Ultrasound Obstet Gynecol 2008; **32**: 565–572
Published online 31 March 2008 in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/uog.5205

Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound

Y.-L. HUANG*, D.-R. CHEN†, Y.-R. JIANG*, S.-J. KUO†, H.-K. WU‡ and W. K. MOON§

**Department of Computer Science and Information Engineering, Tunghai University, Taichung and Departments of †Surgery and ‡Diagnostic Radiology, Changhua Christian Hospital, Changhua, Taiwan and §Department of Diagnostic Radiology, Seoul National University Hospital, Seoul, South Korea*

KEYWORDS: breast cancer; breast sonography; computer-aided diagnosis; morphological analysis; support vector machine; tumor classification; tumor segmentation

ABSTRACT

Objectives *To develop and evaluate a computer-aided diagnosis (CAD) system with automatic contouring and morphological analysis to aid in the classification of breast tumors using ultrasound.*

Methods *We evaluated 118 breast lesions (34 malignant and 84 benign tumors). Each tumor contour was automatically extracted from the digitized ultrasound image. Nineteen practical morphological features from the extracted contour were calculated and principal component analysis (PCA) was applied to find independent features. A support vector machine (SVM) classifier utilized the selected principal vectors to identify the breast tumor as benign or malignant. In this study, all the cases were sampled with k-fold cross-validation ($k = 10$) to evaluate the performance by receiver–operating characteristics (ROC) curve analysis.*

Results *The areas under the ROC curves for the proposed CAD systems using all morphological features and the lower-dimensional principal vector were 0.91 and 0.90, respectively. The classification ability for breast tumors using morphological information was good.*

Conclusions *This system differentiates benign from malignant breast tumors well and therefore provides a clinically useful second opinion. Moreover, the morphological features are nearly setting-independent and thus available to various ultrasound machines. Copyright © 2008 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

Early diagnosis and treatment is the most effective way of reducing mortality caused by breast cancer[1], and mammography and sonography are utilized to achieve this. Although mammography can visualize non-palpable and small tumors, when evaluating breast masses in daily clinical practice, sonography is more convenient as it is real-time, and is less expensive and time-consuming than mammography, and is safer for the patient[2]. Additionally, mammography typically has a low negative predictive value, so many patients with benign tumors are subjected to needless breast biopsies[2,3]. Breast sonography has been employed as an adjunct to mammography and overcomes this drawback[4].

Analysis of sonographic characteristics can assist in differentiating between benign and malignant lesions. Experienced physicians evaluate a breast tumor via a sonogram according to tumor morphology and the contrast of internal echoes. However, an ultrasound image always includes noise, visible as speckles, and tissue-related textures, and image interpretation is operator-dependent. To avoid unnecessary biopsies, the extra information provided by computer-aided diagnosis (CAD) algorithms might enhance diagnostic accuracy[5–8].

The textural variation between benign and malignant tumors is deemed a useful characteristic for their differentiation on ultrasound[9–11]. Common weaknesses of the previous CAD systems employing textural analysis are that they only work effectively with specific ultrasound systems. Information regarding shape, provided by a tumor contour, is also valuable to physicians when making diagnostic decisions[12]. Several previously proposed CAD algorithms have been shown to distinguish effectively and reliably between benign and malignant lesions by analyzing a tumor's shape[13,14]. The appearance of the morphological features was almost independent of sonographic gain setting and could tolerate reasonable variation in contour segmentation associated with the different machines used.

Correspondence to: Prof. Y.-L. Huang, Department of Computer Science and Information Engineering, Tunghai University, Taichung, Taiwan 407 (e-mail: ylhuang@thu.edu.tw)

Accepted: 9 July 2007

This study utilized a previously proposed segmentation algorithm[15] to automatically extract the contour of breast tumors from ultrasound images. This approach integrates the advantages of an adaptive initial contouring procedure and the 'level set' segmentation techniques to extract contours of a breast tumor from ultrasound images with very high reliability. After the segmentation procedure obtained the tumor contour from an ultrasound image, significant morphological features were calculated and then formed a feature vector. The proposed approach applies a dimension-reduction method, principal component analysis (PCA), to identify independent features. The original morphological feature vector is then transformed into the principal vector with fewer dimensions. The projected vector (i.e. the principal vector) thus summarizes the original vector. For support vector machine (SVM) analysis[16], the principal vector was employed to distinguish between benign and malignant lesions. The SVM is a reliable choice for this study because it is rapid and has excellent classification capability.

MATERIALS AND METHODS

Data acquisition

For this study we used an ultrasound image database including 118 images of pathologically proven benign breast tumors from 84 patients and carcinomas from 34 patients. The database included one image only from each patient. The ultrasound images were captured at the largest diameter of each tumor (which was > 1 cm in all cases). The patients' ages ranged from 22 to 67 (mean, 46) years. All digital ultrasound images were obtained using a Philips HDI 5000 system (Advanced Technology Laboratories, Bothell, WA, USA) with a L12-5 small parts transducer, which is a linear-array transducer with a frequency of 5–12 MHz and a scan width of 38 mm. Breasts were scanned by SonoCT® real-time compound imaging (Philips Medical Systems) in the survey mode but without using the harmonic technique. No acoustic standoff pad was used. The capturing resolution of ultrasound images was 640×476 pixels. Unless adjustment was necessary to obtain an adequate view, the sonographic gain setting remained unchanged throughout the entire study. Each monochrome ultrasound image was quantized into eight bits with 256 gray levels. The entire database was collected from December 2002 to May 2003 by an experienced radiologist (W.K.M.). The contours of the tumor were determined manually by a breast surgeon familiar with breast ultrasound interpretation (D.R.C.) and then saved in files for comparison with the automatically generated contours.

Image segmentation

Our method was based on previous work on contouring of breast tumors using sonography[15]. Here we provide a brief overview of the proposed segmentation algorithm.

Preprocessing is an important step before segmentation as medical ultrasound images include considerable noise and tissue textures that render contour segmentation difficult[17]. An effective preprocessing method for contouring should reduce noise and retain useful information. The proposed segmentation algorithm utilizes a sophisticated filter, the modified curvature diffusion equation (MCDE)[18], to enhance ultrasound images. The MCDE filter has been shown to be better than anisotropic diffusion at enhancing and preserving edges in a low-contrast image. After an image was preprocessed through the MCDE filter, an automatic thresholding method[19] was applied to approximately locate the tumor from an ultrasound image. This method can automatically identify the threshold that minimizes the intraclass variance of black and white pixels. The threshold-segmented area was then utilized as a reference from which we defined an initial circular tumor contour (Figure 1).

The level set method[20], a newly developed deformable model, is applied to reduce the time required to sketch a precise contour. It is a numerical technique for computing and analyzing curve propagation that has been applied successfully to solve a wide range of difficult problems in image segmentation. It offers a highly robust and accurate technique for tracking interfaces with complex motions. In the proposed segmentation algorithm, the level set method was employed to segment a tumor contour in an ultrasound image using the defined initial circular contour (Figure 1).

Feature extraction

The shape variation between benign and malignant tumors in an ultrasound image is an effective feature for classifying breast tumors. Nineteen practical morphological features[12,13,21] from contours extracted from images were used as features for classifying breast tumors. These features were defined as follows.

- *Perimeter*. The *Perimeter* feature represents the length of the tumor perimeter. As malignant tumors usually have irregular shapes, a large tumor perimeter is associated with the likelihood that a tumor is malignant.
- *Area*. The *Area* feature is the area of a breast tumor. Malignant tumors frequently have a large area compared with benign tumors.
- *NSPD* (number of substantial protuberances and depressions). The *NSPD* feature can be utilized to calculate the level of boundary irregularity. If p_i is a point in the contour, the k-curve angle of p_i, i.e. θ_i, can be obtained by p_i, p_{i+k} and p_{i-k}, where k is defined as 7 in this study, following Chen *et al.*[13]. The point p_i is a gradual point if θ_i is $\leq 40^\circ$ and is defined as a convex point or a concave point if θ_i is $> 40^\circ$. If there was no concave point between any two convex points, the convex point with the smallest k-curve angle would be eliminated. Similarly, if there was no convex point between any two concave points, the concave point with the smallest k-curve angle would be eliminated.

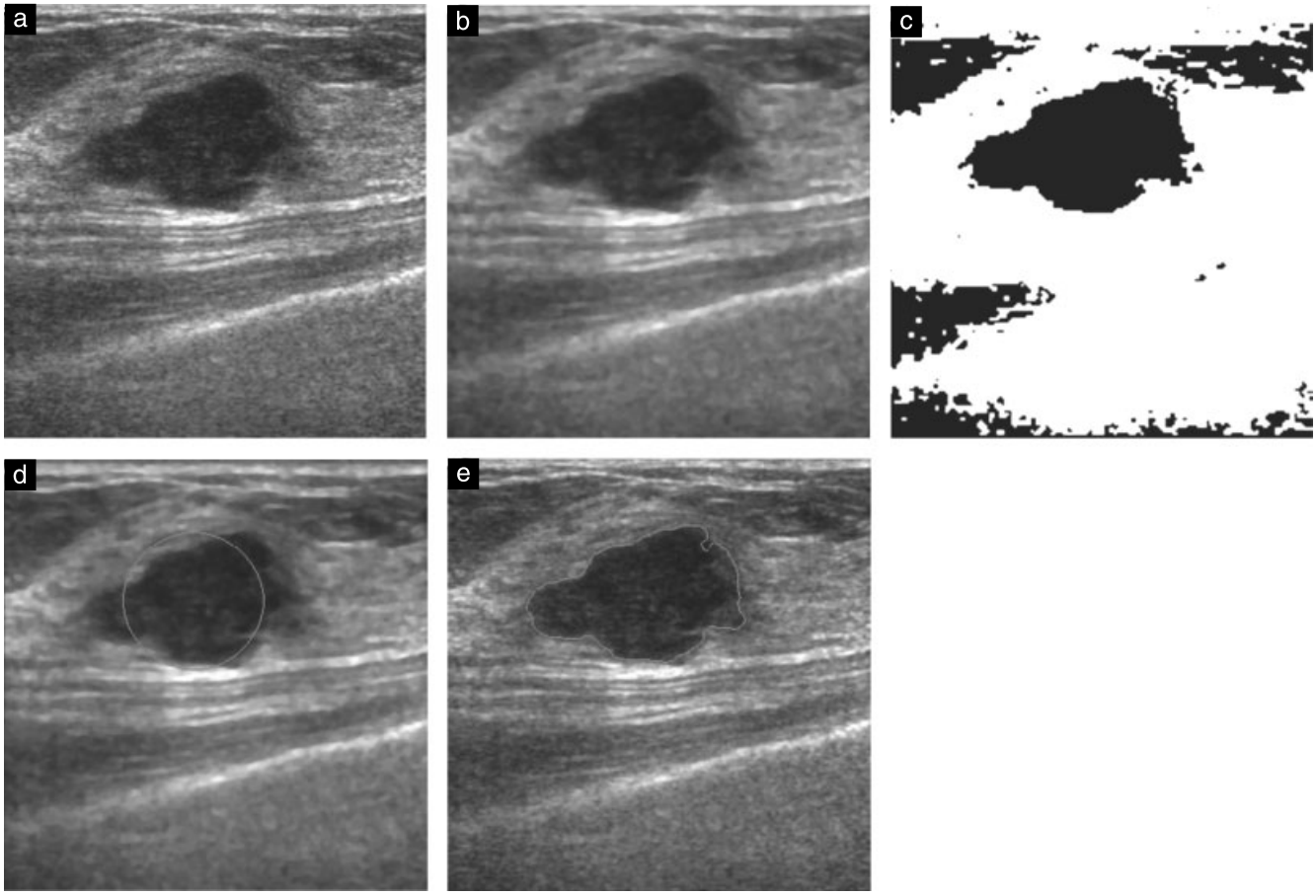

Figure 1 Examples of image segmentation: (a) original breast ultrasound image (malignant case); (b) after modified curvature diffusion equation filtering; (c) after the automatic thresholding method[19]; (d) the circular initial contour; (e) result of the 'level set' segmentation procedure.

Figure 2 gives an example of convex and concave points in a tumor contour. The *NSPD* is then defined as:

$$NSPD = 2 \times n, \tag{1}$$

where n is the number of concave points. Malignant lesions usually occur at an anfractuous boundary, and so malignant tumors have a larger *NSPD*.

- *LI* (lobulation index). According to the definition for a concave point from the *NSPD*, the lobe region enclosed by a lesion contour and a line connected by any two adjacent concave points can be obtained. If $A_{\max}$ and $A_{\min}$ denote the sizes of the largest and the smallest lobe regions, and A_{average} denotes the average size of all lobe regions, then *LI* can be defined as:

$$LI = \frac{A_{\max} - A_{\min}}{A_{\text{average}}}. \tag{2}$$

Usually, a malignant tumor has a larger *LI* than does a benign one.

- *ENC* (elliptic-normalized circumference). The angle of inclination for each tumor, with respect to the $x - y$ coordinate plane, can be obtained by using the second-order moment. The equivalent ellipse for each tumor with the same area, center and angle of inclination can then be generated. If *Equivalent_Ellipse_Perimeter* is the perimeter of the equivalent ellipse, the *ENC* can be

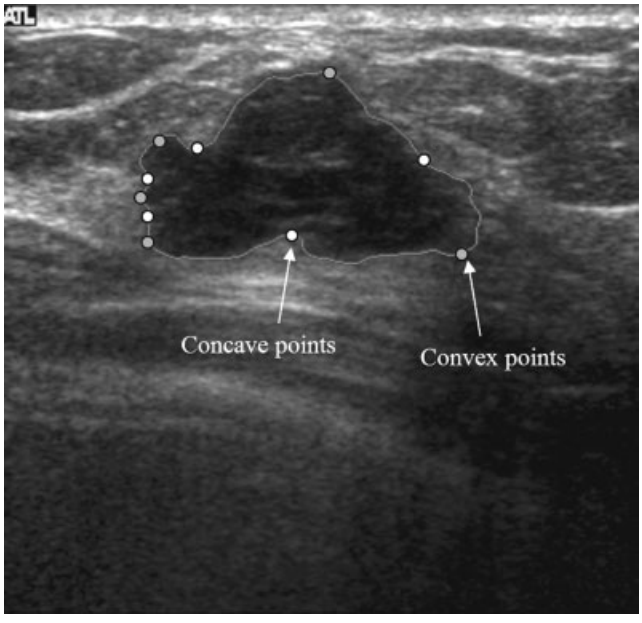

Figure 2 Example of convex points (gray) and concave points (white) in a malignant breast tumor contour.

defined as:

$$ENC = \frac{Equivalent\_Ellipse\_Perimeter}{Perimeter}. \quad (3)$$

When the *ENC* value of a suspected breast tumor is close to 1, the boundary of the tumor is smooth, and there is a high possibility that it is benign.

- *ENS* (elliptic-normalized skeleton). The skeleton of a tumor region expresses a set *S*, and *ENS* is defined as the sum of the skeleton points in *S*. When a tumor has a twisted boundary, the skeleton is also complex. Figure 3 shows an example skeleton of a malignant tumor. A malignant lesion always has a twisted boundary and generates a large *ENS*.
- *LS_Ratio* (long axis to short axis ratio). The *LS_Ratio* is the length ratio of the major (long) axis and minor (short) axis of the equivalent ellipse defined in the *ENC* feature.
- *Aspect_Ratio*. The *Aspect_Ratio* is the length ratio of a tumor's depth and width. If a tumor's depth exceeds its width, the *Aspect_Ratio* is greater than 1 and the tumor has a high probability of being malignant.
- $$Form\_Factor = \frac{4\pi \times Area}{Perimeter^2}. \quad (4)$$

When *Form_Factor* is close to 1, the tumor shape is nearly round.

- $$Roundness = \frac{4 \times Area}{\pi \times Max\_Diameter^2}, \quad (5)$$

where *Max_Diameter* denotes the length of the major axis from the equivalent ellipse of a tumor.

- $$Solidity = \frac{Area}{Convex\_Area}, \quad (6)$$

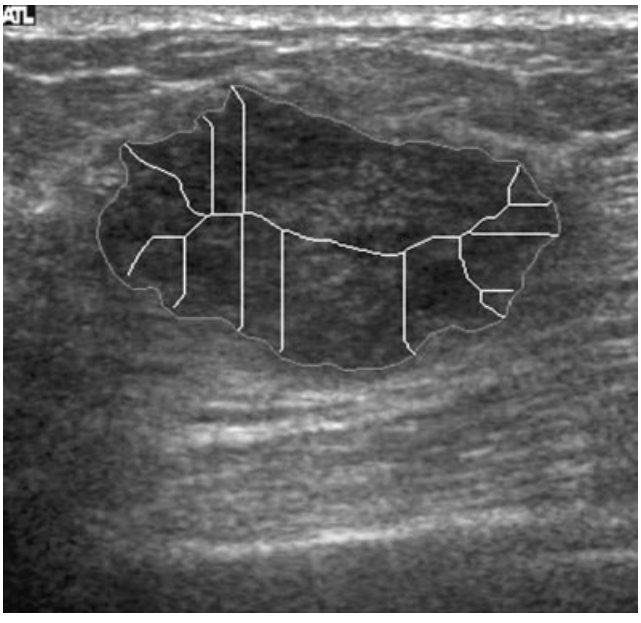

Figure 3 Example skeleton (internal lines) of a malignant breast tumor.

where *Convex_Area* is the area of the convex hull of a tumor. When *Solidity* is close to 0, the tumor is malignant.

- $$Convexity = \frac{Convex\_Perimeter}{Perimeter}, \quad (7)$$

where *Convex_Perimeter* is the perimeter of the convex hull of a tumor.

- $$Extent = \frac{Area}{Bounding\_Rectangle}, \quad (8)$$

where *Bounding_Rectangle* is the smallest rectangle containing the tumor.

- *TCA_Ratio*. The *TCA_Ratio* (tumor area to convex area ratio) is defined as:

$$TCA\_Ratio = \frac{Area}{Convex\_Area}. \quad (9)$$

- *TEP_Ratio* (tumor perimeter to ellipse perimeter ratio). The *TEP_Ratio* is the perimeter ratio of a tumor and the corresponding ellipse. The major and minor axes of the corresponding ellipse are calculated based on the proportion of width to depth of a tumor to acquire the same area for the ellipse and tumor.
- *TEP_Difference* (difference between tumor perimeter and ellipse perimeter). The *TEP_Difference* is defined as the difference between tumor perimeter and the corresponding ellipse.
- *TCP_Ratio* (tumor perimeter to circle perimeter ratio). The *TCP_Ratio* is the perimeter ratio of a tumor and the corresponding circle, the corresponding circle having the same area as the tumor.
- *TCP_Difference* (difference between tumor perimeter and circle perimeter). The *TCP_Difference* is defined as the difference between the tumor perimeter and the corresponding circle, the corresponding circle having the same area as the tumor.
- *AP_Ratio* (area to perimeter ratio). The *AP_Ratio* is the ratio of the area and the perimeter of a tumor.

Morphological features *NSPD*, *LI*, *ENC*, *ENS*, *LS_Ratio* and *Aspect_Ratio* were utilized in Chen's CAD system[13]. Chang *et al.* applied the *Form_Factor*, *Roundness*, *Aspect_Ratio*, *Solidity*, *Convexity* and *Extent* features to diagnose breast tumors[14]. The other features are fundamental and clinically useful indicators.

Principal component analysis (PCA) for vector dimension reduction

The above-mentioned morphological features were consistently in a high-dimensional space when determining the shape variety in ultrasound images. Using the high-dimensional vector directly was unsatisfactory when identifying breast tumors. Furthermore, numerous morphological features may be co-dependent. Only independent features should be utilized to attain reliable classification performance. To test the correlations between the

19 morphological features, the variance inflation factors were measured to perform collinearity diagnostics. This investigation found that the features *Perimeter, LS_Ratio, Aspect_Ratio, Roundness* and *Convexity* might follow significant collinearity. The data were analyzed by the Statistical Package for Social Sciences (SPSS for Windows, release 8.0, SPSS Inc, Chicago, IL, USA). Thus, PCA was applied in this study to reduce the dimensions of a feature vector, transforming the original feature vector into a lower dimensional principal vector.

Notably, PCA is a conventionally adopted statistical analytical method that decreases redundancy by projecting the original data over an appropriate basis. The idea behind the PCA is to create a more pertinent representation for reducing the dimensions of the original vectors. The effects of the new feature vector on the sonography database were then analyzed. The first six principal components ($n = 6$) explained over 95% of the total variability (Figure 4). According to the analysis, the ideal n value was 6; thus, each original 19-dimensional morphological feature vector was condensed by PCA into a new six-dimensional feature vector.

All analyses were carried out on a single CPU Intel Pentium-VI 2.4 GHz personal computer (ASUSTek Computer Inc., Taipei, Taiwan) with the Microsoft Windows XP® (Microsoft Corp., Redmond, WA, USA) operating system. The programs were performed using C++ language and compiled using the Microsoft Visual Studio®. The C++ codes of the PCA were performed using Matlab software (The MathWorks, Inc., Natick, MA, USA).

Classification

A SVM is a machine-learning system developed using statistical learning theories to classify data points into two classes[16]. Notably, SVM models have been applied extensively for classification[22,23], image recognition[24,25] and bioinformatics[26,27]. The concept of the SVM classification is described in the Appendix. In this study, the SVM model was used to classify tumors as either benign or malignant. The morphological features from a breast tumor formed a feature vector, which was then utilized as input signals for the SVM classifier. When the output value of a suspicious tumor region was ≥ 0, the CAD system classified the tumor in the ultrasound image as malignant. Conversely, when the output value was < 0, the tumor was diagnosed as benign.

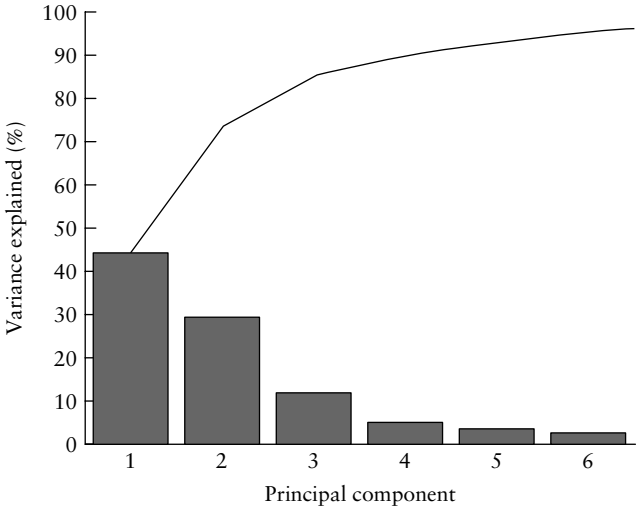

Figure 4 Bar graph of the first six principal components, which explained over 95% of the total variability in the standardized ratings.

Diagnosis evaluation

The k-fold cross-validation method[28] was used to evaluate the performance of the proposed CAD system. All cases were randomly divided into k groups. The first group was excluded and the other $(k - 1)$ groups functioned as the training set. The second group was used as a test group while the sonographic images in the remaining $(k - 1)$ groups were trained. This process was repeated until all k groups had been used in turn as the group used for testing. In the experiment, k was 10 and each group included 11 or 12 ultrasound images.

Two types of performance measure were used to assess the proposed system. The most common means for measuring diagnostic performance for reconstructed images is based on receiver–operating characteristics (ROC) curve analysis[29] with an index of the area (A_Z) under the ROC curve. This gives an indication of the quantitative measure of overall performance of a diagnostic system. The A_Z value can therefore compare performance using different methods to clearly distinguish between positive and negative findings of breast tumors. The second type of performance measure included the diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV).

RESULTS

The ultrasound image database used to test the performance of the proposed CAD system included 118 cases (84 benign breast tumors and 34 malignant ones). Diagnostic performance between different morphological feature subsets was compared by simulation. Table 1 lists the subset notations corresponding to morphological features and notations for CAD algorithms for the different feature sets. Because the radial kernels obtained the best result, kernels with parameter γ, defined as equation (15) in the Appendix, were chosen in the proposed CAD system. Figure 5 presents the diagnostic performance for the SVM system using different γ values. The proposed CAD system obtained the highest accuracy when $\gamma = 0.0001$.

Table 2 shows the classification of the 118 breast nodules using the proposed system and feature sets. Table 3 shows the performance of the feature subsets. Figure 6 presents ROC analysis of the proposed CAD system using the different feature sets.

Table 1 Subset notations corresponding to morphological features and notations for the proposed computer-aided diagnosis (CAD) algorithms for the different feature sets

Notation	*Description*
F_a	Feature subset (6-D) including the features: *NSPD, LI, ENC, ENS, LS_Ratio* and *Aspect_Ratio*.
F_b	Feature subset (6-D) including the features: *Form_Factor, Roundness, Aspect_Ratio, Solidity, Convexity* and *Extent*.
F_c	Feature subset (8-D) including the features: *Perimeter, Area, TCA_Ratio, TEP_Ratio, TEP_Difference, TCP_Ratio, TCP_Difference* and *AP_Ratio*.
F_e	Entire feature set (19-D) including all morphological features of F_a, F_b and F_c
F_{PCA}	Feature set (6-D) including all morphological features of F_a, F_b and F_c with principal component analysis (PCA)
AC_F_e	CAD algorithm with feature set F_e using automatic contours
AC_F_a	CAD algorithm with feature set F_a using automatic contours
AC_F_b	CAD algorithm with feature set F_b using automatic contours
AC_F_c	CAD algorithm with feature set F_c using automatic contours
AC_F_{PCA}	CAD algorithm with feature set F_{PCA} using automatic contours
MC_F_e	CAD algorithm with feature set F_e using manual contours
MC_F_{PCA}	CAD algorithm with feature set F_{PCA} using manual contours

-D, dimensional.

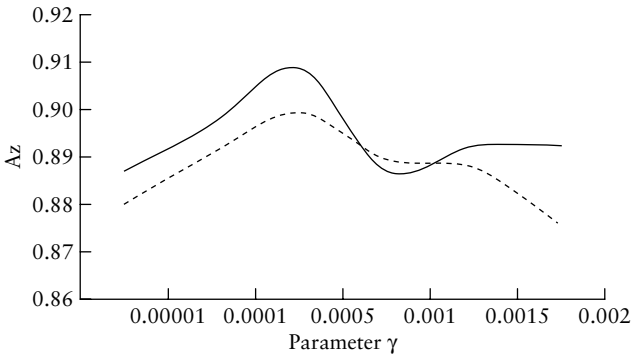

Figure 5 Performance of the proposed computer-aided diagnosis system with different parameter γ values, showing F_e (——) and F_{PCA} (----). For definitions of F_e and F_{PCA}, see Table 1.

Table 2 Classification of breast nodules by support vector machine (SVM) model

	Histological finding	
SVM F_e (F_{PCA}) output	*Benign*	*Malignant*
< 0	TN: 65 (66)	FN: 2 (3)
≥ 0	FP: 19 (18)	TP: 32 (31)
Total	84	34

FN, false negatives (no. malignant cases misdiagnosed); FP, false positives (no. benign cases misdiagnosed); TN, true negatives (no. benign cases diagnosed correctly); TP, true positives (no. malignant cases diagnosed correctly).

DISCUSSION

This study proposed an automatic diagnostic system that utilizes practical morphological features to effectively distinguish between benign and malignant breast lesions. The 19 morphological features from the extracted contours within the ultrasound images were applied as classification criteria. However, selected morphological features were correlated, and only independent features should be used to ensure reliable performance of such a system. PCA was applied to transform the original morphological features into a low-dimensional principal vector. The SVM model utilized the principal vector to classify the breast lesions. The training and diagnostic procedure of SVM is faster and more stable than that of multilayer feed-forward neural networks[30]. With database expansion, new cases can be gathered easily and used as references.

Rapid development of ultrasound technology has led to the use of many different ultrasound systems for medical diagnosis. The primary concerns when designing a CAD system for various sonographic systems are resolution and contrast. Thus, setting independence is important for a CAD algorithm. Morphological features based on tumor shape provided important information for the CAD system using medical ultrasound images. From the viewpoint of a CAD system applied to sonograms,

Table 3 Performance of the proposed computer-aided diagnosis (CAD) system

Algorithm	A_Z	*SD*	*Accuracy (%)*	*Sensitivity (%)*	*Specificity (%)*	*PPV (%)*	*NPV (%)*
AC_F_a	0.8667	0.0330	77.12	91.18	71.43	56.36	95.24
AC_F_b	0.8715	0.0343	75.42	88.24	70.24	54.55	93.65
AC_F_c	0.8984	0.0294	78.81	88.24	75.00	58.82	94.03
AC_F_e	0.9087	0.0265	82.20	94.12	77.38	62.75	97.01
AC_F_{PCA}	0.8993	0.0279	82.20	91.18	78.57	63.27	95.65
MC_F_e	0.8891	0.0327	82.20	91.18	78.57	63.27	95.65
MC_F_{PCA}	0.8711	0.0384	82.20	88.24	79.76	63.83	94.37

Accuracy = (TP + TN)/(TP + TN + FP + FN); sensitivity = TP/(TP + FN); specificity = TN/(TN + FP); positive predictive value (PPV) = TP/(TP + FP); negative predictive value (NPV) = TN/(TN + FN). FN, false negatives (no. malignant cases misdiagnosed); FP, false positives (no. benign cases misdiagnosed); TN, true negatives (no. benign cases diagnosed correctly); TP, true positives (no. malignant cases diagnosed correctly).

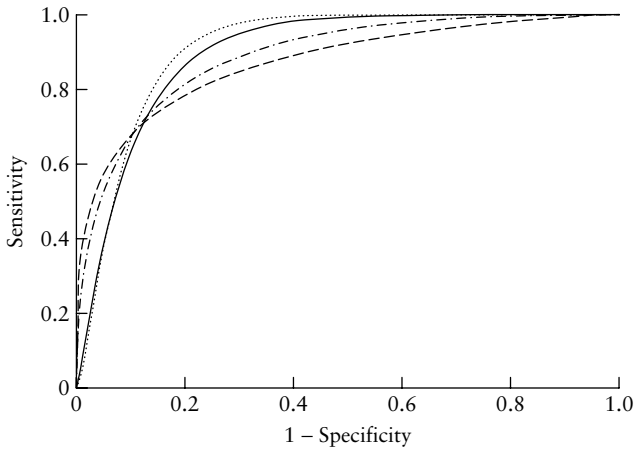

Figure 6 Receiver–operating characteristics (ROC) curve analysis of the proposed computer-aided diagnosis system with different feature sets (defined in Table 1). AC_F_{PCA} (———): area under the ROC curve (A_Z) = 0.8993; MC_F_{PCA} (– – – –): A_Z = 0.8711; AC_F_e (· · · · ·): A_Z = 0.9090; MC_F_e (– · – · –): A_Z = 0.8891.

morphological features are nearly setting-independent and can tolerate reasonable variation in contour segmentation.

According to the diagnostic results (Table 3), the AC_F_a experiment (using feature set F_a that was selected in the CAD of Chen *et al.*[13]) attained A_Z values and a classification accuracy of 0.87 and 77.1%, respectively. The AC_F_b experiment (using feature set F_b, as in the CAD of Chang *et al.*[14]) attained A_Z values and a classification accuracy of 0.87 and 75.4%, respectively. The AC_F_c experiment attained A_Z values and a classification accuracy of 0.90 and 78.8%, respectively. These experimental results indicate that morphological features are helpful when classifying benign and malignant tumors via sonography. The proposed CAD system with the entire feature set (the AC_F_e experiment) for classifying malignancies achieved good accuracy (82.2%) and a relatively high sensitivity (94.1%), meaning that for a practical CAD system, various feature combinations should be applied as classification criteria. However, 19 features for 118 breast lesions are too many and have a detrimental effect on classification. The AC_F_{PCA} algorithm obtained a similar diagnostic result by using the lower-dimensional principal vector. Notably, PCA can help to find the best features for classifying benign and malignant tumors.

Our experimental results suggest that using morphological features based on tumor contour for classifying benign and malignant tumors is effective and reliable. Morphology-based diagnosis of breast tumors has the advantage of being practically independent of both the ultrasound system settings and the type of ultrasound machine. In terms of decision-making, the high sensitivity and NPV demonstrated that the proposed CAD system can recognize malignant tumors and reduce the need for unnecessary biopsies for benign tumors. Thus, this automated CAD system is useful for differential diagnosis of breast tumors based on ultrasound images, and could lead to a decreased need for breast biopsies. Medical costs and adverse reactions will be reduced as well. In the future, we hope to improve the performance of the proposed CAD system by adding other features (such as echotexture, spiculations, blood flow)[9,31,32] of breast tumors. Additionally, three-dimensional sonography is being used increasingly in the clinical setting. Future work should also apply the proposed CAD system to three-dimensional sonograms.

APPENDIX

The aim of a SVM is to find a hyperplane to separate the training data with a maximal margin. Given a set of training vectors with m vectors belonging to separate classes, (x_1, y_1), (x_2, y_2), (x_3, y_3), . . ., (x_m, y_m), let (x_i, y_i) belong to the training vectors, where x_i denotes the value of the input vector and $y_i \in \{+1, -1\}$ is the corresponding desired output. The maximal margin of the separating hyperplane aims to find a pair (w, b) that can achieve an optimal hyperplane:

$$wx + b = 0, \tag{10}$$

where $w \in R^n$ and $b \in R$. The input vectors must be satisfied by

$$y_l = wx_l + b \geq 1, \tag{11}$$

$$y_l = wx_l + b \leq -1, \tag{12}$$

which combine to give:

$$y_l(wx_1 + b) \geq 1. \tag{13}$$

Figure 7 shows an optimal separating hyperplane with the largest possible margin, and the points outside of the

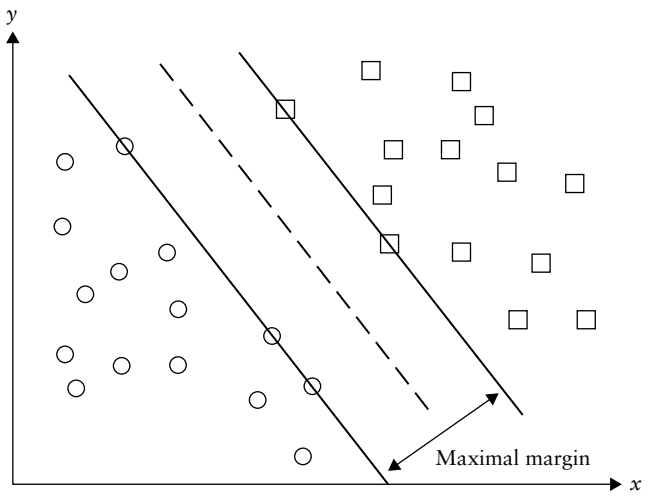

Figure 7 Example of a support vector machine with an optimal separating hyperplane (— —), defined by the equation $wx + b = 0$. The lines above and below are defined by $y = wx + b = 1$ and $y = wx + b = -1$, respectively. Benign (○) and malignant (□) test samples are plotted.

margin border denote the support vectors. The solution to the classification is given by the decision function:

$$f(x) = sign\left(\sum_{i=1}^{l} \alpha_i\, y_i k(s_i, x) + b\right), \qquad (14)$$

where α_i is the positive Lagrange multiplier, s_i are the support vectors (l in total), and $k(s_i, x)$ is the function for convolution of the kernel of the decision function. When $f(x) \geq 0$, the support vector is the same class as $y = 1$, and when $f(x) < 0$, the support vector is the same class as $y = -1$. The radial kernel performed best in our experimental comparison, hence was chosen in the proposed diagnosis system. The radial kernel was defined as:

$$k(x, y) = \exp\left(-\gamma(x - y)^2\right), \qquad (15)$$

where $\gamma \in R$ is a non-zero parameter.

ACKNOWLEDGMENT

We thank the National Science Council of the Republic of China for financially supporting this research, under Contract No. NSC95-2213-E-029-003.

REFERENCES

1. *Breast Cancer Facts and Figures 2003–2004*. American Cancer Society: Atlanta, Georgia, 2005.
2. Jackson VP, Bassett LW. Breast sonography. *Breast Dis* 1998; **10**: 55–66.
3. Rahbar G, Sie AC, Hansen GC, Prince JS, Melany ML, Reynolds HE, Jackson VP, Sayre JW, Bassett LW. Benign versus malignant solid breast masses: US differentiation. *Radiology* 1999; **213**: 889–894.
4. Gefen S, Tretiak OJ, Piccoli CW, Donohue KD, Petropulu AP, Shankar PM, Dumane VA, Huang L, Kutay MA, Genis V, Forsberg F, Reid JM, Goldberg BB. ROC analysis of ultrasound tissue characterization classifiers for breast cancer diagnosis. *IEEE Trans Med Imaging* 2003; **22**: 170–177.
5. Chen DR, Chang RF, Huang YL. Computer-aided diagnosis applied to US of solid breast nodules by using neural networks. *Radiology* 1999; **213**: 407–412.
6. Giger ML. Computerized analysis of images in the detection and diagnosis of breast cancer. *Semin Ultrasound CT MR* 2004; **25**: 411–418.
7. Drukker K, Giger ML, Vyborny CJ, Mendelson EB. Computerized detection and classification of cancer on breast ultrasound. *Acad Radiol* 2004; **11**: 526–535.
8. Collins MJ, Hoffmeister J, Worrell SW. Computer-aided detection and diagnosis of breast cancer. *Semin Ultrasound CT MR* 2006; **27**: 351–355.
9. Huang YL, Kuo SJ, Chang CS, Liu YK, Moon WK, Chen DR. Image retrieval with principal component analysis for breast cancer diagnosis on various ultrasonic systems. *Ultrasound Obstet Gynecol* 2005; **26**: 558–566.
10. Chang RF, Kuo WJ, Chen DR, Huang YL, Lee JH, Chou YH. Computer-aided diagnosis for surgical office-based breast ultrasound. *Arch Surg* 2000; **135**: 696–699.
11. Chen DR, Chang RF, Huang YL, Chou YH, Tiu CM, Tsai PP. Texture analysis of breast tumors on sonograms. *Semin Ultrasound CT MR* 2000; **21**: 308–316.
12. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology* 1995; **196**: 123–134.
13. Chen CM, Chou YH, Han KC, Hung GS, Tiu CM, Chiou HJ, Chiou SY. Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks. *Radiology* 2003; **226**: 504–514.
14. Chang RF, Wu WJ, Moon WK, Chen DR. Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. *Breast Cancer Res Treat* 2005; **89**: 179–185.
15. Huang YL, Jiang YR, Chen DR, Moon WK. Level set contouring for breast tumor in sonography. *J Digit Imaging* 2007; [Epub ahead of print].
16. Christianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press: Cambridge, 2000.
17. Drukker K, Giger ML, Mendelson EB. Computerized analysis of shadowing on breast ultrasound for improved lesion detection. *Med Phys* 2003; **30**: 1833–1842.
18. Whitaker RT, Xinwei X. Variable-conductance, level-set curvature for image denoising. *Proc Int Conf Image Process* 2001; **3**: 142–145.
19. Otsu N. Threshold selection method from gray-level histograms. *IEEE T Syst Man Cy* 1979; **9**: 62–66.
20. Osher S, Sethian J. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J Comput Phys* 1988; **79**: 12–49.
21. Tohno E, Cosgrove DO, Sloane JP. *Ultrasound Diagnosis OF Breast Diseases*. Churchill Livingstone: Edinburgh, 1994.
22. Kim KI, Jung K, Park SH, Kim HJ. Support vector machines for texture classification. *IEEE T Pattern Anal* 2002; **24**: 1542–1550.
23. Song Q, Hu WJ, Xie WF. Robust support vector machine with bullet hole image classification. *IEEE T Syst Man Cy C* 2002; **32**: 440–448.
24. El Naqa I, Yang YY, Wernick MN, Galatsanos NP, Nishikawa RM. A support vector machine approach for detection of microcalcifications. *IEEE Trans Med Imaging* 2002; **21**: 1552–1563.
25. Yang MH, Roth D, Ahuja N. A tale of two classifiers: SNoW vs. SVM in visual recognition. In *Computer Vision – ECCV 2002*, Vol 2353. Springer: Berlin/Heidelberg, 2002; 685–699.
26. Sun YF, Fan XD, Li YD. Identifying splicing sites in eukaryotic RNA: support vector machine approach. *Comp Biol Med* 2003; **33**: 17–29.
27. Song MH, Breneman CM, Bi JB, Sukumar N, Bennett KP, Cramer S, Tugcu N. Prediction of protein retention times in anion-exchange chromatography systems using support vector regression. *J Chem Inf Comp Sci* 2002; **42**: 1347–1357.
28. Weiss SM, Kapouleas I. An empirical comparison of pattern recognition neural nets and machine learning classification methods. *Proceedings of the 11th International Joint Conference on Artificial Intelligence* 1989; 234–237.
29. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29–36.
30. Huang YL, Wang KL, Chen DR. Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines. *Neural Comput Appl* 2006; **15**: 164–169.
31. Huang SF, Chang RF, Chen DR, Moon WK. Characterization of spiculation on ultrasound lesions. *IEEE Trans Med Imaging* 2004; **23**: 111–121.
32. Chang RF, Huang SF, Moon WK, Lee YH, Chen DR. Computer algorithm for analysing breast tumor angiogenesis using 3-D power Doppler ultrasound. *Ultrasound Med Biol* 2006; **32**: 1499–1508.

