·技术与方法· doi:10.3969/j.issn.1671-8348.2014.30.022

Morphologic and texture features in classifying the malignant and benign breast nodules in ultrasonography*

Chen Qiuxia[1], *Xiang Jun*[1], *Liu Qi*[2], *Liu Jian*[1△]
(1. *Department of Equipment, the First Affiliated Hospital of Chongqing Medical University, Chongqing* 400016, *China*; 2. *College of Electrical Engineering and Information, Sichuan University, Chengdu, Sichuan* 610065, *China*)

Abstract: Objective To develop a computer-aided diagnosis(CAD) system with automatic contouring and morphologic and textural analysis to aid on the classification of breast nodules on ultrasound images. **Methods** A modified Level Set method was proposed to automatically segment the breast nodules(46 malignant and 60 benign nodules). Following, 16 morphologic features and 17 texture features from the extracted contour were calculated and principal component analysis(PCA) was applied to find the optimal feature vector dimensions. Fuzzy C-means classifier was utilized to identify the breast nodule as benign or malignant with selected principal vectors. **Results** The performance of morphologic features was 78.30% for accuracy, 67.39% for sensitivity and 86.67% for specificity, while the latter was 72.64%, 58.70% and 83.33%, respectively. After the combination of the two features, the result was exactly the same with the morphologic performance. **Conclusion** This system performs well in classifying the malignant breast nodule from the benign breast nodule.

Key words: computer-aided diagnosis; breast neoplasms; morphologic feature; texture feature

中图分类号:R319 **文献标识码**:A **文章编号**:1671-8348(2014)30-4046-04

Early diagnosis and treatment is the most effective way of reducing mortality caused by breast cancer. Ultrasonic diagnosis is currently widely used in assisting examination of breast nodule. The morphologic and textural variation between benign and malignant nodules is deemed a useful characteristic for their differentiation in ultrasound. Researchers have developed many advanced algorithms to extract the morphologic and texture features from breast nodule on ultrasound images, but the selection of the best features is still on debate. Therefore, it is desirable to develop computer-based classification methods to provide radiologists with a dependable second reading, hence to reduce the number of misdiagnosis.

Computer-aided diagnosis(CAD) has been largely studied in recent years. Joo et al[1]. designed a CAD algorithm identifying breast nodule malignancy using multiple ultrasonography features and artificial neural network classifier, 53.3% of biopsies on benign nodules can be avoided with 99.3% sensitivity. Kim KG et al. evaluated various spiculate and jagged margin shape features and they achieved a result of sensitivity 95.1% and specificity 97.2%[2]. Huang et al[3]. presented a CAD system to prove that textural features are valuable in nodule diagnosis. In this study, we aim at developing a CAD system with automatic contouring and morphologic and textural analysis to aid in the classification of breast nodules on ultrasound images. 16 morphologic features are calculated based on the nodule contour, while 17 texture features are calculated based on nodule region from the integrated spatial gray level co-occurrence matrix(GLCM). Then principal component analysis(PCA) is applied to reduce feature vector dimension and the principal vector is employed to distinguish between benign and malignant lesions by means of Fuzzy C-means(FCM). The FCM is a reliable choice due to its fast speed and excellent classification capability.

1 MATERIALS AND METHODS

1.1 Materials This study contains 106 ultrasonic breast nodule images arbitrarily selected from Department of Ultrasonography in West China Hospital, where 46 malignant and 60 benign. Before quantitative analysis, we employ a modified Level Set method to extract the breast nodule[4] on ultrasound image for morphologic and textural analysis.

1.2 Methods

1.2.1 Morphologic features analysis According to the clinical diagnosis, the shape of the nodules and the regularity of their contours are distinct in benign and malignant breast nodules[5]. A typical benign mass has an elliptical, smooth, and well-circumscribed boundary, whereas a malignant nodule usually has a speculated, rough, and blurry boundary with taller-than-wide shape[6-7]. In addition, area is an important feature for diagnosing nodules as large nodules usually have higher possibility to be malignant than small nodules. In the study, 16 practical morphologic features are extracted to classifiy breast nodules. These features are defined as follows[8-10].

(1) Area, Area feature is the actual number of pixels in the nodule region. Malignant nodules frequently have a larger area

* 基金项目:西华大学四川省信号与信息处理重实实验室基金项目(SGXZD0101-10-1)。 作者简介:陈秋霞(1985-),硕士,主要从事医学图像及信号处理。 △ 通信作者,E-mail:453375917@qq.com。

compared with benign ones.

(2) Perimeter, perimeter feature represents the length of the nodule perimeter. As malignant nodules usually have irregular shapes, a large nodule perimeter is associated with the likelihood that a nodule is malignant.

(3) $Form\_Factor = \frac{4\pi \times Area}{Perimeter^2}$

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

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

Where Max_Diameter denotes the length of the major axis from the equivalent ellipse of a nodule.

(5) $Solidity = \frac{Area}{Convex\_Area}$

The Solidity feature represents the proportion of the pixels in the convex hull that are also in the nodule region, where Convex_Area is the area of the convex hull of a nodule. When Solidity is close to 0, the nodule is malignant.

(6) $Convexity = \frac{Convex\_Perimeter}{Perimeter}$

Where Convex_Perimeter is the perimeter of the convex hull of a nodule.

(7) $Extent = \frac{Area}{Bounding\_Box}$

The Extent feature represents the proportion of the pixels in the bounding box that are also in the nodule region, where Bounding_Box is the smallest rectangle containing the nodule.

(8) TEP_Ratio. The TEP_Ratio is the perimeter ratio of a nodule 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 nodule to acquire the same area for the ellipse and nodule.

(9) TEP_Difference, TEP_Difference is defined as the difference between nodule perimeter and the corresponding ellipse.

(10) TCP_Ratio, TCP_Ratio is the perimeter ratio of a nodule and the corresponding circle which has the same area as the nodule.

(11) TCP_Difference, TCP_Difference is defined as the difference between the nodule perimeter and the corresponding circle.

(12) AP_Ratio, AP_Ratio is the ratio of the area and the perimeter of a nodule.

(13) Eccentricity, Eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1.

(14) Aspect_Ratio, Aspect_Ratio is the length ratio of a nodule's depth and width. If a nodule's depth exceeds its width, the Aspect_Ratio is greater than 1 and the nodule has a high probability of being malignant.

(15) LS_Ratio, LS_Ratio is the length ratio of the major axis and minor axis of the equivalent ellipse.

(16) Roughness, Roughness is represented by the entropy of the normalized radial length. A nodule with a larger value of the entropy indicates the more complexity of its boundary.

Chang et al[11] applied the Form_Factor, Roundness, Aspect_Ratio, Solidity, Convexity and Extent features to diagnose breast nodules. The other features are fundamental and clinically useful indicators for the classification of breast nodules.

1.2.2 Texture features analysis Nodules on ultrasound images have their own characteristics in image textures, so it can be viewed as criteria for discriminating between benign and malignant breast nodules. Texture features are defined as a statistical analysis of the spatial variation within adjacent pixel intensities of an image. In the study, we apply the integrated GLCM to extract texture features.

GLCM reflects some comprehensive information about image in the aspects of direction, interval, variation extent and variation speed[12]. With respect to the relationship between arbitrary two points, one should be concerned about the direction and the distance of two points. In the study, we take distance as one, and directions are 0°, 45°, 90° and 135° respectively, so that each direction corresponds to a GLCM and also each texture feature has four different values. In order to integrate texture features in each direction into a comprehensive one, weighted coefficients on each direction are calculated over contrast in that contrast has the maximal correlation. Then an integrated GLCM is obtained by summing the weighted GLCM in four directions, and under which 17 texture features are calculated. However, six texture features are dismissed due to their incompleteness and having little meaning for classification. The remaining 11 texture features are f1 - angular second-moment, f2 - contrast, f3 - correlation, f4 - variance, f5 - sum variance, f6 - entropy, f7 - sum entropy, f8 - difference entropy, f9 - information measure of correlation 1, f10 - information measure of correlation 2, f11 - inverse difference moment. The equations defining these features can be seen in[8]. According to their particular properties they are sorted into three groups as follows, group 1: f1, f6, f7, f8, f9 and f10 have the invariant property under monotonic gray-tone transformation; group 2: f1, f2, f3, f6 and f11 have the maximal correlation; group 3: f4 and f5 have greater discrepancy as the texture of image are different. Therefore these special texture features can be used as important indict for categorizing breast nodules.

1.2.3 Principal component analysis Principal component analysis (PCA) for vector dimension reduction. Using the high-dimensional vector directly is unsatisfactory when identifying breast nodules. Furthermore, numerous morphologic and texture features may be co-dependent. Only independent features should be utilized to attain reliable classification performance. It is well kwon that PCA is a usually adopted statistical analytical method that decreases redundancy by projecting the original data over an appropriate basis. PCA involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The idea behind the PCA is to create a more pertinent representation for reducing the dimensions of the original vectors. According to the experiment, the ideal dimension is 4; thus, each original 16-dimensional morphologic and 11-dimensional texture feature vector is condensed by PCA into a new 4-dimensional feature vector, hence to find the best classification performance.

1.2.4 Classification (Fuzzy C-means) The Fuzzy C-means clustering algorithm is based on the minimization of an objective function called C-means functional[13]. It is defined by Dunn as:

$$J(X;U,V)=\sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^{m}D_{ikA}{}^{2} \qquad (1)$$

where

$$D_{ikA}=\|x_k-v_i\|_A=(x_k-v_i)^{T}A(x_k-v_i)$$

is a squared inner-product distance norm.

The minimization problem can be solved by means of Lagrange multipliers. If $D_{ikA}{}^2>0$, $\forall\, i,k$, and $m>1$, then (1) may minimize only if

$$u_{ik}=\frac{1}{\sum_{j=1}^{C}(D_{ikA}/D_{jkA})^{2/(m-1)}} \qquad (2)$$

where $(1\leq i\leq C, 1\leq k\leq N)$

and

$$v_i=\frac{\sum_{k=1}^{N}\mu_{ik}{}^{m}x_k}{\sum_{k=1}^{N}\mu_{ik}{}^{m}}, 1\leq i\leq C \qquad (3)$$

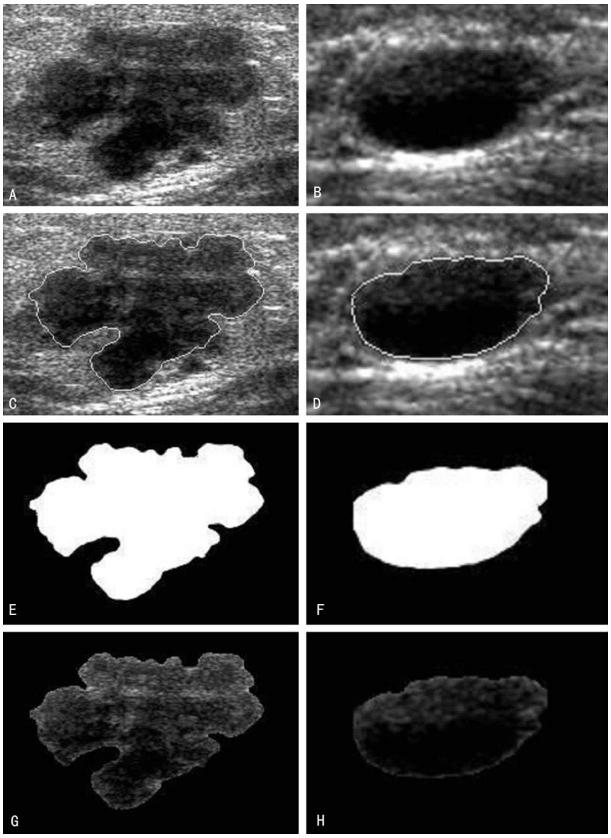
where $X=\{x_i, x_2, \cdots, x_n\}$ is the data setc, is the number of clusters, $m\in(1,\infty)$ is a weighted index (which is usually assigned by 2), v_i is the clustering center, μ_{ik} is the fuzzy membership of whose clustering center is v_i, D_{ik} is the standard Euclidean distance from x_k to v_i.

2 RESULTS

2.1 Image segmentation The original and segmented breast nodule images are presented in Fig. 1.

2.2 Classification results As for texture features, previous experiments have demonstrated the effectiveness of correlation, sum variance, information measurement of correlation 1 and information measurement of correlation 2 for classifying breast nodules, producing a result of accuracy 72.64%, sensitivity 58.70% and specificity 83.33%. Both of information measurement of correlation 1 and information measurement of correlation 2 describe the linear dependency of the intensity. The correlation reflects the correlation of local intensity in images, indicating the textural directionality. The sum variance reflects the variation speed and periodicity of the texture which varies greatly with different texture of the image. So the four texture features can be used as important indexes for discriminating breast nodules.

According to the morphologic characteristic difference between benign and malignant nodules, in various combinations of morphologic features, the one containing TEP_Ratio, TEP_Difference, AP_Ratio and Aspect_Ratio obtains the best result with accuracy 78.30%, sensitivity 67.39% and specificity 86.67%, which is slightly better than that of texture features. TEP_Ratio and TEP_Difference are concerned about the ellipticity of the nodule contour which is the significant characteristic of benign breast nodules, whereas Aspect_Ratio is the length ratio of a nodule's depth and width, corresponding to the typical characteristic of malignant breast nodules which usually has taller-than-wide shape. In addition, AP_Ratio is related to a nodule's area and perimeter. Since large nodules usually have higher possibility to be malignant, this parameter is also an important feature for diagnosing nodules. Nevertheless, taking the combination of morphologic and texture features as input feature vector has not given any higher accuracy, and the best performance is exactly the same with the morphologic features do. All of the results are shown in Table 1.



A、B: the original image; C、D: the segmented image; E、F: the nodule contour for calculating morphologic features; G、H: the nodule region for calculating texture features.

Fig. 1 Tipical case diagram of the malignant breast nodule (left) and the benign breast nodule (right).

Table 1 Performance of the proposed CAD system (%)

Feature	Accuracy	Sensitivity	Specificity	PPV	NPV
Morphologic	78.30	67.39	86.67	79.49	77.61
Texture	72.64	58.70	83.33	72.97	72.46
Combination	78.30	67.39	86.67	79.49	77.61

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).

3 DISCUSSION

We have presented in the study a CAD system with a computationally quick procedure for segmenting the breast nodule on ultrasound images, extracting morphologic and texture features, and identifying the most effective features for categorizing the benign and the malignant breast nodules.

The results of 78.30% accuracy rate for morphologic features and 72.64% accuracy rate for texture features are promising, but the overall accuracy diagnosis is still not high enough to rely on confidently which is ascribed to some limitations on both features. As for morphologic features, some benign lesions may have a speculated appearance or blurred periphery, and round and well-defined malignant lesions do exist, which attributes to the misclassification of breast nodules. In addition, the segmentation process may lead to inaccuracy in the boundary detection, which causes the deteriorated performance of selected features. On the other hand, texture feature calculation has sensibility to nonlinear variations with ultrasound system setting, the time-gain compensation and focal depth. Furthermore, texture features mainly describe internal echo structure and the architecture of surrounding tissue, which leads to their highly sensitivity to the noise in ultrasound images, so the classification only relying on texture features are not reliable enough.

In the study, the false positive fraction in three groups is lower than 30%, and correspondingly, the true positive fraction is higher than 70%, which can prevent unnecessary biopsies. However the selected features may be tuned to the specific cases in the study, we need to integrate additional features, such as clump thickness, uniformity of cell size, uniformity of cell shape, bare nuclei, bland chromatin and mitosis, et al., along with additional patient history data, to further improve the accuracy of diagnosis. Also, the proposed CAD system is required to test on a larger data set.

In summary, an effective CAD system based on an optimal feature selection proposed in the study, which can be used as the "second opinion" for radiologists to make more accurate diagnosis of breast nodules. In the move toward more advanced CAD system for breast cancer detection, many techniques have to be improved: image processing algorithms, calculation of morphologic and texture features with excellent discriminatory ability and classification methods. Thus, enhancing the overall performance of CAD system still remains to be a key issue for future research and development.

REFERENCE:

[1] Joo S, Yang YS, Moon WK, et al. Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features[J]. IEEE Trans Med Imaging, 2004, 23(10): 1292-1230.

[2] Kim KG, Kim JH, Min BG. Classification of malignant and benign nodules using boundary characteristics in breast ultrasonogramsment[J]. J Digit Imaging, 2002, 15 Suppl 1: 224-227.

[3] Huang YL, Lin SH, Chen DR. Computer-aided diagnosis applied to 3-D US of solid breast nodules by using principal component analysis and image retrieval[J]. Conf Proc IEEE Eng Med Biol Soc, 2005, 2: 1802-1805.

[4] Li CM, Xu CY, Gui CF, et al. Level set evolution without re-initialization: a new variational formulation[J]. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, 2005, 1: 430-436.

[5] Nie K, Chen JH, Yu HJ, et al. Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI[J]. Acad Radiol, 2008, 15(12): 1513-1525.

[6] Guliato D, Rangayyan RM, Carvalho JD, et al. Polygonal modeling of contours of breast nodules with the preservation of spicules[J]. IEEE Trans Biomed Eng, 2008, 55 (1): 14-20.

[7] Tsui PH, Liao YY, Chang CC, et al. Classification of benign and malignant breast nodules by 2-D analysis based on contour description and scatterer characterization[J]. IEEE Trans Med Imaging, 2010, 29(2): 513-522.

[8] Yang W, Zhang S, Chen Y, et al. Shape symmetry analysis of breast nodules on ultrasound images[J]. Comput Biol Med, 2009, 39(3): 231-238.

[9] Chen CY, Chiou HJ, Chou SY, et al. Computer-aided diagnosis of soft-tissue nodules using sonographic morphologic and texture features[J]. Acad Radiol, 2009, 16(12): 1531-1538.

[10] Huang YL, Chen DR, Jiang YR, et al. Computer-aided diagnosis using morphological features for claasifying breast lesions on ultrasound[J]. Ultrasound Obstet Gynecol, 2008, 32 (4): 565-572.

[11] Chang RF, Wu WJ, Moon WK, et al. Automatic ultrasound segmentation and morphology based diagnosis of solid breast nodules[J]. Breast Cancer Res Treat, 2005, 89(2): 179-185.

[12] Liao YY, Tsui PH, Li CH, et al. Classification of scattering media within benign and malignant breast nodules on ultrasound texture-feature-based and Nakagami-parameter images[J]. Med Phys, 2011, 38(4): 2198-2207.

[13] Wang H, Fei B. A modified fuzzy C-means classification method using a multiscale diffusion filtering scheme[J]. Med Image Anal, 2009, 13(2): 193-202.

(收稿日期:2014-05-26　修回日期:2014-06-17)