

# Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis

Berkman Sahiner, Heang-Ping Chan, Nicholas Petrick, Mark A. Helvie, and Mitchell M. Goodsitt

*The University of Michigan, Department of Radiology, Ann Arbor, Michigan 48109-0030*

(Received 11 October 1996; accepted for publication 27 January 1998)

A new rubber band straightening transform (RBST) is introduced for characterization of mammographic masses as malignant or benign. The RBST transforms a band of pixels surrounding a segmented mass onto the Cartesian plane (the RBST image). The border of a mammographic mass appears approximately as a horizontal line, and possible spiculations resemble vertical lines in the RBST image. In this study, the effectiveness of a set of directional texture features extracted from the RBST images was compared to the effectiveness of the same features extracted from the images before the RBST. A database of 168 mammograms containing biopsy-proven malignant and benign breast masses was digitized at a pixel size of  $100\ \mu\text{m} \times 100\ \mu\text{m}$ . Regions of interest (ROIs) containing the biopsied mass were extracted from each mammogram by an experienced radiologist. A clustering algorithm was employed for automated segmentation of each ROI into a mass object and background tissue. Texture features extracted from spatial gray-level dependence matrices and run-length statistics matrices were evaluated for three different regions and representations: (i) the entire ROI; (ii) a band of pixels surrounding the segmented mass object in the ROI; and (iii) the RBST image. Linear discriminant analysis was used for classification, and receiver operating characteristic (ROC) analysis was used to evaluate the classification accuracy. Using the ROC curves as the performance measure, features extracted from the RBST images were found to be significantly more effective than those extracted from the original images. Features extracted from the RBST images yielded an area ( $A_z$ ) of 0.94 under the ROC curve for classification of mammographic masses as malignant and benign. © 1998 American Association of Physicists in Medicine. [S0094-2405(98)00904-3]

Key words: mammography, computer-aided diagnosis, masses, classification, texture analysis, discriminant analysis, ROC analysis

## I. INTRODUCTION

Mammography is the most effective method for detection of early breast cancer.<sup>1</sup> However, the positive predictive value of mammographic diagnosis is only about 15%–30%.<sup>2–5</sup> Biopsies performed for mammographically suspicious nonpalpable breast masses had positive predictive values of 29%,<sup>6</sup> 29%,<sup>7</sup> and 21%<sup>8</sup> in three studies. As the number of patients who undergo mammography increases, it will be increasingly important to improve the positive predictive value of the procedure in order to reduce costs and patient discomfort. A computerized algorithm that can assist radiologists in classification of mammographic abnormalities may reduce benign biopsies.

Masses are important indicators of malignancy on mammograms. In recent years, considerable effort has been devoted to the development of computerized methods for detection and classification of mammographic masses.<sup>5,9–23</sup> Methods for classification of mammographic masses can be categorized into two groups: one based on features extracted by a radiologist,<sup>5,16–18</sup> and the other based on computer-extracted features.<sup>19–23</sup>

Classification methods based on features extracted by a radiologist are usually designed to include all mammographic signs such as masses and microcalcifications. Al-

though mammographic features are essential components of these methods, age<sup>5</sup> and the personal and family history of the patient<sup>16</sup> are also sometimes used. Getty *et al.*<sup>17</sup> designed a classifier based on discriminant analysis and 12 mammographic features extracted by radiologists, and showed that the classifier can substantially increase the radiologist's diagnostic accuracy. Wu *et al.*<sup>18</sup> designed a neural network classifier based on 14 mammographic features extracted by an experienced radiologist, and showed that its performance in classifying benign and malignant lesions was higher than the average performance of attending and resident radiologists. Recently, Baker *et al.*<sup>16</sup> reported the development of a classifier based on BI-RADS features of the American College of Radiology and the personal and family history of the patient. The specificity of their neural network classifier was shown to be significantly higher than that of the radiologists at high sensitivity levels.<sup>16</sup>

Mass classification methods based on computer-extracted features have the advantages of objectivity and consistency, since they rely on computerized methods for the entire analysis. However, they may also be more difficult to design. These methods usually first extract the lesion shape using interactive or automatic methods, and then extract features from the shape and gray-level characteristics of the lesion,

and the surrounding tissue. Brzakovic *et al.*<sup>19</sup> classified computer-detected suspicious regions into one of three categories, namely, benign tumor, malignant tumor, or nontumor, using their shape and intensity variations. Kilday *et al.*<sup>20</sup> extracted mass shapes using interactive gray-level thresholding, and classified them into cancer, cyst, or fibroadenoma categories using shape features and patient age. Pohlman *et al.*<sup>22</sup> used a region growing algorithm for tumor segmentation, and morphological features extracted from the segmented masses for classification. Huo *et al.*<sup>23</sup> developed a technique to quantify the degree of spiculation of a lesion and classified masses as malignant or benign using the spiculation measures. Their computer-extracted spiculation measure was shown to yield higher classification accuracy than the spiculation rating of an experienced radiologist.<sup>23</sup>

Typical characteristics of malignant masses include high density, spiculated margins, and indistinct, irregular or fuzzy contours. Benign breast masses tend to have sharper, well-circumscribed borders.<sup>24</sup> Automatic characterization of the region surrounding a mass is therefore very important in computer aided diagnosis. An important factor in analyzing the gray-level, gradient, spiculation, and texture characteristics of the area around a mass is their directional dependence. For some of these characteristics, it is difficult to preserve significant directional information from the region surrounding the mass. For example, the gradient of the opacity caused by a mass is radially oriented, and this makes it difficult to extract gradient and texture features from the region surrounding the mass without some preprocessing. Similarly, detection of spiculations is complicated by the fact that the search direction for the spiculation changes with the shape of the mass and the curvature of its margin. To overcome this problem, we have designed a rubber band straightening transform (RBST) which maps a band of pixels surrounding the mass onto the Cartesian plane (a rectangular region). In the transformed image, the border of a mass is expected to appear approximately as a horizontal edge, and spiculations are expected to appear approximately as vertical lines.

The classification algorithm in this paper consisted of four main steps, which were (1) automatic extraction of the mass shape; (2) computation of the RBST image; (3) extraction of texture features; and (4) classification using linear discriminant analysis (LDA). To study the potential advantage of texture feature extraction using the RBST images, the effectiveness of texture features extracted from the RBST images for classification was compared to the effectiveness of the same features extracted from the region surrounding the mass in the original image.

The rest of the paper is organized as follows. In the next section, we describe our image database, and the four steps of the classification algorithm specified above. In Sec. III, we present the classification results using texture features extracted from different image representations (with or without the RBST). Section IV contains a discussion of these results. Finally, Sec. V concludes the investigation and provides a scope for further research.

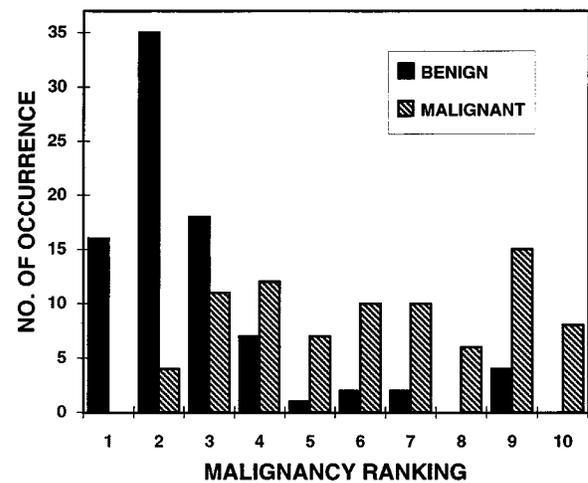


FIG. 1. The distribution of the malignancy ranking of the masses in our dataset, by an experienced radiologist. 1: Very likely benign, 10: Very likely malignant.

## II. MATERIALS AND METHODS

### A. Data set

The mammograms used in this study were randomly selected from the files of patients who had undergone biopsy in the Department of Radiology at the University of Michigan. The criterion for inclusion of a mammogram in the data set was that the mammogram contained a biopsy-proven mass. To avoid the effect of repetitive grid lines on the image texture, mammograms that contained grid lines caused by the stationary grid of some older mammographic units were excluded. The mammograms were digitized with a LUMISYS DIS-1000 laser scanner at a pixel resolution of  $100 \mu\text{m} \times 100 \mu\text{m}$ , and 4096 gray levels. The digitizer was calibrated so that gray-level values were linearly proportional to the optical density (OD) within the range of 0.1–2.8 o.d. units, with a slope of 0.001 OD/pixel value. Outside this range, the slope of the calibration curve decreased gradually. The o.d. range of the digitizer was 0 to 3.5.

The data set in this study included 168 mammograms from 72 patients. Of the 168 mammograms, 83 contained malignant masses, and 85 contained benign masses. Six of the benign masses and 45 of the malignant masses were spiculated, as determined visually by a radiologist experienced in mammographic interpretation. Regions of interest (ROIs) containing the biopsied masses were extracted by the same radiologist from each mammogram. The size of each ROI was  $256 \times 256$  pixels. Our data set contained a range of obvious to subtle masses. The probability of malignancy of each mass, based on its mammographic appearance, was ranked by the radiologist on a scale of 1 to 10, where a ranking of 1 corresponded to the masses with the most benign mammographic appearance. The distribution of the malignancy ranking of the masses is shown in Fig. 1.

### B. Mass shape extraction

We used a pixel-by-pixel clustering algorithm followed by object selection for segmentation of the ROI into a mass

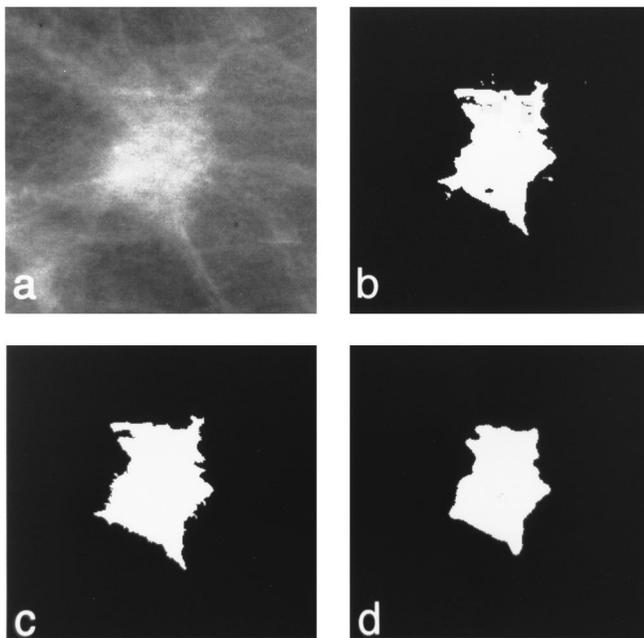


FIG. 2. (a) The original ROI; (b) the result of the initial segmentation; (c) segmented and grown mass object; and (d) smoothed mass object.

object and background tissue. Our segmentation algorithm is described in detail elsewhere.<sup>25,26</sup> Briefly, we obtained several filtered images from the original ROI pixel values, and used the original and filtered pixel values as the elements of a feature vector in the clustering algorithm. The inclusion of spatially filtered images incorporated neighborhood information in the classification of a given pixel.

Figure 2(a) shows an ROI with a spiculated mass. The segmented objects which resulted from the clustering algorithm are shown in Fig. 2(b). After clustering, the largest connected object among all detected objects was selected, filled, and grown in a small region outside its boundary. Details of the region growing algorithm can be found in our previous publications.<sup>25,26</sup> Figure 2(c) shows the result of object selection, filling, and object growing applied to Fig. 2(b). Finally, the borders of the grown object were smoothed by using a morphological opening operation.<sup>27</sup> The opening operation for a binary image consists of the successive application of erosion and dilation operations. In this study,  $11 \times 11$  and  $7 \times 7$  pixel circular masks were used for the erosion and dilation operations, respectively. The final smoothed mass object for the ROI in Fig. 2(a) is shown in Fig. 2(d).

In this study, we chose the parameters in the clustering and region growing algorithms such that the mass object was segmented to be slightly smaller than that which could be visually determined on the mammogram. Thus a thin border region along the mass margin was included in the RBST image. Important texture and gradient information at the mass margin was therefore included in the analysis of the region surrounding the mass.

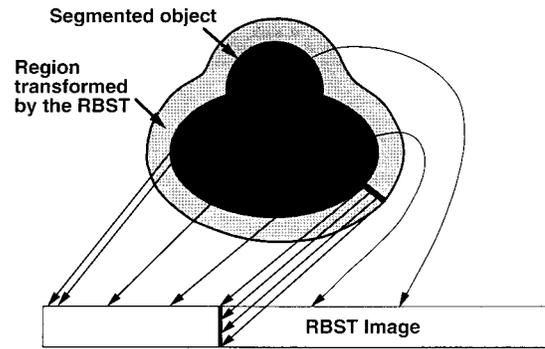


FIG. 3. An illustration of the RBST. The pixels along the object boundary are mapped to the first row of the RBST image. Pixels along a normal to the object boundary are mapped to a column of the RBST image.

### C. The RBST image

The RBST maps the pixel values in a band of pixels surrounding the mass onto the Cartesian plane. The mapping designed in this study had the following properties: (i) traversing a closed path at a constant distance from the detected object border in the original image approximately corresponded to moving along a row of the RBST image; and (ii) traveling in a direction normal to the boundary of the detected object in the original image approximately corresponded to moving along a column of the RBST image (Fig. 3). These properties make the RBST well-suited for extracting texture features that radiate from the borders of the mass.

The RBST consists of three main components, edge enumeration, computation of normals, and computation of RBST pixel values. These steps are explained in detail below.

#### 1. Edge enumeration and computation of normals

The border pixels of an object form a closed chain, i.e., starting at an arbitrary pixel, it is possible to move along the chain and return to the starting pixel. Conceptually, the edge enumeration algorithm removes pixels, one at a time, from the edge contour of the object, and places the  $x$  and  $y$  coordinates of each border pixel on an edge enumeration list. Thus each pixel in the chain is assigned a number, which corresponds to the placement of the pixel in the list.

The algorithm starts by choosing a relatively smooth location on the edge contour, as illustrated in Fig. 4. One pixel (pixel number 1 in Fig. 4) is removed from the edge chain so that the chain is broken. Starting at this break point, pixels are sequentially removed from the chain, and the  $x$  and  $y$  coordinates of a removed pixel are placed on an enumeration list. Edge enumeration terminates when one returns to the starting pixel after every pixel has been removed from the chain. Since pixel removal is sequential, consecutive pixels in the enumeration list have to be 8-connected neighbors<sup>28</sup> on the edge contour of the object. The algorithm tries to keep the chain in one piece as long as it is possible. Thus referring to Fig. 4, pixel number 12 is followed in the list by pixel number 13, and not pixel number 24. However, when the object shape is complicated, for example, if the object consists of two subobjects joined together with a single bridge

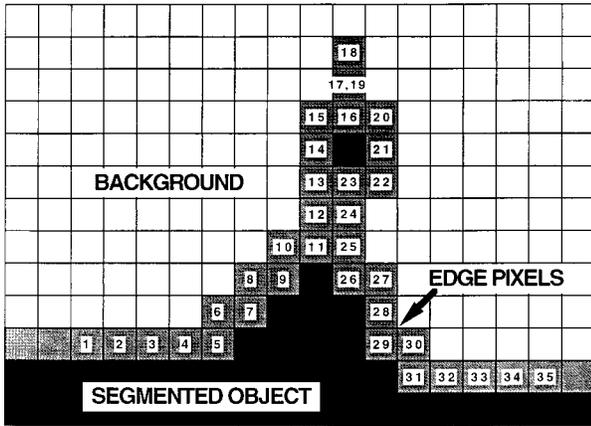


FIG. 4. The edge enumeration algorithm.

pixel, it may not be possible to keep the chain in one piece. When the chain has to be broken into two, some of the pixels in the chain have to be repeated in the list so that one can return to the starting pixel after removing all the pixels in the chain. The algorithm will then choose a path such that only a small number of pixels in the list are repeated. Thus referring to Fig. 4, pixel number 17 is repeated as pixel number 19 in the list. The number of pixels in the edge enumeration list is denoted as  $N_e$ . Since some of the pixels may be repeated in the list,  $N_e$  is larger than or equal to the number of edge pixels in the object.

The computation of the normal direction to the object is based on the object shape and the result of the edge enumeration. For a given pixel  $i$  in the enumeration list, pixels  $i + K$  and  $i - K$ , occurring  $K$  places before and after pixel  $i$  are located in the list. The normal direction to the object at edge pixel  $i$  is determined as the normal to the line joining edge pixels  $i + K$  and  $i - K$ . This procedure is illustrated graphically in Fig. 5. If  $K=1$  as in Fig. 5, only a small neighborhood of a pixel is considered for normal computation, and the computed normals may be noisy. In addition,

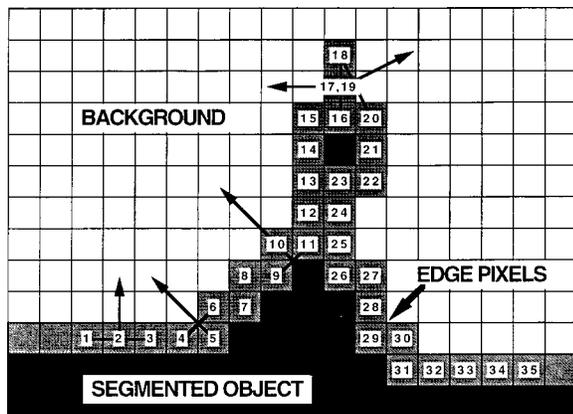


FIG. 5. Computation of normals. For each pixel  $i$ , the normal direction  $L(i)$  is perpendicular to the line joining pixels  $i + K$ ,  $i - K$ . For the purpose of illustration,  $K$  is set to 1 in this figure.  $K=12$  was used in the actual calculation.

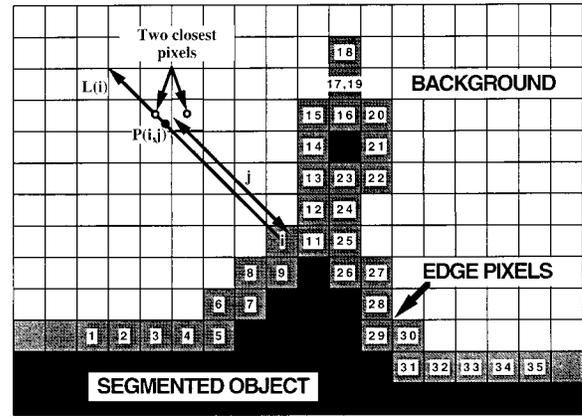


FIG. 6. Computation of RBST pixel values.  $p(i,j)$  has a distance  $j$  from the pixel  $i$  along the normal line  $L(i)$ . The  $(i,j)$ th pixel value in the RBST image is a distance-weighted average of the two closest pixels to  $p(i,j)$ .

$K=1$  confines the number of normal directions to only a small number, since the line joining two neighboring pixels of a given edge pixel can occur in only a small number of directions. On the other extreme, a large value of  $K$  may introduce too much smoothing, and some of the fine direction changes in the mass contour may be missed. In this study, it was found experimentally that  $K=12$  resulted in a satisfactory normal estimation for most of the mass shapes, and this value was used in the computation of all the RBST images.

### 2. Computation of RBST pixel values

The basic idea behind the computation of RBST pixel values is as follows. Let  $L(i)$  denote the normal to the object at edge pixel  $i$ , and let  $p(i,j)$  denote the point on the line  $L(i)$  which has a distance  $j$  from edge pixel  $i$  (see Fig. 6.). The value of the pixel in row  $j$ , column  $i$  of the RBST image is defined as the distance-weighted average of the two closest pixels to  $p(i,j)$  in the original image. With this definition, the number of pixels in the enumeration list  $N_e$  is equal to the number of columns in the RBST image. The width of the region desired to be transformed determines the number of rows in the RBST image. This definition of the RBST will be referred to as the short RBST.

One difficulty with the short RBST is that as the distance  $j$  in the normal direction increases, the length of the closed path surrounding the object at a constant distance  $j$  from the object boundary also increases. This may result in undersampling and possibly a loss of information in the RBST image. For example, each of the object border pixels in the original ROI are mapped to the first row of the RBST image. Thus at the first row, transformation from the original image to the RBST image does not result in any information loss. However, when  $j$  is large, some pixels in the original image do not contribute to any of the pixels in the RBST image, and the information carried by these pixels will be lost. To reduce the information loss, we increased the number of columns of the RBST image from  $N_e$  (defined in the previous paragraph) to  $2N_e$ . Normals were drawn from each edge

pixel of the object, as well as the midpoints between every two pixels, and the computation of the RBST image was performed as described in the previous paragraph. This definition of the RBST will be referred to as the regular RBST. In this study, we implemented the regular RBST as our main transform. The classification results using the regular RBST are presented in Sec. III.

Depending on the size and shape of the mass, the regular RBST image may contain more pixels than the band of pixels surrounding the mass in the rows adjacent to the segmented mass border. The RBST pixels are computed from the original pixel values using distance-weighted as described above. Therefore, these extra pixels can be considered as the result of an interpolation process. To test whether these extra pixels resulting from interpolation contribute to the performance of the RBST, two options are available. The first option is to interpolate the  $256 \times 256$  pixel ROI to a larger size by cubic spline interpolation, and to compare the classification accuracy of the texture features extracted from band of pixels in the interpolated image to that of features extracted from the regular RBST image. The second option is to implement the short RBST. The short RBST contains half as many pixels as the regular RBST, and always has fewer pixels than the band of pixels surrounding the mass for convex mass shapes. The classification accuracy using the short RBST can then be compared to that using the original ROI. In this work, we have implemented this second option for comparison, which will simplify RBST implementation if it is found to be as effective as the regular RBST. The results of the comparison are presented in Sec. IV.

Other implementation issues are as follows. A 40-pixel-wide region surrounding the mass object, which corresponds to a 4-mm-wide band, was used to determine the RBST image. The size of the regular RBST image was thus  $2N_e$  columns by 40 rows. As discussed in the previous subsection, the distance  $K$  used in the computation of normals was 12. For some large masses, some pixels in a 40-pixel-wide band around the mass might fall outside the boundary of the  $256 \times 256$  pixel ROI. In this study, if  $p(i, j)$  fell outside the ROI, the  $(i, j)$ th pixel value of the RBST image was flagged as an “invalid” pixel. This in effect reduced the size of the region for extraction of the texture features, as described below. However, since the RBST image of a large mass had a large value of  $N_e$ , the reduction in region size did not have a strong effect on the statistical properties of the texture features. An example of an original ROI, segmented mass object, and the RBST image is shown in Fig. 7.

#### D. Texture features

The texture features used in this study were calculated from spatial gray-level dependence (SGLD) matrices,<sup>13,14,29</sup> and run-length statistics (RLS) matrices.<sup>30</sup> For comparison purposes, these matrices were computed for three image representations: (i) the entire  $256 \times 256$  ROI, denoted as  $R1$ ; (ii) a 40-pixel-wide band surrounding the extracted mass boundary, denoted as  $R2$ ; and (iii) the RBST image obtained by applying the RBST to the 40-pixel-wide band, denoted as

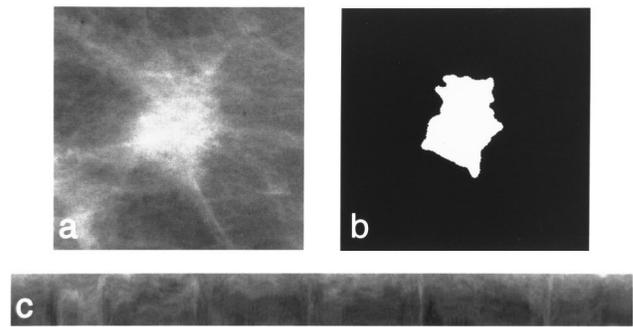


Fig. 7. (a) Original image; (b) segmented mass object; and (c) RBST image.

$R3$ . SGLD matrices were constructed from the gray-level images for  $R1$ ,  $R2$ , and  $R3$ , and RLS matrices were constructed from the vertical and horizontal gradient images derived from the three image representations, as described below.

#### 1. SGLD texture features

An SGLD matrix can be considered to be a two-dimensional histogram. The element in row  $r$ , column  $c$  of an SGLD matrix is the joint probability that gray levels  $r$  and  $c$  occur in a direction  $\theta$  and at a pixel pair distance of  $d$  in the image. The distribution of the SGLD matrix elements reflects the average spatial relationship of pairs of gray-level tones with respect to the distance  $d$  and direction  $\theta$  used in SGLD matrix construction. For example, if the image texture is coarse, and the distance  $d$  is small in comparison to the texture element, then pairs with similar gray levels are expected to occur relatively frequently, and pairs with dissimilar gray levels are expected to occur relatively infrequently. Thus the SGLD matrix will be mainly concentrated along the main diagonal. If, in addition, the image is relatively bright (indicated by high pixel values), then the SGLD matrix will be concentrated around the lower main diagonal. SGLD texture features, described in the next paragraph, extract this information from the SGLD matrix.

Based on our previous studies,<sup>13</sup> a bit depth of eight bits was used in the SGLD matrix construction, i.e., the least significant four bits of the 12-bit pixel values were discarded. Eight texture measures, namely, correlation, energy, difference entropy, inverse difference moment, entropy, sum average, sum entropy, and inertia were extracted from each SGLD matrix at eight different pixel pair distances, ( $d=1, 2, 3, 4, 6, 8, 12,$  and  $16$ ) and in four directions ( $\theta=0^\circ, 45^\circ, 90^\circ,$  and  $135^\circ$ ). Therefore, a total of 256 SGLD features were calculated for each image representation. The formulation of these texture measures has been described in the literature.<sup>13,14,29</sup> These features contain information about image characteristics such as homogeneity, contrast, and the complexity of the image.<sup>29</sup> For example, the energy feature, which is the sum-of-squares of the SGLD matrix elements, is smallest when all the elements of the SGLD matrix are equal, i.e., when all the pixel pairs occur with equal probability. This would indicate that the image does not have a

lot of structure. As another example, the inertia feature, which is the moment of inertia of the SGLD matrix around its main diagonal, measures the spread of the matrix elements around the main diagonal. A high value of this feature means that the spread is high, which indicates that the size of the image texture elements are comparable to, or smaller than the pixel pair distance  $d$ . Although such examples provide an idea about the meaning of these features, it is difficult to establish a one-to-one correspondence between the qualitative image characteristics and the extracted features.<sup>29</sup>

In this study, special care was taken in the construction of the SGLD matrix, since some of the image representations contained invalid pixel values as specified in the previous subsection. When the SGLD matrices were constructed, pixel pairs involving invalid pixel values were not accumulated in the SGLD matrix.

## 2. RLS texture features

RLS texture features were extracted from vertical and horizontal gradient magnitude images, which were obtained by filtering the image representation of interest by the horizontally or vertically oriented Sobel filters, and computing the absolute value of the filtered image.

A gray-level run is a set of consecutive, colinear pixels in a given direction which have the same gray-level value. A run length is the number of pixels in a run.<sup>30</sup> The RLS matrix describes the run-length statistics for each gray-level value in the image. The element in row  $r$ , column  $c$  of an RLS matrix is the number of times that the gray level  $r$  in the image possesses a run length of  $c$  in a given direction.

Analogous to SGLD matrix computation, invalid pixel values were excluded from the RLS matrix computation. If a large bit depth is used in RLS matrix computation, the resulting run lengths are very short for all of the images, and the discriminatory power may not be high. Conversely, if the bit depth is too small, then run lengths become predominantly long. In this study, it was found experimentally that a bit depth of 5 bits in RLS matrix computation resulted in a good compromise.

Five texture measures, namely, short runs emphasis, long runs emphasis, gray-level nonuniformity, run-length nonuniformity, and run percentage were extracted from the vertical and horizontal gradient images in two directions,  $\theta=0^\circ$ , and  $\theta=90^\circ$ . Therefore, a total of 20 RLS features were calculated for each image representation. The definition of the RLS texture measures used in this study can be found in the literature.<sup>30</sup> It is possible to crudely describe the dependence of these features on the image characteristics, e.g., the run percentage feature value is small for images with long linear structures, and the gray-level nonuniformity feature value is small for images where runs are equally distributed throughout the gray levels. However, it is again difficult to establish a one-to-one correspondence between the qualitative image characteristics and the extracted features.

## E. Classification

Linear discriminant analysis<sup>31,32</sup> was used to classify malignant and benign masses based on the extracted texture features. A stepwise feature selection procedure with the minimization of Wilks' lambda (the ratio of within-group sum of squares to the total sum of squares) was used as the optimization criterion to select effective predictor variables. Stepwise feature selection is an iterative procedure, where one feature is entered into or removed from the selected feature pool at each step by analyzing its effect on the selection criterion. In the feature entry phase of a step, the available features are entered into the selected feature pool one at a time. The significance of the change in Wilks' lambda, as measured by  $F$ -statistics, when a feature is entered into the selected feature pool is compared to a threshold  $F_{in}$ . The feature with the highest significance is entered into the selected feature pool only if the significance is higher than  $F_{in}$ . Likewise, in the feature removal phase, features that were already selected are removed from the selected feature pool one at a time, and the significance of change in Wilks' lambda is compared to a threshold  $F_{out}$ . The feature with the least significance is removed from the selected feature pool only if the significance is lower than  $F_{out}$ . Since the optimal values of the  $F_{in}$  and  $F_{out}$  parameters are not known *a priori*, we varied both parameters, and tried to obtain the feature combinations that yielded the highest classification accuracy for each of the three image representations. Details about the application of stepwise linear discriminant analysis to CAD can be found in our previous publications.<sup>13,14,26</sup>

A leave-one-case-out method was used to train and test the classifier. In this method, all films belonging to one patient were left out from the classifier design group at the same time. A linear discriminant function was formed using the design group, and test discriminant scores were computed for the left-out films using the linear discriminant function. This process cycled through the data set until every patient's films were used as test films once. The test discriminant scores of all films were analyzed using receiver operating characteristic (ROC) methodology<sup>33</sup> to evaluate the classifier performance. The discriminant scores of the malignant and benign masses were used as the decision variable in the LABROC1 program,<sup>34</sup> which provided the ROC curve based on maximum likelihood estimation. The classification accuracy was evaluated as the area  $A_z$  under the ROC curve. The CLABROC program<sup>35</sup> was used to test the statistical significance of the difference between pairs of ROC curves obtained using texture features extracted from  $R1$ ,  $R2$ , and  $R3$  under corresponding conditions.

## F. Computational considerations

Segmentation, image transformation, feature extraction, and classifier design steps of our algorithm were executed on an AlphaStation 500 (400-MHz Alpha chip), and the feature selection step was performed on a PC compatible computer with a 90-MHz Pentium processor. The classification for the entire data set of 168 images took less than an hour, which meant that the classifier design and classification for each

TABLE I. Classifier performance with SGLD texture features, extracted from (a) *R1* (the original ROI), (b) *R2* (the 40-pixel-wide region surrounding the mass), and (c) *R3* (the RBST image).  $F_{in}$  and  $F_{out}$  values are thresholds used in the stepwise feature selection method for entering and removing features from the selected feature pool. In general, lower thresholds result in a larger number of selected features.

(a)				
$F_{in}$	$F_{out}$	Num. of features	Training $A_z$	Test $A_z$
1.2	1.0	13	$0.85 \pm 0.03$	$0.77 \pm 0.04$
1.1	1.3	17	$0.87 \pm 0.03$	$0.79 \pm 0.03$
1.1	1.2	20	$0.88 \pm 0.03$	$0.78 \pm 0.04$
0.8	0.6	25	$0.91 \pm 0.02$	$0.81 \pm 0.03^*$
0.6	0.4	26	$0.91 \pm 0.02$	$0.79 \pm 0.03$
(b)				
$F_{in}$	$F_{out}$	Num. of features	Training $A_z$	Test $A_z$
0.6	0.8	5	$0.78 \pm 0.03$	$0.74 \pm 0.04$
0.73	0.73	17	$0.85 \pm 0.03$	$0.79 \pm 0.03$
0.7	0.7	21	$0.90 \pm 0.02$	$0.83 \pm 0.03$
0.6	0.4	32	$0.96 \pm 0.01$	$0.87 \pm 0.03^*$
0.4	0.2	34	$0.96 \pm 0.01$	$0.86 \pm 0.03$
(c)				
$F_{in}$	$F_{out}$	Num. of features	Training $A_z$	Test $A_z$
2.4	2.2	9	$0.92 \pm 0.02$	$0.89 \pm 0.03$
2.2	2.0	12	$0.94 \pm 0.02$	$0.91 \pm 0.02^*$
0.6	0.4	18	$0.95 \pm 0.02$	$0.90 \pm 0.02$
0.4	0.2	23	$0.95 \pm 0.02$	$0.89 \pm 0.02$

mass was performed in less than 30 s. If a trained classifier is implemented, the feature selection and classifier design steps will not be needed for classifying an unknown case, and the computation time will be shorter.

### III. RESULTS

In this section, we present classification results with texture features derived from the *R1*, *R2*, and *R3* image representations. Since the optimal number of features is not known *a priori*, we varied the  $F_{in}$  and  $F_{out}$  parameters in the stepwise linear discriminant analysis and tried to obtain a range in the number of selected features for each image representation. The  $F_{in}$  and  $F_{out}$  values, as well as the number of features are tabulated for different conditions in the following subsections. After feature selection and classifier design were completed, each designed classifier was applied to its design samples, and a training  $A_z$  value was obtained. Since our database contained images from 72 different patients, 72 classifiers were trained for each feature combination in a leave-one-case-out paradigm. The training  $A_z$  values and their standard deviations in the following tables represent the averages of these quantities from the output of the LABROC1 program over the 72 classifiers. After training and testing were completed on all of the films for a feature combination, the test  $A_z$  and its standard deviation were estimated by the LABROC1 program using the test scores.

#### A. SGLD feature space

Tables I(a)–I(c) show the training and test classification

TABLE II. Classifier performance with RLS texture features, extracted from (a) *R1* (the original ROI), (b) *R2* (the 40-pixel-wide region surrounding the mass), and (c) *R3* (the RBST image).  $F_{in}$  and  $F_{out}$  values are thresholds used in the stepwise feature selection method for entering and removing features from the selected feature pool. In general, lower thresholds result in a larger number of selected features.

(a)				
$F_{in}$	$F_{out}$	Num. of features	Training $A_z$	Test $A_z$
1.6	1.4	3	$0.74 \pm 0.04$	$0.70 \pm 0.04$
1.4	1.2	4	$0.74 \pm 0.04$	$0.70 \pm 0.04$
1.2	1.0	5	$0.75 \pm 0.03$	$0.70 \pm 0.04^*$
0.2	0.1	9	$0.75 \pm 0.03$	$0.67 \pm 0.04$
(b)				
$F_{in}$	$F_{out}$	Num. of features	Training $A_z$	Test $A_z$
1.2	1.0	2	$0.73 \pm 0.04$	$0.71 \pm 0.04$
0.8	0.6	5	$0.76 \pm 0.04$	$0.72 \pm 0.04^*$
0.6	0.4	6	$0.77 \pm 0.04$	$0.72 \pm 0.04$
(c)				
$F_{in}$	$F_{out}$	Num. of features	Training $A_z$	Test $A_z$
5.2	5.0	5	$0.86 \pm 0.03$	$0.84 \pm 0.03^*$
3.8	2.7	6	$0.87 \pm 0.03$	$0.83 \pm 0.03$
1.2	1.0	7	$0.87 \pm 0.03$	$0.84 \pm 0.03$
1.0	0.8	9	$0.88 \pm 0.03$	$0.83 \pm 0.03$

accuracies using the SGLD features derived from the *R1*, *R2*, and *R3* image representations, respectively. The highest test classification result for each representation is marked with an asterisk. It can be observed that the range of selected features for each representation was large enough so that the maximum occurred within the range, and not at the highest or lowest number of selected features. The test classification results in Table I(a), as well as those in Table I(c) were within one standard deviation of each other. The results in Table I(b) had a larger variation, due to the wider range in the number of selected features. The difference between the best classification results using *R1* and *R3* was statistically significant ( $p < 0.03$ ). The difference between the best classification results using *R2* and *R3* did not achieve statistical significance. The texture features that were selected most frequently in the SGLD feature space were difference entropy and inverse difference moment. Both of these features measure the spread of the SGLD matrix along lines parallel to the main diagonal. Therefore, they are measures of the local nonhomogeneity of the image.

#### B. RLS feature space

Tables II(a)–II(c) show the training and test classification accuracies using the RLS features derived from the *R1*, *R2*, and *R3* image representations, respectively. The highest test classification results are marked with an asterisk. Since we had only 20 RLS texture features, the variation in the number of features in each table was smaller compared to that for the SGLD texture features. The test classification results within each table were again within one standard deviation of each other. The difference between the best classification results using *R1* and *R3*, as well as *R2* and *R3* were statistically

TABLE III. Classifier performance with combined texture features, extracted from (a)  $R1$  (the original ROI), (b)  $R2$  (the 40-pixel-wide region surrounding the mass), and (c)  $R3$  (the RBST image).  $F_{in}$  and  $F_{out}$  values are thresholds used in the stepwise feature selection method for entering and removing features from the selected feature pool. In general, lower thresholds result in a larger number of selected features.

(a)				
$F_{in}$	$F_{out}$	Num. of features	Training $A_z$	Test $A_z$
1.8	1.6	8	$0.82 \pm 0.03$	$0.77 \pm 0.04$
1.35	1.2	14	$0.87 \pm 0.03$	$0.80 \pm 0.03$
1.3	1.2	16	$0.88 \pm 0.03$	$0.81 \pm 0.03^*$
1.2	1.0	19	$0.90 \pm 0.02$	$0.79 \pm 0.03$
1.0	0.8	20	$0.90 \pm 0.02$	$0.78 \pm 0.03$
0.8	0.6	30	$0.92 \pm 0.02$	$0.80 \pm 0.03$
(b)				
$F_{in}$	$F_{out}$	Num. of features	Training $A_z$	Test $A_z$
1.8	1.6	15	$0.91 \pm 0.02$	$0.86 \pm 0.03$
1.4	1.2	19	$0.93 \pm 0.02$	$0.86 \pm 0.03$
1.2	1.0	20	$0.93 \pm 0.02$	$0.87 \pm 0.03^*$
1.1	1.1	21	$0.93 \pm 0.02$	$0.86 \pm 0.03$
1.0	0.8	25	$0.94 \pm 0.02$	$0.86 \pm 0.03$
0.8	0.8	27	$0.94 \pm 0.02$	$0.85 \pm 0.03$
(c)				
$F_{in}$	$F_{out}$	Num. of features	Training $A_z$	Test $A_z$
3.0	2.8	11	$0.92 \pm 0.02$	$0.89 \pm 0.02$
2.6	2.4	14	$0.96 \pm 0.01$	$0.94 \pm 0.02$
2.2	2.0	18	$0.97 \pm 0.01$	$0.94 \pm 0.02$
1.6	1.4	20	$0.98 \pm 0.01$	$0.94 \pm 0.02^*$
1.0	1.0	22	$0.97 \pm 0.01$	$0.93 \pm 0.02$

significant ( $p < 0.01$  for both). Long runs emphasis and short runs emphasis were the two features that were selected most frequently in the RLS feature space. These features emphasize long and short runs in the image, and therefore indicate the existence of long or short linear structures in the image, respectively.

### C. Combined SGLD and RLS feature space

Tables III(a)–III(c) show the training and test classification accuracies using both the SGLD features and the RLS features derived from the  $R1$ ,  $R2$ , and  $R3$  image representations, respectively. In analogy to SGLD feature selection, the range of selected features in this subsection was large enough so that the maximum occurred within the range. Almost all of the test classification results within each table were within one standard deviation of each other. The ROC curves for the classifiers with the highest test accuracy, marked by an asterisk in the tables, are plotted in Fig. 8. The difference between the best classification results using  $R1$  and  $R3$ , as well as  $R2$  and  $R3$  were again statistically significant ( $p < 0.01$  for both). The distribution of the test discriminant scores obtained by using features extracted from the RBST images is shown in Fig. 9. By choosing an appropriate decision threshold on the test discriminant scores, more than 30% of the benign masses could correctly be identified without missing any malignant masses. Difference en-

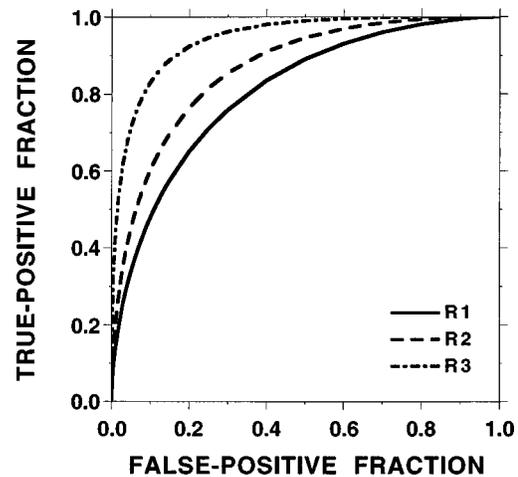


FIG. 8. ROC curves for  $R1$  (the original ROI),  $R2$  (the 40-pixel-wide region surrounding the mass), and  $R3$  (the RBST image). Classification was performed in the combined SGLD and RLS feature space.

trophy, inverse difference moment, and long runs emphasis were the three features that were selected most frequently in the combined feature space.

### IV. DISCUSSION

We have designed and implemented a new rubber band straightening transform, and used this transformation for classifying malignant and benign breast masses. Our results showed that both SGLD features and RLS features, as well as the combined feature set extracted from the RBST images ( $R3$ ) were significantly more effective than similar features extracted from the entire  $256 \times 256$  ROI containing the mass ( $R1$ ). The RBST image was obtained by transforming a 40-pixel (4 mm) wide band surrounding the segmented mass. For this reason, we compared the classification effectiveness of texture features extracted from a 40-pixel-wide band surrounding the segmented mass ( $R2$ ) with those from the RBST image ( $R3$ ). Our results showed that RLS features extracted from  $R3$  were significantly more effective than RLS features extracted from  $R2$ . The classification accuracy using SGLD features extracted from  $R3$  was also higher

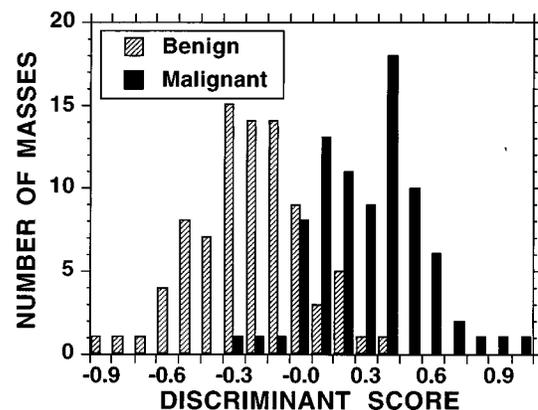


FIG. 9. The distribution of the test discriminant scores obtained by using combined SGLD and RLS features extracted from  $R3$  (the RBST images).

than  $R2$ , although the difference did not achieve statistical significance. In the combined feature space, we again observed significantly higher classification accuracy with the use of the RBST images.

It is expected that the texture of the region surrounding a mass has a radial dependence, because possible speculations and the gradient of the opacity caused by the mass are approximately radially oriented. However, most texture extraction methods are designed for texture orientations in a uniform direction (horizontal, vertical, or at a certain angle between these two directions). By transforming the region surrounding a mass into an RBST image, we have attempted to create a transformed image in which texture orientations become more suitable for feature extraction using existing techniques. The results of this study indicate that our approach is promising.

The width of the region transformed by the RBST was selected as 40 pixels (4 mm) in this paper. In another publication on classification of masses,<sup>36</sup> the same size was used inside and outside the mass for feature extraction. If the width of this band is too small, then the RBST image may exclude some of the border regions with useful texture in the original image. If the width is too large, then the statistical feature variations of the structures far away from the mass, which carry little or no information on its probability of malignancy, may be included and degrade the classifier performance. We did not perform a systematic study of the effect of the size of this region on the classification accuracy. However, to test whether this size was a critical parameter, we obtained RBST images for 30- and 50-pixel-wide bands, and extracted the same set of features as discussed in Sec. II from these images. With 30- and 50-pixel-wide bands, the test classification accuracy  $A_z$  using the combined feature space was 0.93 and 0.92, respectively. The difference between these results and the best result in Table III(c) ( $A_z=0.94$ ) was not statistically significant. We therefore surmise that the classification accuracy will not be very sensitive to this size. It is reasonable to expect that the size of the region surrounding the mass that contains useful information about its malignancy will change with the size of the mass. Therefore, one may improve the classification results obtained in this paper by adaptively changing the size of the region transformed by the RBST depending on the size of the mass. This will be investigated in the future.

The length of the RBST image in this paper was  $2N_e$  pixels, where  $N_e$  is the number of edge pixels of the segmented mass. Depending on the size and shape of the mass, the RBST image thus defined may contain more pixels than the 40-pixel-wide band area surrounding the mass. To test whether these extra pixels contribute to the performance of the RBST, we implemented a variation of the RBST termed the short RBST, which produces an RBST image having a length of  $N_e$  pixels. For a convex mass shape, the short RBST image will always have fewer pixels than the band of pixels surrounding the mass.

After the computation of the short RBST images, feature extraction, selection, and classification were performed in the same way as the regular images, as discussed in Secs. II and

III. The test  $A_z$  scores using the SGLD, RLS, and combined feature spaces were 0.91, 0.81, and 0.93, respectively. These results are equal to, or slightly worse than the best test results in Tables I(c), II(c), and III(c) marked with an asterisk. The difference between the  $A_z$  values obtained using the corresponding feature spaces was not statistically significant. The statistical differences between the classification results obtained using the short RBST and the  $R1$  or  $R2$  image representations were similar to the differences between the regular RBST and the  $R1$  or  $R2$  image representations. More precisely, the classification results obtained using the short RBST were significantly better than those obtained using both  $R1$  and  $R2$  representations in the RLS and combined feature spaces ( $p<0.05$ ). In the SGLD feature space, the difference between the classification results using the short RBST and the  $R1$  image representation was statistically significant ( $p<0.05$ ), but the difference between the short RBST and the  $R2$  image representation did not achieve statistical significance. These results show that the extra pixels resulting from the interpolation in the computation of the regular RBST do not provide an advantage to the RBST over the other image representations. This is consistent with the expectation that interpolation generally does not increase image information.

The test  $A_z$  values obtained from a given representation in a given feature space were within one standard deviation of each other. This meant that the optimal values of  $F_{in}$  and  $F_{out}$ , and therefore the number of selected features, were not critical for designing the classifiers. However, the feature selection process itself is a critical component in classification, as shown in our previous study.<sup>26</sup> In many of the tables, one can observe the so-called peaking phenomenon,<sup>37</sup> which means that when a moderate number of design samples is available for classifier design, the test accuracy first increases, but later starts to decrease as the number of features is increased.

As discussed in Sec. II, the probability of malignancy of each mass, based on the mammographic appearance, was ranked by a radiologist experienced in mammography (Fig. 1). Based on this ranking, an ROC curve was estimated using the LABROC1 program, and plotted in Fig. 10. The figure also plots the ROC curve obtained by using the combined texture features extracted from the RBST images. The  $A_z$  value obtained by the malignancy ranking of the radiologist was  $0.89\pm 0.03$ . The difference between the ROC curves using the computerized classification algorithm ( $A_z=0.94\pm 0.02$ ) and the malignancy rating of the radiologist was statistically significant (two-tailed  $p$  level=0.03). This result also highlights the promise of our approach.

In this study, the ranking by the radiologist, as well as the computer scores, were based only on the appearance of the mass on a single mammogram. Other views of the patient, such as different views of the same breast, films of the other breast, previous mammograms, spot, and magnification views were not used to assist either the radiologist or the computer. Therefore, the discussion in the previous paragraph only compares the performances of the radiologist and the computer under specific laboratory conditions. The ma-

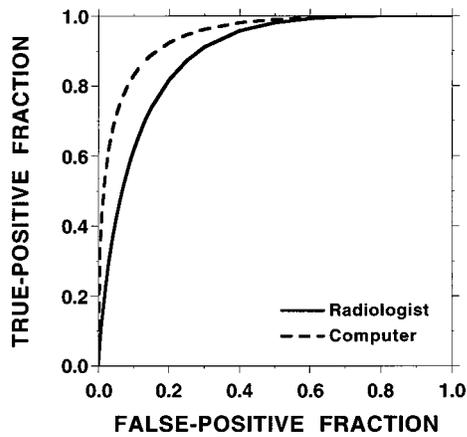


FIG. 10. ROC curves obtained by using the radiologist's malignancy rating ( $A_z = 0.89 \pm 0.03$ ) and the computer's discriminant score output ( $A_z = 0.94 \pm 0.02$ ) with features extracted from *R3* (the RBST images).

lignant and benign classification by radiologists can be expected to be more accurate when different views of the same mass are examined. The accuracy of computerized characterization is also expected to improve when the features or discriminant scores obtained from different mammograms of the same patient are combined. However, this was not performed in this study since our purpose was to compare the usefulness of the RBST with other image representations. Similarly, the ROC curves and the  $A_z$  scores in Sec. III do not necessarily reflect the accuracy expected to be obtained under clinical conditions, but they show the trend that the RBST is useful.

The segmentation, feature extraction and classification methods used in this work and that of Huo *et al.*<sup>23</sup> are different. However, in both investigations, features extracted from the area surrounding the segmented mass resulted in better classification accuracy compared to features extracted from other regions. Since the data sets are different, it is difficult to compare the performances of the two methods. The data set used in our study was almost twice as large as that used by the other study.<sup>23</sup> Huo *et al.* used an *ad hoc* method for geometric shape correction, and employed the maximum of the corrected measure in four different neighborhoods for better classification results. It remains to be seen whether these methods are generalizable to larger data sets. Similarly, when our feature selection and classification methods are applied to a larger data set, the selected features and the coefficients of the selected features in linear discriminant analysis are likely to change. It remains to be seen

whether the classification accuracy will decrease under these conditions.

An advantage of our approach compared to some recent publications<sup>22,23</sup> is that the mass characterization method proposed in this study is applicable to both spiculated and nonspiculated masses. As summarized in Table IV, at a 95% overall sensitivity level, our algorithm was able to correctly diagnose 100% of the spiculated malignant masses, and 89% of the nonspiculated malignant masses. At the same overall sensitivity level, the radiologist's rankings also showed 100% and 89% true-positive rates for spiculated malignant and nonspiculated malignant masses, respectively. However, at this sensitivity level, the computer had a 81% specificity (69 true negatives—68 nonspiculated and 1 spiculated) and the radiologist had a 60% specificity (51 true negatives—50 nonspiculated and 1 spiculated).

## V. CONCLUSION

We have developed a new image transformation method, referred to as RBST, for the characterization of mammographic masses. The results of our classification study indicate that texture features extracted from the transformed images are useful in differentiation of malignant and benign masses. With the best combination of texture features, the test  $A_z$  value on our database of 168 mammograms reached 0.94. It was found that texture features extracted from the transformed images were significantly more effective than features extracted from the ROIs before the transformation. This demonstrates the usefulness of the RBST. Before the applicability of our approach can be tested in a clinical setting, further studies need to be performed with a larger database to investigate the generalizability of these results. The combination of information from mammograms of different views obtained from the same patient will be investigated. The combination of texture and morphological features for benign and malignant characterization of masses will also be studied.

## ACKNOWLEDGMENTS

This work is supported by a Career Development Award (B.S.) from the USAMRMC (DAMD 17-96-1-6012) and a USPHS Grant CA 48129. The content of this publication does not necessarily reflect the position of the government, and no official endorsement of any equipment and product of any companies mentioned in the publication should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for providing the LABROC1 and CLABROC programs.

TABLE IV. Sensitivity (Sens.) and specificity (Spec.) for spiculated (S), and nonspiculated (NS) masses at a 95% overall sensitivity level.

	Malignant ( $n = 83$ )			Benign ( $n = 85$ )		
	S ( $n = 45$ )	NS ( $n = 38$ )	Overall Sens. ( $n = 83$ )	S ( $n = 6$ )	NS ( $n = 79$ )	Overall Spec. ( $n = 85$ )
Computer	100% ( $n = 45$ )	89% ( $n = 34$ )	95% ( $n = 79$ )	17% ( $n = 1$ )	86% ( $n = 68$ )	81% ( $n = 69$ )
Radiologist	100% ( $n = 45$ )	89% ( $n = 34$ )	95% ( $n = 79$ )	17% ( $n = 1$ )	63% ( $n = 50$ )	60% ( $n = 51$ )

- <sup>1</sup>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.
- <sup>2</sup>D. B. Kopans, "The positive predictive value of mammography," *Am. J. Roentgenol.* **158**, 521–526 (1992).
- <sup>3</sup>D. D. Adler and M. A. Helvie, "Mammographic biopsy recommendations," *Curr. Opin. Radiol.* **4**, 123–129 (1992).
- <sup>4</sup>M. Moskowitz, "Impact of a priori medical decisions on screening for breast cancer," *Radiology* **171**, 605–608 (1989).
- <sup>5</sup>C. J. D'Orsi, D. J. Getty, J. A. Swets, R. M. Pickett, S. S. Seltzer, and B. J. McNeil, "Reading and decision aids for improved accuracy and standardization of mammographic diagnosis," *Radiology* **184**, 619–622 (1992).
- <sup>6</sup>G. Hermann, C. Janus, I. S. Schwartz, B. Krivisky, S. Bier, and J. G. Rabinowitz, "Nonpalpable breast lesions: Accuracy of prebiopsy mammographic diagnosis," *Radiology* **165**, 323–326 (1987).
- <sup>7</sup>F. M. Hall, J. M. Storella, D. Z. Silverstone, and G. Wyshak, "Nonpalpable breast lesions: Recommendations for biopsy based on suspicion of carcinoma at mammography," *Radiology* **167**, 353–358 (1988).
- <sup>8</sup>H. G. Jacobson and J. Edeiken, "Biopsy of occult breast lesions: Analysis of 1261 abnormalities," *JAMA* **263**, 2341–2343 (1990).
- <sup>9</sup>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).
- <sup>10</sup>W. P. Kegelmeyer, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiology* **191**, 331–337 (1994).
- <sup>11</sup>H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark, "Markov random field for tumor detection in digital mammography," *IEEE Trans. Med. Imaging* **14**, 565–576 (1995).
- <sup>12</sup>B. Zheng, Y.-H. Chang, and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," *Acad. Radiol.* **2**, 959–966 (1995).
- <sup>13</sup>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).
- <sup>14</sup>D. Wei, H.-P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograms: Multiresolution texture analysis," *Med. Phys.* **22**, 1501–1513 (1995).
- <sup>15</sup>N. Petrick, H.-P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *IEEE Trans. Med. Imaging* **15**, 59–67 (1996).
- <sup>16</sup>J. A. Baker, P. J. Kornguth, J. Y. Lo, M. E. Williford, and C. E. Floyd, "Breast cancer: Prediction with artificial neural network based on BI-RADS standardized lexicon," *Radiology* **196**, 817–822 (1995).
- <sup>17</sup>D. J. Getty, R. M. Pickett, C. J. D'Orsi, and J. A. Swets, "Enhanced interpretation of diagnostic images," *Invest. Radiol.* **23**, 240–252 (1988).
- <sup>18</sup>Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C. E. Metz, "Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer," *Radiology* **187**, 81–87 (1993).
- <sup>19</sup>D. Brzakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammography," *IEEE Trans. Med. Imaging* **9**, 233–241 (1990).
- <sup>20</sup>J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computerized image analysis," *IEEE Trans. Med. Imaging* **12**, 664–669 (1993).
- <sup>21</sup>M. L. Giger, C. J. Vyborny, and R. A. Schmidt, "Computerized characterization of mammographic masses: Analysis of spiculation," *Cancer Lett.* **77**, 201–211 (1994).
- <sup>22</sup>S. Pohlman, K. A. Powell, N. A. Obuchowski, W. A. Chilcote, and S. G. Broniatowski, "Quantitative classification of breast tumors in digitized mammograms," *Med. Phys.* **23**, 1337–1345 (1996).
- <sup>23</sup>Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick, P. Lu, D. E. Wolverton, and R. A. Schmidt, "Analysis of spiculation in the computerized classification of mammographic masses," *Med. Phys.* **22**, 1569–1579 (1995).
- <sup>24</sup>L. Tabar and P. B. Dean, *Teaching Atlas of Mammography* (Thieme, New York, 1985).
- <sup>25</sup>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: An artificial neural network with morphological features," in *Proceedings of the World Congress on Neural Networks* (INNS Press, New Jersey, 1995), pp. 876–879.
- <sup>26</sup>B. Sahiner, H.-P. Chan, D. Wei, N. Petrick, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue," *Med. Phys.* **23**, 1671–1684 (1996).
- <sup>27</sup>J. Serra, *Image Analysis and Mathematical Morphology* (Academic, London, 1982).
- <sup>28</sup>R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973).
- <sup>29</sup>R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **3**, 610–621 (1973).
- <sup>30</sup>M. M. Galloway, "Texture analysis using gray level run lengths," *Comput. Graph. Image Process.* **4**, 172–179 (1975).
- <sup>31</sup>P. A. Lachenbruch, *Discriminant Analysis* (Hafner, New York, 1975).
- <sup>32</sup>M. J. Norusis, *SPSS Professional Statistics 6.1* (SPSS, Chicago, 1993).
- <sup>33</sup>C. E. Metz, "ROC methodology in radiographic imaging," *Invest. Radiol.* **21**, 720–733 (1986).
- <sup>34</sup>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 (1990).
- <sup>35</sup>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," in *Information Processing in Medical Imaging: Proceedings of the 8th Conference*, edited by F. Deconinck (Martinus Nijhoff, Boston, Brussels, 1984), pp. 432–445.
- <sup>36</sup>R. M. Rangayyan, N. El-Faramawy, J. E. L. Desautels, and O. A. Alim, "Discrimination between benign and malignant breast tumors using a region-based measure of edge profile acutance," in *Digital Mammography '96*, edited by K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt (Elsevier, Amsterdam, 1996), pp. 213–218.
- <sup>37</sup>R. F. Wagner, D. G. Brown, J.-P. Guedon, K. J. Myers, and K. A. Wear, "Multivariate Gaussian pattern classification: Effects of finite sample size and the addition of correlated or noisy features on summary measures of goodness," in *Information Processing in Medical Imaging*, edited by H. H. Barrett and A. F. Gmitro (Springer-Verlag, Berlin, 1993), pp. 507–524.