Sonography Images for Breast Cancer Texture classification in Diagnosis of Malignant or Benign Tumors

P. Babaghorbani
Young Researches Club
Tehran, Iran

S. Parvaneh
Islamic Azad University, Science and Research branch
Tehran, Iran

AR. Ghassemi
Department of Diagnostic Imaging, Tehran University,
Tehran, Iran

K. Manshai
Islamic Azad University, Science and Research branch
Tehran, Iran

Abstract— This work aims at selecting useful features in critical angles and distances by Gray Level Co-occurrence Matrix (GLCM). In this project, images were labeled based on physician opinion in two groups (malignant or benign). These labeled images were used in classification analysis. Images were opened and read in Matlab software. The tumors were cropped in rectangular shape manually; then graycomatrix and GLCM have been calculated in 4 angels (0, 45, 90 and 135 degree) and 4 distances (1, 2, 3 and 4) for cropped tumor images. Since each angle and distance pair include 22 features, each image had 352 final features (22 features * 4 angles * 4 distances =352). At the final step, features were classified using Kmeans method into 2 classes of malignant and benign; then the confusion matrix was made and qualitative comparison was used to select important features and critical distances and angles in each one. Some special features, angels and distances which had the best classification result and high percentages of accuracy were selected as useful features. These finding suggested that texture parameters can be useful to help in distinguishing between malignant and benign breast tumors.

Keywords-Breast Sonography; Texture; Gray Level Co-occurrence Matrix

I. INTRODUCTION

Breast sonography is one of the first steps in the breast cancer diagnosis in women under 35. It is one of the best methods to find a tumor and the type of the tumor which can be malignant or benign. Diagnostic ultrasound is a useful clinical device for imaging of the human soft tissue. Some features in sonography such as shape, D/W (depth/width), tumor margin, shadow, etc. could separate malignant tumors from benign ones [1].

Many researchers have been done in this field. Ke Nie and colleagues evaluated MRI images for breast cancer analysis. They cropped pictures automatically and used 8 morphologic and 10 GLCM texture features. They utilized angle 0 and distance 1 for their GLCM features [2]. Weijie Chen and colleagues utilized contrast-enhanced MRI images for breast analysis and automatically segmented 3D breast lesions. They used nondirectional GLCM (addition of 13 directional GLCMs) and investigate extracted texture features by ROC analysis [3]. Andre Victor et al. extracted five parameters from the complexity curve (CC) and other five parameters from the GLCM in ultrasound images. These features calculated in a rectangular region of interest (ROI) containing the tumor and its neighborhood and internal tumor region. The most significant individual parameters were the contrast from the GLCM over the ROI and the maximum value from the CC for the tumor internal region. Their results propose that the texture parameters can be useful in distinguishing between benign or malign breast tumors on ultrasound images. They didn't focus on angles and distances in their work [4].

K. Holli et al. considered features provided by MaZda texture application. Their aim was to discriminate between patients, different imaging series and two histological types of carcinomas (ductal vs. lobular). Their result shown that the textures in every imaging series are different when non-cancer and cancer tissue compared and also between the two histological groups [5]. Jie Wu et al. used 2D Co-occurrence texture features in combination with Gabor texture features and both 2D and 3D Semi-variogram texture features for segmentation of abdominal MRI images. In their work, a GLCM is computed for a 3*3 neighborhood of each pixel (instead of the whole image). They calculated 14 statistics of GLCM in 4 directions (=0, 45, 90, 135) and one distance (=1) in order to find pair of pixels [6].

In this project, we studied GLCM matrix in 4 angles and 4 distances. We extract 22 features in each angle and each distance by GLCM toolbox in Matlab and analyzed extracted features independent of other features.

The main aim of this work is to choose superior individual textural features calculated from GLCM in classification of malignant or benign breast tumor.

978-1-4244-4713-8/10/$25.00 ©2010 IEEE

II. METHODS

A. Database

Twenty five benign and fifteen malignant lesions were used on this study, which gathered from Imam Khomeini Hospital in Tehran using Siemens sonography set in 13MHz frequency. The format of images was jpeg.

B. Software and computer used for analysis

We used Matlab version 7.8 for our analysis and used a computer with Intel® Core™2 duo 2 GHz CPU and 2 Gigabyte of memory.

C. Reading and pre-processing of images

Images were opened using Matlab software and detected tumors have been cropped manually. Figure 1 shows a sample of the original and cropped image.

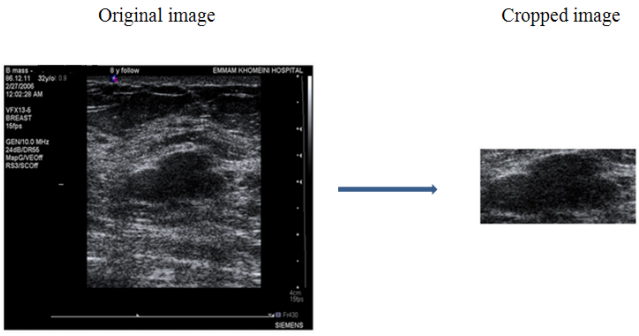

Figure 1. Sample of Original and Cropped Image

D. Texture analysis and feature extraction

Texture is an alteration and variation of data in small scale. Texture analysis calculates sensitive qualities such as rough, smooth or silky as a function of the spatial variation in pixel intensities [7, 8].

At the first step, we considered and plotted histogram of images, but could not obtain useful information because positional information was not considered. So we decided to use Graycomatrix and extract features from that.

GLCM calculates the probability of a pixel with the gray-level value i occurring in a specific spatial relationship to a pixel with the value j. The number of gray levels in the image determines the size of the GLCM. So, we have 256*256 matrices [7, 8]. We considered angle and distance as the main two parameters in our research.

Although there is a function in Matlab Image Processing toolbox that computes four parameters Contrast, Correlation, Energy, and Homogeneity, the paper by Haralick suggests a few more parameters that are also computed here [9]. It is easy to add new features based on the GLCM using this code.

By the GLCM toolbox 22 features can be obtained, which are listed below:

Autoc : Autocorrelation

contr : Contrast

corm : Correlation Matlab

corrp : Correlation [1, 2]

cprom : Cluster Prominence

cshad : Cluster Shade

dissi : Dissimilarity

energy : Energy

entro : Entropy

homom : Homogeneity Matlab

homop : Homogeneity [2]

maxpr : Maximum Probability

sosvh : Sum of Squares Variance

savgh : Sum Average

svarh : Sum Variance

senth : Sum Entropy

dvarh : Difference Variance

denth : Difference Entropy

inf1h : Information Measure of Correlation 1

inf2h : Information Measure of Correlation 2

Indnc : Inverse Difference Normalized (INN)

Idmnc : Inverse Difference Moment Normalized

We calculated co-occurrence matrix in 4 angles (0, 45, 90 and 135) and 4 distances (1, 2, 3 and 4) and extracted mentioned features in each angle/distance pair. Figure 2 shows the block diagram of steps of evaluation and comparison.

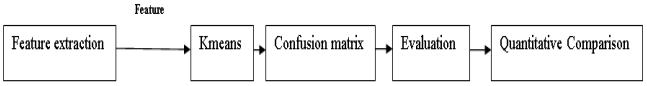

Figure 2. Block diagram of steps of evaluation and comparison

Some main features that could be defined and calculated are mentioned below:

1) Contrast

It returns a measure of the power difference between a pixel and its neighbor over the total image.

$$\sum_{i,j} |i-j|^2 \, p(i,j)$$

2) Correlation

It returns a measure of how linked a pixel is to its neighbor over the whole image [7].

$$\sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j}$$

3) Energy

It shows the sum of squared elements in the GLCM [7].

$$\sum_{i,j} p(i,j)^2$$

4) Homogeneity

It returns a value that measures the closeness of the division of elements in the GLCM to the GLCM crosswise [7].

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|}$$

TABLE I. FEATURES AND EQUATIONS

Features	Equations
Angular Second Moment	$\sum_i \sum_j \boldsymbol{p(i,j)^2}$
Sum of Squares Variance	$\sum_i \sum_j (i-\mu)^2\ p(i,j)$
Inverse Difference Moment	$\sum_1 \sum_j \frac{1}{1+(i-j)^2}\ p(i,j)$
Sum Average	$\sum_{i=2}^{2Ng} i\ Px+y(i)$ x : row y : column Px+y : probability of co-occurance matrix coordinates summing x+y
Sum Variance	$\sum_{i=2}^{2Ng} (i-f\delta)^2\ Px+y(i)$
Sum Entropy	$-\sum_{i=2}^{2Ng} Px+y(i)\ log\{Px+y(i)\} = f\delta$
Diffrence Variance	$\sum_{i=0}^{Ng-1} i^2\ Px-y(i)$
Diffrence Entropy	$-\sum_{i=0}^{Ng-1} Px-y(i)\log\{Px-y(i)\}$
Info. Measure of Correlation 1	$\frac{HXY-HXY1}{max\{HX,HY\}}$
Info. Measure of Correlation 2	$(1-exp[-2(HXY2-HXY)])^{\frac{1}{2}}$ HXY= $-\sum_i \sum_j p(i,j)log(p(i,j))$ HXY1= $-\sum_i \sum_j p(i,j)log(p(i)p(j))$ HXY2= $-\sum_i \sum_j p(i)p(j)log(p(i)p(j))$

5) Entropy

It is a scalar value representing the irregularity of grayscale image that can be used to characterize the texture of the input image [7].

$$-\sum_i \sum_j p(i,j)\log(p(i,j))$$

Table I shows some other features and their equations that count them [9].

E. Feature Evaluation

Selecting useful features is important in breast cancer analysis (malignant/benign classification). In this research, we used feature evaluation to get to our goal. We supplied a matrix with 40*352 dimension, which means 40 images and 352 features (4angles*4distances*22features) that each row of the matrix represents one feature in specific angle and distance. Then we organized the matrix by Kmeans method to separate 2 groups of malignant and benign tumors.

We made a 2*2 confusion matrix (con); position (1, 1) shows the true classification of benign, position (2, 2) represents the true classification of malignant, position (1, 2) wrongly distinguishes malignant instead of benign and position (2, 1) wrongly distinguishes benign instead of malignant tumors. From the confusion matrix following definitions have been defined:

Error = (100/40) * (con (1, 2) +con (2, 1))

Accuracy = 100 - Error

We calculated Error and Accuracy for each angle and distance pair and selected high percentages of accuracy as useful features. Table II shows the confusion matrix and its cell definition.

TABLE II. CONFUSION MATRIX

Confusion Matrix	
position (1, 1) shows the true classification of benign	***position (1, 2) wrongly distinguishes malignant instead of benign***
position (2, 1) wrongly distinguishes benign instead of malignant	***position (2, 2) represents the true classification of malignant***

III. RESULT

In this paper, we analyzed extracted features independently from other angles and distances (other features).

66 features were selected as useful features, which had high accuracy (up to 85%). The following features are useful in specific angles:

- Autocorrelation in angle 135, in distances 1, 2 and 3.
- Correlation in angle 90, in distances 1, 2 and 3.
- Dissimilarity in angle 45, in distances 1, 2 and 3.
- Homogeneity in angle 0, in distances 1, 2 and 3.
- Maximum Probability in angle 135, in distances 1, 2 and 3.
- Sum Average in angle 90, in distances 1, 2 and 3.

- Difference Entropy in angle 45, in distances 1, 2 and 3.
- Inverse Difference Normalized in angle 0, in distances 1, 2 and 3.

Table III shows the number of important features in each angle and distance. From the last column, we could conclude that in odd distances we had more informative features. In the last row, there was the nearly equal number of high accuracy features in each angle, so there was low sensitivity about angles in classification. Sum of important features in all survived angles/ distances was 66.

TABLE III. ANGLES IN SPECIAL DISTANCE

Distance/Angles	*0*	*45*	*90*	*135*	*Total*
1	6	5	6	6	23
2	2	3	2	2	9
3	6	6	6	6	24
4	2	2	3	3	10
Total	16	16	17	17	66

IV. CONCLUSION

In this study sonography, images have been used and then tumors were cropped manually. Features were extracted under GLCM matrix in 4 angles and 4 distances. They were classified by Kmeans and based on calculated confusion matrix, superior angels, distances and features in texture analysis of malignant and benign tumors have been identified. As described in result section, some features such as autocorrelation, correlation, dissimilarity, homogeneity, maximum probability, sum average, difference entropy and inverse difference normalized in distance 1, 2 and 3 have high accuracy in different angles. We found useful features in most of the distances and angles but we could not obtain a very high sensitivity by assessing specific angles only. We also found useful results by evaluating only distances, especially distance 1 and 3 without considering angles.

REFERENCES

[1] M. Giti, et al., Principles of Diagnostic and Cure Maladies of Breast, 2002 (In Persian).

[2] K. Nie, et al., "Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI," Academic Radiology, vol. 15, pp. 1513-1525, 2008.

[3] W. Chen, et al., "Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images," Magnetic Resonance in Medicine, vol. 58, pp. 562-571, 2007.

[4] A. Alvarenga, et al., "Complexity curve and grey level co-occurrence matrix in the texture evaluation of breast tumor on ultrasound images," Medical Physics, vol. 34, p. 379, 2007.

[5] K. Holli, et al., "Detection of characteristic texture parameters in breast MRI," 2008.

[6] J. Wu, et al., "Texture Feature based Automated Seeded Region Growing in Abdominal MRI Segmentation," 2008, pp. 263-267.

[7] "Matlab Software Help"

[8] The GLCM Tutorial Home Page, *http://www.fp.ucalgary.ca/mhallbey/tutorial.htm*.

[9] A. Uppuluri. GLCM Texture Features, *http://www.mathworks.com/matlabcentral/fileexchange/22187-glcm-texture-features*.