowed the number of pixels used to construct the SGLD matrices to be kept at 64×64. The SGLD matrices at scale 8 were then constructed with distances of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12. These distances between pixel pairs were equivalent to the distances of 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, and 48 in the original image. Therefore, a total of 224 features were calculated from each ROI [8 features×2 angles×(1 pixel distance at scales 1, 2, 4+11 pixel distances at scale 8)] in this feature space. The feature vectors in this 224-dimensional feature space were based on wavelet transform and variable distances, and were denoted as $\mathbf{F}_{WV}$.

To evaluate the effect of the wavelet transform on the classification results, we compared the features described above to those extracted from the SGLD matrices of the original image. The SGLD matrices were constructed with pixel distances of 1, 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, and 48. These distances corresponded to those used in the calculation of $\mathbf{F}_{WV}$ when the latter were converted to equivalent pixel distances in the original image. Therefore, a total of 224 features were calculated from each ROI (8 features×2 angles×14 pixel distances) in this feature space. These feature vectors based on SGLD matrices from the original images with variable distances were denoted as $\mathbf{F}_{VD}$. From the 224 features, we could also select a subset of 80 features at $d=1$, 2, 4, 8, and 16. The pixel distances in this subset corresponded to the pixel distances used for the calculation of the features in $\mathbf{F}_{WT}$. The 80-dimensional feature vectors obtained from this subset of features with variable distances were denoted as $\mathbf{F}_{VDS}$. The impact of the wavelet transform on the discriminant power of the texture features was studied by comparing the classification results obtained with $\mathbf{F}_{VD}$ and $\mathbf{F}_{VDS}$ to those obtained with $\mathbf{F}_{WV}$ and $\mathbf{F}_{WT}$, respectively.

E. Linear discriminant analysis

Linear discriminant analysis[31] is a systematic statistical technique to classify individuals or cases into one of the mutually exclusive classes based on certain indices or predictor variables. These indices or predictor variables may have certain correlations with one another. In a two-class classification problem, for example, a linear combination of these variables is formed and the coefficients are determined based on certain optimization criteria. One of such criteria, proposed by Fisher, is that the ratio of the difference of the means of the linear combination in the two classes to its variance is maximized.[31,32]

The discriminant analysis in the SPSS software package [M. J. Norusis, *SPSS for Windows Professional Statistics*, Release 6.0 (SPSS Inc., Chicago, IL, 1993)] was used in this study. The extended feature spaces as explained above were each used as a pool of predictor variable candidates for a two-class discriminant analysis that contained a mass class and a normal tissue class. Similar to the situation of multiple linear regression, including a large number of possible predictor variables in the linear model of the discriminant function is not a good strategy. Inclusion of irrelevant variables will not improve the classification accuracy and will decrease the generalization capability of the classifier. Because of the large number of features in the pools, it is a formidable task to test all different feature combinations at different numbers of feature variables to find the best combination. Therefore, we utilized a stepwise feature selection procedure to select predictor variables in each feature space. Five selection criteria are provided in the SPSS package, including (1) the minimization of Wilks' lambda, (2) the minimization of unexplained variance, (3) the maximization of the between-class F statistic, (4) the maximization of Mahalanobis distance, and (5) the maximization of Lawley–Hotelling trace (Rao's V). For each feature space, we tested all available selection criteria. With each criterion, we performed stepwise feature selection on all the 168 cases using the program default values for the inclusion and exclusion threshold parameters and the termination criterion. The selection criterion that provided the best classification result would be chosen. Since the program default values of the parameters might not be the optimal choices for our application, we varied the parameter values of the chosen criterion in an attempt to further improve the classification results. For our data sets, when the thresholds were set higher than the default values, fewer feature variables would be included and the classification accuracy decreased. When the thresholds were set lower than the default values, more features would be included and the classification results might improve. However, when the thresholds were lowered further and too

many features were included, the classification would deteriorate. The set of feature variables that provided the best classification in this selection process were used for the formulation of the discriminant function in the given feature space. For simplicity, we will refer to this stepwise selection procedure with different thresholds as a stepwise or automatic selection process. Our feature selection process was by no means exhaustive. However, it would represent the best selection achievable within reasonable computational requirements.

To evaluate the capability of generalization of a trained classifier, we randomly divided the 168 cases into two groups (G_1 and G_2) of equal size. We used the features selected with the procedure described above as discriminant variables. If a given group was used for training, the feature values of each case from that group were used to optimize the coefficients of the linear discriminant function. The training cases were then classified with the linear discriminant function as a verification of consistency. The other group was used as test cases of which the feature values were input to the classifier and the discriminant score of each case was calculated from the linear discriminant function. One of the two groups was alternately used as the training group so that the variability of the classifier with different training groups could be observed.

Receiver operating characteristic (ROC) analysis[33,34] was used to evaluate the overall performance of the linear discriminant functions, in addition to the classification results reported by the SPSS program under certain prior probability assumptions. For a two-class problem, the ROC curve could be obtained using the Bayes' rule by changing the prior probability. Alternatively, the discriminant score from the canonical discriminant function could be used as the decision variable in the ROC analysis. Figure 4 demonstrates such a distribution of discriminant scores based on the linear combination of features calculated from wavelet coefficients at variable distances. The distribution of the discriminant scores of the ROIs in the training or the test group was input into

TABLE I. Texture features selected by stepwise discriminant analysis.

(a) From $\mathbf{F}_{WT}$ and $\mathbf{F}_{VDS}$

scale	1		2		4		8		16	
pixel distance	1		2		4		8		16	
θ	0°	45°	0°	45°	0°	45°	0°	45°	0°	45°
correlation							Δ		•	•
difference entropy									•	•
energy		•								
entropy										
inertia		•			•	•			•	
inv. dif. moment								•		
sum average	Δ			Δ			•	Δ	•	Δ
sum entropy										•

(b) From $\mathbf{F}_{VD}$ and $\mathbf{F}_{WV}$

scale	1	2	4		8																
Distance	1	2	4		8		12	16	20		24		28	32		36		40	44		48
θ	0°	45°	0°	45°	0°	45°	45°	45°	0°	45°	0°	45°	0°	0°	45°	0°	45°	0°	0°	45°	0°	45°
correlation					□▲		□▲				□▲		□	▲		■□		▲	□	■□	▲	▲
dif. entropy	■		■▲			■						□			□		■□		□		■▲	□
energy						□										□						□
entropy	▲	▲			□		□	□						■	□						□	■□▲
inertia																■		▲				
inv. dif. moment						■□	▲	■		▲								□			■□	
sum average				▲		■	▲		□						▲		■		□	■▲	■	
sum entropy	□												■					□		■		□▲

(a) ● 13 features (automatic)selected from $\mathbf{F}_{WT}$. ∇ 5 features (automatic) selected from $\mathbf{F}_{VDS}$. Note: 0° represents the average of features at 0° and 90°; 45° represents the average at 45° and 135°.

(b) ■ 19 features (automatic) from $\mathbf{F}_{WV}$. □ 29 features (semiautomatic) from $\mathbf{F}_{WV}$. ▲ 20 features (automatic) from $\mathbf{F}_{VD}$. Note: Some distances/angles are not shown if no feature was selected. 0° represents the average of features at 0° and 90°; 45° represents the average at 45° and 135°.

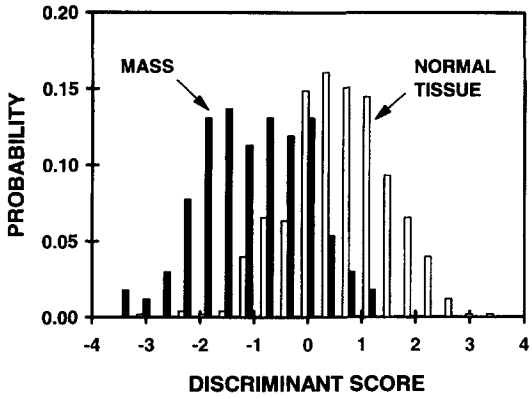


FIG. 4. An example of the probability density distribution of the discriminant scores of the masses and normal tissue. The discriminant scores were calculated from the canonical discriminant function that was optimized with all 672 ROIs with 19 features selected from multiresolution texture feature space $\mathbf{F}_{WV}$.

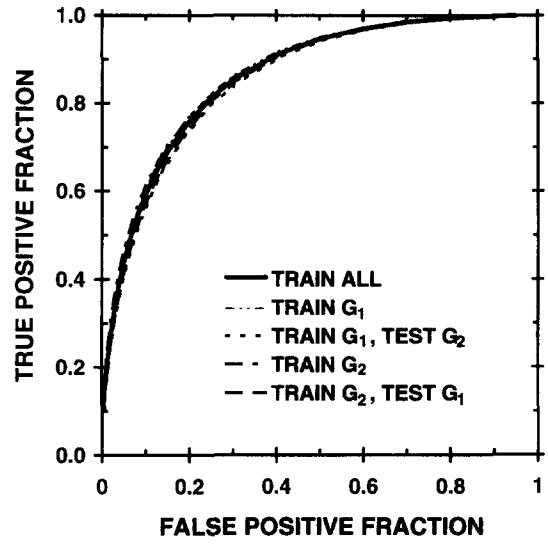


FIG. 5. ROC curves for classifying masses from normal tissue with discriminant function based on 13 features selected from texture feature space $\mathbf{F}_{WT}$.

the LABROC1 program,[35] which provided a maximum-likelihood estimation of a binormal ROC curve for training or testing, respectively. The area under the fitted ROC curve, A_z, was used as a performance index for evaluating the different sets of features selected from different multiresolution feature pools. The standard deviation (SD) of A_z estimated by LABROC1 was also reported. The CLABROC program was employed to test the statistical significance of the difference between A_z values of different sets of selected features.[36] The two-tailed p values were reported in the following comparisons.

III. RESULTS

A. Texture features based on wavelet coefficients

Stepwise feature selection was performed with the multiresolution texture features extracted from the feature space $\mathbf{F}_{WT}$. Thirteen features were selected as shown in Table I(a). The A_z and the estimated SD of the ROC curves are summarized in Table II. Figure 5 shows the ROC curves for the classification using the features derived from the wavelet coefficients. The A_z values of 0.858 and 0.854 for testing of G_1 and G_2, respectively, are higher than those of 0.817 ± 0.027 ($p=0.02$) and 0.829 ± 0.026 ($p=0.10$) obtained with texture features calculated from the SGLD matrix at a single distance of 20 pixels.[25]

B. Texture features based on original images with variable distances

To evaluate whether the improvement of classification over the results using features based on a single distance[25] is caused by the low-pass filtering in the wavelet transform or by the changes in the pixel distances, we used the same 13 features variables selected from $\mathbf{F}_{WT}$ but the feature values were calculated from the SGLD matrices based on the origi-

TABLE II. Comparison of the area under the ROC curves, A_z, obtained from different feature spaces.

Number of Features	Feature Space	Features extracted from scales	Training on G_1 and G_2	Training on G_1 Testing on G_2		Training on G_2 Testing on G_1	
			A_z (Train)	A_z (Train)	A_z (Test)	A_z (Train)	A_z (Test)
13*	$\mathbf{F}_{WT}$	1, 2, 4, 8, 16	0.864±0.016	0.869±0.021	0.854±0.023	0.868±0.022	0.858±0.022
13♦	$\mathbf{F}_{VDS}$	1	0.796±0.019	0.808±0.026	0.781±0.027	0.798±0.027	0.787±0.027
5*	$\mathbf{F}_{VDS}$	1	0.758±0.021	0.766±0.028	0.747±0.029	0.754±0.029	0.760±0.028
20*	$\mathbf{F}_{VD}$	1	0.885±0.014	0.834±0.024	0.837±0.024	0.905±0.018	0.857±0.022
19▼	$\mathbf{F}_{VD}$	1	0.871±0.015	0.883±0.019	0.836±0.025	0.878±0.021	0.859±0.022
19*	$\mathbf{F}_{WV}$	1, 2, 4, 8	0.884±0.014	0.899±0.018	0.853±0.025	0.887±0.021	0.859±0.022
29△	$\mathbf{F}_{WV}$	1, 2, 4, 8	0.887±0.014	0.904±0.018	0.840±0.026	0.903±0.018	0.855±0.022

*Automatic feature selection.
♦Features corresponding to those automatically selected from $\mathbf{F}_{WT}$.
△Features corresponding to those automatically selected from $\mathbf{F}_{WV}$.
▼Semiautomatic feature selection.

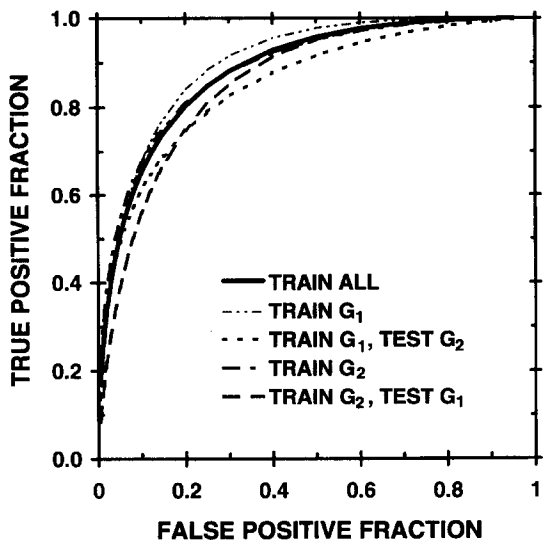


FIG. 6. ROC curves for classifying masses from normal tissue with discriminant function based on 19 features selected from texture feature space $\mathbf{F}_{WV}$.

nal images at equivalent distances, $\mathbf{F}_{VDS}$. The A_z values of the ROC curves with the same training and test groups (Table II) are significantly lower than the corresponding A_z values ($p<0.006$) with features extracted from the wavelet coefficients.

Stepwise feature selection was also performed on the entire data set of 168 cases from the feature space $\mathbf{F}_{VDS}$. The five features selected are listed in Table I(a). When this set of features was used to formulate the discriminant function, there was no improvement in A_z (Table II), compared with the results using the 13 features with feature values from the same $\mathbf{F}_{VDS}$ space. The differences between the A_z values obtained with 13 features from $\mathbf{F}_{WT}$ and the corresponding A_z values obtained with 5 features from $\mathbf{F}_{VDS}$ were statistically significant ($p<0.0002$). When the entire feature space of $\mathbf{F}_{VD}$ was used in the stepwise feature selection, 20 features were selected as listed in Table I(b). The A_z values for classification in both the training and test groups are significantly higher than those obtained with 5 features from $\mathbf{F}_{VDS}$ ($p<0.025$). As can be seen from Table I(b), 12 out of the 20 features were selected from distances greater than 16 pixels. This indicates that the information at larger distances which is not present in $\mathbf{F}_{VDS}$ is important in the classification of mass and normal tissue. Some of the A_z values obtained with these 20 features from $\mathbf{F}_{VD}$ are higher than those with 13 features from $\mathbf{F}_{WT}$, while the others are lower than those obtained from $\mathbf{F}_{WT}$, with p values ranging from 0.06 to 0.77. This is an indication that the discriminant power of the features from $\mathbf{F}_{VD}$ is comparable to that of the features from $\mathbf{F}_{WT}$.

C. Texture features based on wavelet coefficients at variable distances

Figure 6 illustrates the ROC curves for training and testing when stepwise feature selection was performed on the texture features extracted from the feature space $\mathbf{F}_{WV}$. The 19 features selected are listed in Table I(b). As shown by the A_z values in Table II, when the selected features were used to formulate the discriminant function, the classification results for the training sets improved in general, with p values ranging from 0.08 to 0.40, whereas the test results were almost the same as those obtained with the 13 features from $\mathbf{F}_{WT}$, with p values of 0.93 and 0.88. As can be seen from the same table, if these 19 variables were used on the feature values from $\mathbf{F}_{VD}$, the A_z values were similar to or slightly lower than those obtained with $\mathbf{F}_{WV}$. The differences are statistically significant for A_z (training on G_1 and G_2) and for A_z (testing on G_2) at $p<0.03$, and are insignificant for the other A_z values with p values ranging from 0.23 to 0.60.

We also selected the features in two steps, referred to as semiautomatic selection. First, we input texture features of the same type, e.g., correlation, calculated at all scales and distances into the discriminant analysis program. By using the stepwise selection method with reduced thresholds for the F values for variable entry and removal, we found the scales and distances that are important for classification for each texture feature. Then we applied the stepwise procedure again to all features at their selected scales and distances to further reduce the number of features. In this way, 29 features were selected as shown in Table I(b). Although most of them were different from the 19 features selected automatically, the overall classification results did not show much difference, indicating that some of the features used in one discriminant function might be linearly correlated with some of the features in the other discriminant function. The classification results (Table II) improved slightly in the training groups (with p values ranging from 0.08 to 0.74) but deteriorated in the testing groups (with p values of 0.24 and 0.54), probably because the increased number of features used in the discriminant function limited its capability for generalization.

IV. DISCUSSION

A. Multiresolution texture analysis

Textures are generally recognized as being fundamental to perception, although there is no precise definition or characterization of textures available in practice. Intuitively, texture descriptors provide measures of properties such as smoothness, coarseness, and regularity. When an image is composed of elements of texture primitives, the description of the image by texture features can be very effective. One of the advantages is that the texture features are shift invariant and can be made orientation invariant by averaging over various angles. This is very important since the location and orientation of the mass in the ROI can be arbitrary.

The masses found in clinical mammograms have very different shapes and sizes. It is a challenge to find a universal feature or a set of features that can differentiate the masses from the normal tissue and parenchymal structures in the breast. It is also difficult to define *a priori* an optimal resolution for the ROIs. A multiresolution approach could provide a scale-invariant interpretation of an image.

The wavelet transform is closely related to the well-known Fourier transform through the short-time Fourier transform or Gabor transform. It is considered a natural way of decomposing the image energy into different frequency

bands through convolution with the translated and dilated version of a function called the "mother wavelet."[29] Unlike the Fourier transform where the coefficients in the transform domain do not reflect the local spatial variations, the wavelet coefficients retain the spatial variations of the original image.

In the multiresolution framework using wavelet decomposition proposed by Mallat,[29] the transform domain contains a minimum set of coefficients from which the reconstruction of the decomposed image is perfect or lossless. In the successive image decomposition, the approximation image in the current scale is decomposed into an approximation image and three detail images in the larger scale. Once the mother wavelet is chosen, the coefficients, which contain one approximation image and a series of detail images at different scales, are nonredundant and the transform is one-to-one. The extraction and condensation of image information through Mallat's framework are very efficient. Thus the wavelet transform is often used for image compression.[18,37] In the classification and pattern recognition problem, however, the focus is on the extraction of those features that can provide maximum distinction among different classes rather than on the minimal representation of the original image. In our current texture analysis, we used the approximation images at different scales, which are redundant representations of the original image. Such representations may be helpful in classification and pattern recognition applications, as demonstrated by the improvement in classification accuracy in comparison to the results obtained with features at a single distance,[25] or to the results obtained with features at variable distances without wavelet transform.

The discrete wavelet transform can be described as a cascaded process with two basic operations: filtering and down sampling. There are certain requirements for a filter to be wavelet filter.[26] Although it is possible to find optimal wavelet filters for certain types of images, our focus in this work is on the feasibility of multiresolution features for classification of masses from normal tissue rather than the optimization of this procedure. Therefore, an orthonormal four-weight Daubechies' filter with compact support[27] was used for our image decomposition. When the down-sampling process effectively reduces the image size by a factor of 2 in each direction as the scale increases, the reduced size of the distortion at the boundary will help keep as much useful image information as possible for texture calculation.

B. Comparison of classification accuracy with features from different feature spaces

To compare the discriminant power of texture features calculated from the wavelet coefficients to those from the original images, we used the feature variables with equivalent distances ($\mathbf{F}_{VDS}$). Using the features selected by the stepwise procedure, the classification results based on the features from $\mathbf{F}_{WT}$ were significantly better than those based on the features from $\mathbf{F}_{VDS}$. If we used the 13 features automatically selected from $\mathbf{F}_{WT}$ but formulated the discriminant functions based on the texture feature values from $\mathbf{F}_{VDS}$, the classification results demonstrated similar differences. This indicates that the texture features at equivalent distances from the wavelet transform domain have better discriminant power than those from the original images. However, when texture features up to distances of 48 ($\mathbf{F}_{VD}$, corresponding to 4.8 mm for 0° features and 6.79 mm for 45° features) are available for feature selection, the discriminant power of the texture features from the original images can reach as high as that of the features from $\mathbf{F}_{WT}$ or $\mathbf{F}_{WV}$. As can be seen from the features selected from each space shown in Tables I(a) and I(b), the texture information at large distances is important for the classification task. The feature space $\mathbf{F}_{VDS}$ does not provide such important information, resulting in poor classification. On the other hand, although the features in $\mathbf{F}_{WT}$ were calculated at distances equivalent to those of $\mathbf{F}_{VDS}$, the low-pass filtering effectively increases the correlation distances of the features. The structural information and energy of the original image obtainable at larger distances than the maximum equivalent distance of 16 pixels are condensed into the wavelet coefficients used for the calculation of $\mathbf{F}_{WT}$. The fact that the features from $\mathbf{F}_{WV}$ do not provide significant improvement (at least for the test groups) in the classification results indicates that the compression of image information is efficiently accomplished by the wavelet transform so that the additional information in $\mathbf{F}_{WV}$ is redundant as expected.

The overall operations of the discrete wavelet transform can be summarized as bandpass filtering (including low-pass and high-pass filtering) and downward sampling (decimation). The approximation images with the wavelet coefficients are the result of the low-pass filtering from convolution with the orthogonal scaling function.[29] The detail images obtained through convolution with the orthogonal wavelet function contain the edge (or high-frequency) information of the images. The texture features based on the multiresolution approximation images demonstrate improvement compared with those based on the original images for the classification of masses from normal tissue. This seems logical since, unlike microcalcifications that contain high-frequency components, the masses usually have relatively lower frequency contents. The frequency components of the background normal tissue are also in the low-frequency region, which makes the differentiation much more difficult. As the scale increases (by downward sampling or by increasing the distances in SGLD matrix formulation), the spatial resolution becomes lower while the low-frequency bands becomes narrower. The texture features based on the wavelet coefficients with decreasing low-frequency bandwidth demonstrate statistical difference between masses and normal tissue. At the same time, the effect of the noise with relatively high frequency is eliminated. The subtle differences between the masses and the normal tissue in the low-frequency range are therefore revealed when the difference in the changes of the low-frequency bands between them is utilized through multiresolution analysis. This may explain our finding that classification results with the multiresolution textures are better than those with single distance textures,[25] except for the results obtained with features selected from $\mathbf{F}_{VDS}$. It may be noted that the maximum distance of 16 pixels used in $\mathbf{F}_{VDS}$ is lower than the selected distances of 20 pixels in the single resolution texture analysis.

It is expected that the detail images in the wavelet trans-

form domain contain valuable information about the difference between masses and normal tissue. When radiologists observe some large, suspicious structure, they will usually inspect it in more detail to determine whether it is a mass. However, we found that using the texture features based on the detail images in the wavelet transform domain to formulate the discriminant function did not result in proper classification. It seems that the statistical summary of the textures used here is not effective for the detail images. We will explore the use of other statistical features to extract the information contained in the detail images in future studies.

The reason that the features from the wavelet transform improve the classification results can also be explained as the result of the low-pass filtering operation. In this sense, other low-pass filters can also be used. This provides the possibility of designing optimal filters for the masses so that the classification results can be further improved. An advantage of the wavelet transform over other low-pass filters is that it provides an integral multiresolution framework with great computational efficiency.

The down-sampling process in the wavelet transform effectively reduces the number of pixels in the approximation image at each scale. The reduced size of the approximation images at larger scales will cause more variability in SGLD matrix formulation, thereby affecting the accuracy of the textures estimated at lower image resolution. As the scale increases, the statistical fluctuations of the SGLD matrix based on the smaller and smaller images cannot be neglected due to the random sample errors. In fact, when the approximation images at scale 32 with 8×8 pixels (equivalent to a pixel distance of 32) were used, the texture features did not show any differences between ROIs containing mass and ROIs containing normal tissue due to the small number of pixel pairs for the SGLD matrix formulation. To improve the statistical accuracy of the SGLD matrices, we used the information contained in the decimated coefficients in the wavelet transform and increased the number of discrete distances at which the SGLD matrices, thereby texture features in $\mathbf{F}_{WV}$, could be calculated. Equivalently, this implies that features based on the information in the low-frequency bands with different bandwidths are used for classification. Although this did not significantly improve the classification results for the current data set, the features from $\mathbf{F}_{WV}$ may be statistically superior to those from $\mathbf{F}_{WT}$ because of the reduced uncertainties in the SGLD matrices.

C. Linear discriminant analysis

The classification accuracy is dependent on the feature variables in the linear discriminant function. We observed that when more features were used for the discriminant functions, there was a trend that the training results would improve at the expense of the test results. This is probably because the classifier has too many unknown parameters and is tuned toward the training group when it contains a small number of cases. The resulting discriminant function may not be representative for the general population. Therefore, the generalization capability of the classifier may deteriorate as the number of features used in the linear discriminant function increases. A similar situation arises when other classifiers, e.g., neural network, are used. We also observed that the feature variables selected by the stepwise discriminant analysis was dependent on the case samples in the training set. If we used the training subgroups to select feature variables, the feature variables selected from G_1 were not identical to those from G_2. Therefore, we used the whole data set (G_1 and G_2) to select the feature variables. As the number of case samples increased in the data set for feature selection, the statistical uncertainty of the distributions of the vectors in the feature space was reduced. This is expected to improve the robustness of the selected feature variables.

D. CAD application

One of our goals in the development of CAD methods in mammography is to assist radiologists in detection of suspicious masses on mammograms using computer vision techniques. Before the automated ROI detection method is fully developed, we used manually extracted ROIs to study the feasibility of using texture features for the classification of mass and normal tissue in different types of breast parenchyma. The results of this study demonstrated the potential of using multiresolution texture features for the classification task. The accuracy at an average A_z of 0.86 for the test sets represents a significant improvement over a single resolution approach.[25] Although further improvement in the accuracy is needed before clinical implementation, the algorithm can be incorporated into an automated mass detection program as a step to reduce false-positive ROIs. For example, we can set a decision threshold on the ROC curves (Fig. 6) at a true-positive fraction of 95% and a false-positive fraction (FPF) of 55%, thereby reducing 45% of the FPs while most of the true masses are retained. Alternatively, an accurate classification algorithm, once developed, can also be used independently from an automated detection algorithm. For example, it can be implemented in a CAD workstation and used by radiologists interactively to help differentiate ROIs indicated by the radiologists. The texture information used by the computer analysis may complement the human visual perception. The classification accuracy required, the best operating point on the ROC curve, and the appropriate approach of CAD implementation that can be most useful to radiologists are important topics of investigation in the future.

It is well known that the accuracy of a classifier for FP reduction depends on the specific types of FPs generated in the detection process, which may vary with different automated detection schemes or human observers. The accuracy may also depend to some extent on the properties of the image acquisition system used, such as the amplification mode, dynamic range, or spatial resolution. The coefficients in the linear discriminant function and the selected feature variables are expected to be different when the classifier is used in conjunction with different detection programs. The usefulness of this study lies in the fact that we developed a general approach to the extraction of multiresolution texture features and demonstrated their effectiveness in classification of masses and normal tissue. When this method is applied to a specific task, the classifier must be trained with ROIs representative of the population detected in that process, using the procedures developed in our study as a guide. It is also

important that a much larger number of training samples than that used in this feasibility study is used in order to ensure the generalization capability of the trained classifier.

V. CONCLUSION

In this study, we examined the application of multiresolution texture features in the classification of masses and normal breast parenchyma. With linear discriminant analysis, we demonstrated that multiresolution texture features from the approximation images in the wavelet coefficients at different scales, $\mathbf{F}_{WT}$, provide significant improvement in the classification accuracy over the features from the original images at equivalent distances, $\mathbf{F}_{VDS}$. The features from the combination of wavelet coefficients and variable distances, $\mathbf{F}_{WV}$, can further improve the classification accuracy, although the improvement falls short of statistical significance. The A_z under the ROC curve using 19 features from the $\mathbf{F}_{WV}$ feature space reached an average of 0.89 for training and 0.86 for testing. The approach developed here can be incorporated into a CAD procedure which may assist radiologists in the detection of suspicious lesions on mammograms. While improvement in the classification accuracy is still necessary for clinical applications, our results demonstrate the feasibility of using multiresolution textures for the classification of masses from normal tissue on digital mammograms.

ACKNOWLEDGMENTS

This work is supported by USPHS Grant No. CA 48129, a Faculty Research Award (FRA-334) from the American Cancer Society, and U.S. Army Grant No. DAMD 17-93-J-3007 (through Subgrant No. GU RX 4300-803UM from Georgetown University). The content of this publication does not necessarily reflect the position of the Georgetown University or the government and no official endorsement of any equipment and product of any companies mentioned in the publication should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for providing the LABROC1 and the CLABROC programs.

APPENDIX A: WAVELET TRANSFORM

In the following, we will briefly describe the basic approach of the wavelet transform that is related to this paper. For simplicity, one-dimensional wavelet transform is discussed. Generalization to two-dimensional space is straightforward.

In the wavelet transform, a signal $f(x)$ is decomposed with a family of real orthonormal bases $\psi_{j,n}(x)$ obtained through translation and dilation of a kernel function known as the mother wavelet:

$$\psi_{j,n}(x)=2^{-j/2}\psi(2^{-j}x-n), \tag{A1}$$

where j and n are integers. The wavelet coefficients of the signal $f(x)$ can be obtained through the decomposition

$$c_{j,n}=\int_{-\infty}^{+\infty} f(x)\psi_{j,n}(x)dx. \tag{A2}$$

The signal can be reconstructed from the wavelet coefficients $c_{j,n}$ and the wavelet bases $\psi_{j,n}(x)$ through the synthesis formula

$$f(x)=\sum_{j,n} c_{j,n}\psi_{j,n}(x). \tag{A3}$$

The mother wavelet $\psi(x)$ can be constructed from a scaling function $\phi(x)$, which satisfies the two-scale difference equation[26,27]

$$\phi(x)=\sqrt{2}\sum_k h(k)\phi(2x-k). \tag{A4}$$

The wavelet kernel $\psi(x)$ is related to the scaling function via

$$\psi(x)=\sqrt{2}\sum_k g(k)\phi(2x-k), \tag{A5}$$

where

$$g(k)=(-1)^k h(1-k). \tag{A6}$$

Several conditions have to be met in order for the set of wavelet functions in Eq. (A1) to be unique, orthonormal, and have a certain degree of regularity.[26] Different sets of coefficients satisfying those conditions can be found in the wavelet literature.[27–29]

In the discrete wavelet transform, fast recursive algorithms for wavelet decomposition have been developed. The pyramid wavelet algorithm, which we used for the multiresolution image analysis in this study, decomposes the signal into two parts in the next, larger scale: an approximation signal with the scaling function that has low-pass filter characteristics, and the detail signal with the wavelet function that has the bandpass filter characteristics. In our two-dimensional wavelet transform, we retained the coefficients that corresponded to the scaling function $\phi(x)$ at each scale for texture analysis.

APPENDIX B: TEXTURE FEATURES

An SGLD matrix element $p_{\theta,d}(i,j)$ is the joint probability of the gray level pairs i and j in a given direction θ separated by a distance of d pixels. For each ROI eight features were derived from its SGLD matrix of a given θ and d:

$$\text{energy}=\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p^2(i,j),$$

where n is the number of gray levels of the image;

$$\text{correlation}=\frac{\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(i-\mu_x)(j-\mu_y)p(i,j)}{\sigma_x\sigma_y},$$

where

$$\mu_x=\sum_{i=0}^{n-1} i\sum_{j=0}^{n-1} p(i,j), \quad \sigma_x^2=\sum_{i=0}^{n-1}(i-\mu_x)^2\sum_{j=0}^{n-1} p(i,j),$$

$$\mu_y=\sum_{j=0}^{n-1} j\sum_{i=0}^{n-1} p(i,j), \quad \sigma_y^2=\sum_{j=0}^{n-1}(i-\mu_y)^2\sum_{i=0}^{n-1} p(i,j)$$

are the mean and variance of the marginal distributions $p_x(i)$ and $p_y(j)$, respectively;

$$\text{inertia}=\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(i-j)^2 p(i,j),$$

$$\text{entropy}=-\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}p(i,j)\log_2 p(i,j),$$

$$\text{inverse difference moment}=\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\frac{1}{1+(i-j)^2}p(i,j),$$

$$\text{sum average}=\sum_{k=0}^{2n-2}kp_{x+y}(k),$$

where

$$p_{x+y}(k)=\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}p(i,j),i+j=k,k=0,\ldots,2n-2;$$

$$\text{sum entropy}=-\sum_{k=0}^{2n-2}p_{x+y}(k)\log_2 p_{x+y}(k),$$

$$\text{difference entropy}=-\sum_{k=0}^{n-1}p_{x-y}(k)\log_2 p_{x-y}(k),$$

where

$$p_{x-y}(k)=\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}p(i,j),|i-j|=k,k=0,\ldots,n-1.$$

[a]Author to whom correspondence should be addressed. Department of Radiology, University of Michigan Hospital, 2910 Taubman Center, Ann Arbor, MI 48109-0326; Telephone: (313)-936-4357; Fax: (313)-936-9723.

[1]H. C. Zuckerman, "The role of mammography in the diagnosis of breast cancer," in *Breast Cancer: Diagnosis and Treatment*, edited by I. M. Ariel and J. B. Cleary (McGraw-Hill, New York, 1987), pp. 152–172.
[2]J. N. Wolfe, "Mammography: errors in diagnosis," Radiology **87**, 214–219 (1966).
[3]J. E. Martin, M. Moskowitz, and J. R. Milbrath, "Breast cancer missed by mammography," Am. J. Roentgenol. **132**, 737–739 (1979).
[4]R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," Radiology **184**, 613–617 (1992).
[5]J. A. Harvey, L. L. Fajardo, and C. A. Innis, "Previous mammograms in patients with impalpable breast carcinoma: retrospective vs. blinded interpretation," Am. J. Roentgenol. **161**, 1167–1172 (1993).
[6]H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms: the potential of computer-aided diagnosis," Invest. Radiol. **25**, 1102–1110 (1990).
[7]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).
[8]C. J. Vyborny and M. L. Giger, "Computer vision and artificial intelligence in mammography," Am. J. Roentgenol. **162**, 699–708 (1994).
[9]E. A. Sickles, W. N. Weber, and H. B. Galvin, "Baseline screening mammography: One vs two views per breast," Am. J. Roentgenol. **147**, 1149–1153 (1986).
[10]L. W. Bassett, D. H. Bunnell, R. H. Jahashahi, R. H. Gold, R. D. Arndt, and J. Linsman, "Breast cancer detection: One versus two views," Radiology **165**, 95–97 (1987).
[11]F. F. Yin, M. L. Giger, C. J. Vyborny, K. Doi, and R. A. Schmidt, "Comparison of bilateral-subtraction and single-image processing techniques in the computerized detection of mammographic masses," Invest. Radiol. **28**, 473–481 (1993).
[12]W. P. Kegelmeyer, Jr., "Computer detection of stellate lesions in mammograms," Proc. SPIE **1660**, 446–454 (1992).
[13]L. V. Ackerman and E. E. Gose, "Breast lesion classification by computer and xeroradiography," Cancer **30**, 1025–1035 (1972).
[14]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).
[15]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).
[16]S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," IEEE Trans. Acoust. Speech Signal Process. **37**, 2091–2110 (1989).
[17]A. Laine and S. Song, "Multiscale wavelet representations for mammographic feature analysis," Proc. SPIE **1808**, 610–624 (1992).
[18]R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," IEEE Trans. Inf. Theory **38**, 719–746 (1992).
[19]T. Chang and C. C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," IEEE Trans. Image Process. **2**, 429–441 (1994).
[20]A. Laine, S. Song, J. Fan, W. Huda, and J. Honeyman, "Adaptive multiscale processing for contrast enhancement," Proc. SPIE **1905**, 521–532 (1993).
[21]W. B. Richardson, Jr., "Wavelet packets applied to mammograms," Proc. SPIE **1905**, 504–507 (1993).
[22]W. Qian, L. P. Clarke, K. Maria, H. D. Li, R. Velthuizen, R. A. Clark, and M. L. Silbiger, "Tree-structured filter and wavelet transform for microcalcification segmentation in mammography," Proc. SPIE **1905**, 509–520 (1993).
[23]R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," IEEE Trans. Syst. Man Cybern. **3**, 610–621 (1973).
[24]R. W. Conner, "Towards a set of statistical features which measure visually perceivable qualities of textures," Proceedings of the IEEE Conference On Pattern Recognition and Image Processing, (1979), pp. 382–390.
[25]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).
[26]G. Strang, "Wavelet and dilation equations: A brief introduction," SIAM Rev. **31**, 614–627 (1989).
[27]I. Daubechies, "Orthonormal bases of compactly supported wavelets," Commun. Pure Appl. Math. **41**, 909–996 (1988).
[28]I. Daubechies, "The wavelet transform, time–frequency localization and signal analysis," IEEE Trans. Inf. Theory **36**, 961–1005 (1990).
[29]S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989).
[30]S. G. Mallat and S. Zhong, "Characterization of signals from multiscale edges," IEEE Trans. Pattern Anal. Mach. Intell. **14**, 710–732 (1992).
[31]P. A. Lachenbruch, *Discriminant Analysis* (Hafner, New York, 1975).
[32]M. M. Tatsuoka, *Multivariate Analysis, Techniques for Educational and Psychological Research* (Macmillan, New York, 1988).
[33]J. A. Swets and R. M. Pickett, *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory* (Academic, New York, 1982).
[34]C. E. Metz, "ROC methodology in radiologic imaging," Invest. Radiol. **21**, 720–733 (1986).
[35]C. E. Metz, J. H. Shen, and B. A. Herman, "New methods for estimating a binormal ROC curve from continuously distributed test results," Presented at the 1990 Annual Meeting of the American Statistical Association, Anaheim, CA, August 7, 1990 (unpublished).
[36]C. E. Metz, P. L. Wang, and H. B. Kronman, "A new approach for testing the significance of differences between ROC curves measured from correlated data," *Information Processing in Medical Imaging: Proceedings of the 8th Conference*, edited by F. Deconinck, Boston, Brussels, Aug. 29–Sept. 2, 1983 (The Hague, Nijhoff, 1984), pp. 432–445.
[37]W. R. Zettler, J. Huffman, and D. C. P. Linden, "Application of compactly supported wavelets to image compression," Proc. SPIE **1244**, 150–160 (1990).