34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Fourier-Based Shape Feature Extraction Technique for Computer-Aided B-Mode Ultrasound Diagnosis of Breast Tumor

Jong-Ha Lee[1], Yeong Kyeong Seong[1], Chu-Ho Chang[1], Jinman Park[1], Moonho Park[1], Kyoung-Gu Woo[1], and Eun Young Ko[2]

Abstract— **Early detection of breast tumor is critical in determining the best possible treatment approach. Due to its superiority compared with mammography in its possibility to detect lesions in dense breast tissue, ultrasound imaging has become an important modality in breast tumor detection and classification. This paper discusses the novel Fourier-based shape feature extraction techniques that provide enhanced classification accuracy for breast tumor in the computer-aided B-mode ultrasound diagnosis system. To demonstrate the effectiveness of the proposed method, experiments were performed using 4,107 ultrasound images with 2,508 malignancy cases. Experimental results show that the breast tumor classification accuracy of the proposed technique was 15.8%, 5.43%, 17.32%, and 13.86% higher than the previous shape features such as number of protuberances, number of depressions, lobulation index, and dissimilarity, respectively.**

I. INTRODUCTION

According to the American Cancer Society, more than 178,000 women and 2,000 men are found to be afflicted with breast cancer every year; international statistics report an estimated 1,152,161 new cases annually [1]. This form of the disease is the leading killer of females between 40 and 55 years of age, and is statistically the second leading cause of death overall in women. Clearly, early detection and diagnosis is the key to surviving this fatal disease. The sooner the tumors are detected and diagnosed, the better treatment can be applied.

Screening through mammography is one of the most often used diagnostic tools. Mammography, however, is not effective for dense breasts, with high false negative rates [2]. In recent years, ultrasound imaging has become an important tool for early breast tumor detection, proving complementary diagnostic information for mammography, especially in patients with dense breast tissue [3]. To evaluate breast tumors via ultrasound, radiologists consider several features in the image, such as lesion shape, orientation, echo pattern, and posterior acoustic enhancement [4], [5]. Interpretation of ultrasound images, however, is subjective and variability is very high due to its low image resolution and the different experiences of radiologists who analyze the tumor features. Variations in human perception of the image, different features used in the interpretation of the image, and a lack of the quantitative measures of the features used for image analysis, result in diagnostic variability between radiologists. Consequently, final confirmation is achieved through other modalities, such as CT, MRI, or biopsy [6].

In order to remove operator dependency and increase diagnostic accuracy, the computer-aided diagnosis (CAD) system provides a valuable method for breast tumor detection and classification. To describe the target in the CAD system, sonographic findings should be quantified into computerized values. The effectiveness of echo and texture features have already been shown to improve diagnosis of breast tumor [7]. However, the echo and texture features are sensitive to image noise and modality setting. Mass shape is one of the most significant image features [8]. The shapes of malignant breast tumors are more complex than benign lesions due to their infiltration characteristics into surrounding tissues. Malignant lesions are usually irregular, microlobulated, or spiculated whereas benign lesions are oval, round, or macrolobulated. Thus, if the irregularity of the sonographic finding is quantified correctly, the accuracy of the CAD scheme can be improved significantly. Several boundary-based methods analyzed in the spatial domain have been used to describe mass shape, such as roundness, concaveness, and convexness [8], [9]. However, throughout these methods, irregularity of mass shape is only expressed locally. Global important information to represent shape irregularity are not considered in boundary-based methods. For instance, if irregularity characteristics only exist locally, and most other areas show non-irregular characteristics, boundary-based methods will confirm that the shape is irregular, not regular. Thus, it is necessary to develop a new feature extraction technique to express shape irregularity globally.

In this paper, we propose a Fourier-based shape feature extraction approach for ultrasound images that are able to model subtle shape differences between benignancies and malignancies in the global aspect of view, thus improving CAD classification accuracy. The proposed technique is developed by applying Fourier transform on a shape signature which is derived by two different functions. The first function we utilized is the centroid distance function (CDF). The CDF is expressed by the distance of the mass boundary points from the centroid of a mass shape. The second function we developed is the centroid shape context function (CSCF). The CSCF captures the distribution of the mass boundary points by the log-polar diagram. Finally, the results from CDF and CSCF are combined together to improve classification perfor-

*This work was not supported by any organization.

[1]J.-H. Lee, Y. K. Seong, C.-H. Chang, J. Park, M. Park, and K.-G. Woo are with the Data Analytics Group, Future IT Laboratories, Samsung Advanced Institute of Technology, Yongin-Si, South Korea `jongha3.lee@samsung.com`

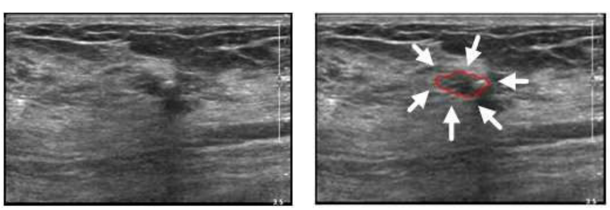
[2]E. Y. Ko is a medical doctor in the Department of Radiology, Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea `ey.ko@samsung.com`

mance. The proposed method takes advantage of considering mass shape irregularity globally, thus even if irregularity of the mass exists locally, shape irregularity can be expressed correctly. To verify the performance of the proposed features for breast tumor classification, experiments were done using 4,107 breast ultrasound images of *in vivo* human subjects containing 2,508 malignant cases.

The reminder of this paper is organized as follows. Section 2 introduces ultrasound images acquisition and data set used in this research. Section 3 presents proposed shape features allowing improved classification accuracy for breast tumors. The support vector machine as a classifier for breast tumors as benign or malignant is also discussed. Section 4 demonstrates performance of the proposed approach in comparison with that of other shape feature extraction approaches. Finally, Section 5 concludes the paper.

II. Ultrasound Image Acquisition and Data Collection

Ultrasound images of 4,107 breast tumors were collected from the Samsung Medical Center, Seoul, South Korea, between 2006 and 2010. The mean age of benign cases was 45 years, and the age range was from 11 to 81. The mean age of malignant cases was 49 years, and the age range was from 24 to 86. The test sets consisted of 1,599 benign and 2,508 malignant tumors. All images were taken using a Philips ATL iU22 ultrasound machine under the approval of the institutional review board of Samsung Medical Center. The scanner was equipped with a 5 to 12 MHz, 6 cm linear probe. The B-mode image size was 1024 × 768 pixels, with a spatial resolution of 0.23 mm/pixel. For each image, tumor boundary was manually outlined by a radiologist. Fig. 1 shows an example of the manual segmentation. Fig. 1(a) shows an example of an original ultrasound image with a malignant tumor. Fig. 1(b) shows the manual segmentation result for the tumor.



(a) (b)

Fig. 1: The example of detection and segmentation of a breast tumor. (a) The original image with a benign tumor, (b) Manual segmentation of the tumor (red line).

III. Fourier-Based Shape Feature Extraction

In the previous section, we described ultrasound image acquisition and manual segmentation of breast tumor. In this section, we describe the proposed Fourier-based technique to extract shape feature from the manually segmented tumor area. Initially, control points are subsampled along the boundary of the segmented tumor area, then the shape features are extracted from the Fourier descriptor (FD) based on centroid distance function (CDF) and centroid shape context function (CSCF). Using the staked vector approach, several feature vectors are combined into a single vector of features. Finally, a support vector machine (SVM) with Radial Basis Function (RBF) classifier was used for classification of the given image as benign or malignant. We now describe the feature extraction and classification steps used in the proposed algorithm in more detail.

A. Centroid Distance Function

Centroid distance function (CDF) is a one-dimensional (1-D) shape signature calculated from shape boundary coordinates. This shape signature captures the perceptual characteristic of the shape. The centroid distance function $r(n)$ is expressed by the distance of the contour points from the centroid (g_x, g_y) of a shape as $r(n) = [(x_n - g_x)^2 + (y_n - g_y)^2]^{1/2}$, where x_n and y_n are coordinates of shape boundary. (g_x, g_y) are calculated by $g_x = \frac{1}{6A}\sum_{i=0}^{N-1}(x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$ and $g_y = \frac{1}{6A}\sum_{i=0}^{N-1}(y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$, where A is the segmented shape area in the image. The centroid distance representation is invariant to translation because of the subtraction of centroid from boundary coordinates. Due to this property, classification performance will not be affected by the position of tumors in the ultrasound image.

B. Centroid Shape Context Function

Consider the vector set originating from a initial point to all other contour points on a shape. Then the vectors show the configuration of the entire shape relative to the reference point. Centroid shape context function (CSCF) captures the entire shape relative to the centroid of the mass using the log-polar diagram and expresses it through the shape histogram [10], [11]. The bins of log-polar diagram are uniform in log-polar space, which makes the diagram more sensitive to positions of adjacent points than to those of points far apart. An example shown in Fig. 2(a) is the log-polar diagram that has 5 bins for polar direction (r) and 12 bins for the angular direction (θ). The histogram of mass centroid is formed by putting the center of the diagram on the centroid of the mass. Then each bin of the histogram contains a count of all other sample points on the shape falling into that bin. The number of points in each bin is then expressed by the two-dimensional (2-D) histogram. 2-D histogram calculated at centroid can be converted into 1-D histogram by ordering with regard to decreasing polar distance. An example of 2-D histogram and converted 1-D histogram is shown in Fig. 2(b) and (c). The example of proposed CDF and CSCF shape signatures extracted from the segmented boundary of Fig. 1(b) are shown in Fig. 3.

C. Fourier Descriptor

For analysis and synthesis of the plane-closed curves, the Fourier descriptor (FD) is used [12]. FD is obtained by applying Fourier transform on a shape signature derived from shape boundary coordinates in Section III-A and III-B.

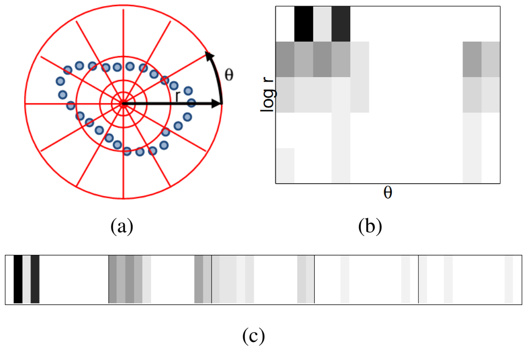


Fig. 2: The centroid shape context and its histogram. (a) The diagram of the log-polar bins, (b) The 2-D histogram for the boundary points, (c) The converted 1-D histogram.

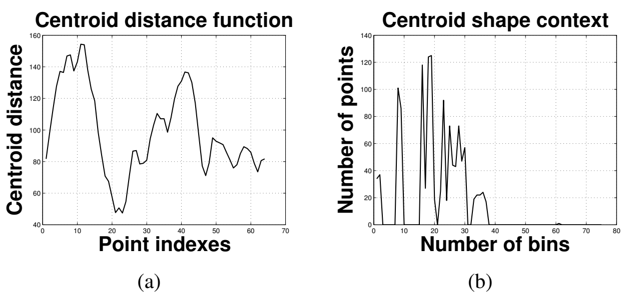


Fig. 3: The shape signature from (a) centroid distance function (CDF), (b) centroid shape context function (CSCF).

The normalized transformed Fourier coefficients are called the Fourier descriptors. The discrete Fourier transform of signature function is given by

$$a_n = \frac{1}{N}\sum_{t=0}^{N-1} r(t)\exp\left(\frac{-j2\pi nt}{N}\right),\ n = 0, 1, \ldots, N-1 \quad (1)$$

Since the CDF and CSCF are invariant to translation and rotation, the Fourier coefficients have to be further to make them scale invariant and start point independent shape descriptors. The relation between Fourier coefficients of original shape and transformed through scaling and change of start point is given by

$$a_n = \exp(jn\varphi) \cdot s \cdot a_n^o \quad (2)$$

where φ is the angles incurred by the start point change and and s is the scale factor. Then normalized Fourier coefficients b_n of the transformed shape is obtained by

$$b_n = \frac{a_n}{a_1} = \frac{a_n^o}{a_1^o}\exp[j(n-1)\varphi] = b_n^o \exp[j(n-1)\varphi] \quad (3)$$

where b_n^o is the normalized Fourier coefficients of the original shape. From Eq. (3), it is clear that, if we ignore phase of the coefficients, then magnitudes $|b_n|$ and $|b_n^o|$ are the same. In other words, $|b_n|$ is invariant to translation, rotation, scaling and start point change. The values of $|b_n|$ are the shape features that we used for the breast tumor classification.

For the improvement in the classification of ultrasound images, we have combined both Fourier descriptors from CDF and CSCF through staked vector approach (SVA) technique, where both Fourier descriptors are concatenated to form a long feature vector.

D. Support Vector Machine (SVM) Classifier

Support vector machine (SVM) is widely used method for classification and regression tasks [13]. In the two class problem like the one in this paper, the aim is to find an optimal separating hyperplane. For a two-class problem with training samples $\{x_i, y_i\}, i = 1, 2, \ldots, N$ in $d-$dimensional feature space, $x_i \in R^d$, with associated targets, $y_i \in \{-1, 1\}$, then the discriminate function of the separating hyperplane for linearly separable classes is given as

$$f(x) = w \cdot \Phi(x) + b \quad (4)$$

where $w \in R^d$ is the vector normal to the hyperplane and $b \in R$ is the bias. SVM fit a hyperplane to the training samples of two classes in the feature space by minimizing the cost function consisted of two criteria, namely margin maximization and error minimization. The data points closest to hyperplane are called support vectors. In this paper, classification is done through a SVM classifier with a Radial Basis Function (RBF).

In summary, the proposed Fourier-based shape feature extraction scheme can be described as follows.

Step 1. Boundary extraction from the input ultrasound image.

Step 2. Find Fourier descriptors from centroid distance function.

Step 3. Find Fourier descriptors from centroid shape context function.

Step 4. Staking of Fourier features extracted in step 2 and 3, to make a single feature vector.

Step 5. SVM classifier is used to classify the input ultrasound image by using staked feature vector found in step 4.

IV. Experimental Results

To test the performance of the computed shape features, classification experiments were performed. In total, 94 features were extracted from each ROI of the 4,107 ultrasound images: 31 CDF features and 63 CSCF features. Feature selection was then done by searching the space of feature subsets, and evaluating each one. In this experiment, a 43 reduced feature dataset was obtained from the full feature dataset using a correlation based feature selection evaluator and bestfirst search method [13]. Every experiment was performed using a 2.93 GHz Intel Xeon CPU workstation with 3GB of RAM and the Windows 7 operating system.

To calculate classification accuracy, the k-fold cross validation method was used and set k as 10. The pathological results were used as the ground truth, and the performance indices were evaluated by five performance indices, including classification accuracy (TP+TN)/(TP+TN+FP+FN), sensitivity (TP/[TP+FN]), specificity (TN/[TN+FP]), positive predictive value (TP/[TP+FP]), and negative predictive value

(TN/[TN+FN]), where TP is the number of true-positive findings correctly classified as positive (i.e., a malignant tumor is considered as malignant); TN, true-negative; FP, false-positive; and FN, false-negative [14].

Fig. 4 represents the performance of the proposed features compared with other shape feature sets: number of protuberances (F_1) [8], number of depressions (F_2) [8], lobulation index (F_3) [8], and dissimilarity (F_4) [9]. From the result, we can see that the performance of the classifier using the proposed features was much higher than that of other previous feature sets. In terms of accuracy, the proposed features increased accuracy to 15.8% for F_1, 5.43% for F_2, 17.32% for F_3, and 13.86% for F_4. In terms of sensitivity, the proposed features increased sensitivity to 5.96% for F_1, 5.6% for F_2, 37.29% for F_3, and 7.69% for F_4. Thus, it is demonstrated that the proposed features can classify malignant breast tumors with respect to the five objective indices more accurately.

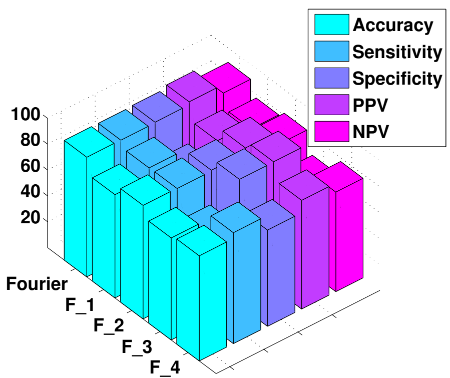


Fig. 4: Classification performance of breast tumor using five objective indices.

The relationship between sensitivity and specificity can be illustrated by the receiver operating characteristic (ROC) curve, which is a graphical plot of the TP rate, against the FP rate [14]. A ROC curve facilitates advanced analysis of the classification accuracy of a diagnostic method. The shape of the ROC curve can be determined by the area under the curve (AUC). Thus, the AUC can be used as a measure of test accuracy. An AUC of 1 represents a perfect test and an AUC of 0.5 represents a worthless test. An experimental result that gives a larger AUC indicates a better method than one with a smaller area. Fig. 5 shows the classification performance of the proposed features compared to the other features using AUC. As shown in the result, the AUC of the proposed features are 15.75%, 12.93%, 5.22%, and 4.04% higher than the other features such as F_1, F_2, F_3, and F_4. Therefore, the proposed technique has the greatest discrimination capacity.

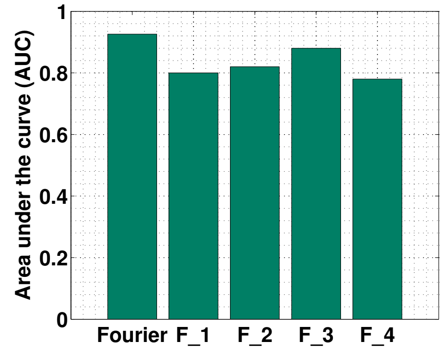


Fig. 5: Comparison of the AUC of the classification results from different feature sets.

V. Conclusions

In this paper, we propose a novel Fourier-based shape feature extraction technique which can provide a high accuracy rate of mass classification in the computer-aided B-mode ultrasound diagnosis of breast tumor. We also tried different kinds of features and compare the performance of each feature in classifying tumors. To demonstrate the performance of the proposed features, 4,107 ultrasound images containing 2,508 malignant cases was used. The experiments demonstrate that the proposed features can represent the shape irregularity of tumors in ultrasound images, result in the better classification performance compared to other boundary-based features in the spatial domain.

References

[1] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, and M. J. Thun, "Cancer statistics, 2011," CA: A Cancer Journal for Clinicians, vol. 61, no. 2, pp. 225-249, 2011.

[2] V. Jackson, R. Hendrick, S. Feig,and D. Kopans, "Imaging of the radiographically dense breast," Radiol., vol. 188, no. 2, pp. 297-301, 1993.

[3] G. J. Rizzatto, "Towards a more sophisticated use of breast ultrasound," Eur. Radiology, vol. 11, no. 12, pp. 2425-2435, 2001.

[4] K. G. Kim, S. W. Cho, S. J. Min, J. H. Kim, B. G. Min, and K. T. Bae, "Computerized scheme for assessing ultrasonographic features of breast masses," Acad. Radiol., vol. 12, no. 1, pp. 58-66, 2005.

[5] R. F. Chang, W. J. Wu, W. K. Moon and D. R. Chen, "Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors," Breast Cancer Res. Treat., vol. 89, no. 2, pp. 179-185, 2005.

[6] L. Papp, N. Zsoter, C. Loh, B. Ole, B. Egeler, I. Garai, and U. Luetzen, "Automated lymph node detection and classification on breast and prostate cancer SPECT-CT images," Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, pp. 3431-3434, 2011.

[7] K. G. Kim, S. W. Cho, S. J. Min, J. H. Kim, B. G. Min, and K. T. Tae, "Computerzied scheme for assessing ultrasonographic features of breast masses," Acad. Radiol., vol. 12, no. 1, pp. 58-66, 2005.

[8] C.-M. Chen, Y.-H. Chou, K.-C. Han, C.-M. Tiu, H.-J. Chiou, and S.-Y. Chiou, "Breast lesions on sonograms: Computer-aided diagnosis with nearly setting-independent features and artificial neural networks," Radiol., vol. 226, no. 2, pp. 504-514, 2003.

[9] W.-C. Shen, R.-F. Chang, W. K. Moon, Y.-H. Chou, and C.-S. Huang, "Breast ultrasound computer-aided diagnosis using BI-RADS features," Acad. Radiol., vol. 14, no. 8, pp. 928-939, 2007.

[10] Y. Mingqiang. K. Kidiyo, and R. Joseph, "A survey of shape feature extraction techniques," Pattern Recogn., pp. 43-90, 2008.

[11] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," IEEE Trans. Pattern Anal. Mach. Intell., VOL. 24, NO. 24, pp. 509-521, 2002.

[12] C. T. Zhan and R. Z. Roskies, "Fourier descriptors for plane closed curves," IEEE Trans. Computers, vol. C-21, no. 3, pp. 269-281, 1972.

[13] V. Vapnik, "Statistical learning theory," Wiley-Interscience; 1 edition, 1998.

[14] R. M. Rangayan, "Biomedical image analysis," CRC Press, 2005.