2010 3rd International Conference on Computer and Electrical Engineering (ICCEE 2010)
IPCSIT vol. 53 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V53.No.2.97

Molybdenum-Ray Image Recognition of Breast Cancer Based on Fractal Texture Analysis

Yang Dong-Feng+

Department of Information Technology, Heilongjiang Bayi Agriculture University

Daqing, China

Abstract: A kind of computer-aided diagnosis method for mammary gland molybdenum palladium X-ray image is put forward. Using fuzzy histogram equalization algorithm to improve the original image contrast in order to extract the interested region and strengthen regional edge. In order to accurately extract lesions the improved fuzzy clustering algorithm is put forward. Fractal box counting dimension is used to describe the texture characteristics of lesions which have been segmented. According to the shape and gray distribution of 10 forms of tumour and calcification spot, 6 characteristic parameters are extracted. A BP neural network of three layers is established. After training on 244 samples, 238 samples are used to validate the accuracy of the system, the result shows average accuracy rate is 96.59%.

Keywords-fuzzy clustering;computer aided diagnosis;fractal dimension;texture analysis; pattern recognition

1. Introduction

Breast cancer is one of the most common malignant tumor of women. According to statistics, every year about 120 million women suffer from breast cancer and 500,000 women die of breast cancer. In the United States, 182,000 cases are diagnosed each year with breast cancer and of which 46,000 cases die. In china, the incidence of breast cancer is rising sharply by a growth rate of 3% ~ 4%. Each year, about 180,000 women suffer from breast cancer and 13,000 women die of breast cancer. Most medical experts believe that successful treatment of breast cancer is closely related with the early diagnosis.

Molybdenum-ray photography is the most important clinical means of detecting breast cancer. According to data, only 70~85 percent of breast cancer cases can be detected by medical experts through its breast X-ray image at the first time. In the rest of the cases, only 2/3 can be tested in the second time. This is mainly due to the benign representation of malignant lesions and too small lesions and visual fatigue of medical experts. So it is necessary to help them by using computer-aided diagnosis.

At present, in the field of intelligent recognition of breast cancer, fruitful results have been produced. Li XiaoFeng proposed an ultrasound breast tumor image computer-aided diagnosis system based on anisotropic diffusion and support vector machines which is given by [1]. Yang Wei put forward a method to classify X-ray images of breast cancer based on hidden markov tree model in wavelet domain which is given by [2]. Hao Xin put forward a medical diagnostic technique based on image content retrieval which is given by [3]. Lu Wen used fuzzy genetic neural network to extract breast microcalcifications that is very important for breast lesion recognition which is given by [4]. Wang ShuYan studied the fuzzy clustering analysis in medical image data mining application and made use of decision tree algorithm to classify image data of breast cancer which is given by [5]. Antonie used neural network and the association rule mining method to classify

+ Corresponding author.
E-mail address: yangsansun@sina.com

mammary gland X-ray image which is given by [6]. A.E.Hassanien used decision tree to do classification and feature extraction on mammary gland X-ray image which is given by [7].

Fuzzy image enhancement algorithm and fuzzy clustering analysis are used to enhance and segment the X-ray image of mammary gland molybdenum palladium in this paper. On texture analysis of lesions, advantages for fractal analysis on irregular random texture are made full use and the fractal dimension that is extracted together with seven other features such as energy, entropy of gray level co-occurrence matrix forms a vector of neural network.

2. Fuzzy Image Enhancement

We use fuzzy histogram hyperbolization algorithm (FHH) to enhance the image contrast, extract the region of interest and enhance the interested region.

A. Sub-block partition

Define attribute set of conditions C={c1,c2}, c1 is the pixel gray value attribute and c2 is the noise properties. X-ray image is commonly composed of lighter area and darker area, so it has two peaks in its histogram, one peak area corresponds to lighter gray value and the other peak corresponds to the dark gray value area. In between two peaks, choose a gray value as threshold P. Gray value attribute cl={0,1},0 represents the gray value of 0 ~ P and 1 represents the (P +1) ~ 255 gray scale. Noise property c2={0,1}, 0 represents the absolute difference value between the average gray value sub-block s and the average gray value of adjacent sub-block is less than a threshold Q and 1 represents the absolute difference value between them is larger than Q.

1) Mark the molecule picture according to c1

Set x represents "lighter" pixels, equivalence relation Rc1 defined as: if gray values of two pixels are more than a certain threshold P, then the two pixels are Rc1 related, that is:

$$R_{c1}(x) \mid \{x : f(x) > P\}$$

F(x) represents pixel gray value of x, Rc1(x) represents set of all the lighter pixel x.

2) Mark the molecule picture according to c2

Rc2 is defined as the equivalence relation: the rounding values of absolute difference of average gray value m(s) between sub-block sij and adjacent blocks are greater than a threshold Q, that is:

$$R_{c2}(s) = \bigcup_i \bigcup_j \{s_{ij} \mid : \text{int} \mid m(s_{ij}) - m(s_{i\pm1,j\pm1}) \mid > Q\} \tag{1}$$

Rc2 (s) represents the set of all the noise pixels, order:

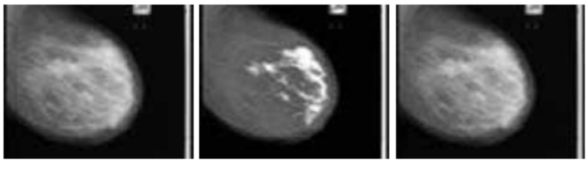
$$A_1 = R_{c1}(x) - R_{c2}(s) \; A_2 = \overline{R_{c1}}(x) - R_{c2}(s) \tag{2}$$

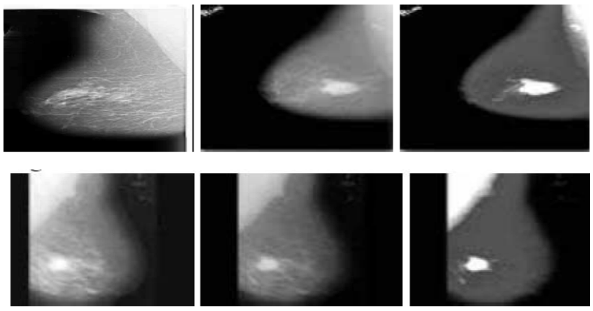
A1 is the set of all lighter pixel x after removing noise, A2 is the set of all darker pixel x after removing noise.

B. Do strengthen transform

- Fill the graph A1, that is: using gray value of threshold P and average gray value of macroblock of noise sub-block to fill the position of all the darker pixels and noise pixels which constitute the B1.
- Fill the graph A2, that is: using gray value of threshold P and average gray value of macroblock of noise sub-block to fill the position of all the brighter pixels and noise pixels which constitute the B2.
- Do histogram equalization for B1 and do histogram exponential transform for B2.
- Overlap the transformed image and output the enhanced image.

Chose some images from a X-ray image database in a hospital and use FHH algorithm to process them, the result shows in Fig.1.





(a)Original image (b) HE enhancement (c) FHH enhancement

Fig.1 Fuzzy image enhancement

3. Fuzzy clustering image segmentation

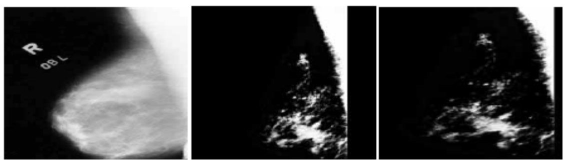
Most edges of benign tumors of are clear and regular, while the malignant tumor edges are blur and irregular. However, the boundary of some benign tumors such as cyst block and fibrous tumor is blurred too. Therefore, the tumor region segmentation is the key of further accurate classification between benign and malignant tumors. The standard fuzzy C-mean (S-FCM) clustering algorithm is as follows:

The standard fuzzy c-means objective function for partitioning data set $Y=\{y1, y2, \cdots, yn\}$ into c clusters is given by [8,9]:

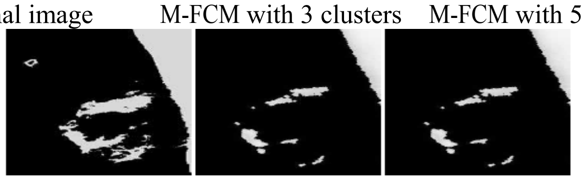
$$J = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^{p} \| y_k - v_i \|^2 \qquad (3)$$

where v_i are the prototypes of the clusters and the array $u_{ik} = U$ represents a partition matrix. The parameter p is a weight exponent determining the amount of fuzziness in the resulting classification. The FCM algorithm is computationally expensive and sensitive to the noise. To solve these problems, we present a modified version of the fuzzy c-mean clustering algorithm which is given by [10].

Figure.2 represents S-FCM and M-FCM visual clustering results with different initiation parameters. The weight parameter of cost function is ranging from 0.001to 0.0000001 and the clusters' number ranging from three to eight. From the obtained results, we observe that both algorithms get the same good results with higher number of clusters and small weight of the cost function. The average of segmentation accuracy of the introduced algorithm is about 3.9837% error, which means that it is robust enough.



Original image M-FCM with 3 clusters M-FCM with 5 clusters



M-FCM with 6 clusters M-FCM with 7clusters S-FCM with 7 clusters

Fig.2. FCM Segmentation results with different number of clusters

4. Fractal Texture Analysis and Feature Extraction

Breast X-ray images contain a lot of information including various tissues, glands, ducts, breast edge, tumour, calcification, etc. Tumour and calcification are the key analysis objects for identification of breast cancer.

TABLE 1 FORMS OF LESIONS

Main Lesions	Shape
Tumour	Conglomeration Shape
	Star-Shaped Spikes
	Cloud Sheet
	Hemispherical Shape

	Comet Shape
Calcification	Round
	Oval
	Irregular Polygon Shape
	Linear
	Bifurcation Shape

These forms are empirical conclusions drawn by medical experts through carefully observing. In order to accurately identify breast cancer, texture and shape features must be extracted of these forms.

Texture is the spatial distribution description of gray scale of an image. There are two types of texture one of which is regular structured texture with deterministic cycle and another is a random-type texture with random structure. Tumour and calcification images have randomness texture feature in grayscale distribution. Fractal is a kind of mathematical model suitable of describing complex and irregular shape of object. For the image texture analysis, fractal dimension is a quantitative description of fractal geometry characteristic and the complexity parameters of fractal geometry, which is better to characterize the rough degree of texture, and it is not sensitive to changes in the scales, this is very similar to the human visual system (HVS). Therefore, the fractal dimension is often used to describe the texture characteristics.

Box dimension is the most widely used one of the fractal dimensions. Let F be any non-empty set. N_δ (F)expresses the minimum number of the set whose maximum diameter is δ which can cover set F.

The upper and lower box dimension of F is defined as:

$$\underline{Dim_B(F)} = \lim_{\delta \to 0} \inf \frac{\log N_\delta(F)}{-\log \delta} \tag{4}$$

$$\overline{Dim_B(F)} = \lim_{\delta \to 0} \inf \frac{\log N_\delta(F)}{-\log \delta} \tag{5}$$

If they are equal, the value is called the box dimension.

$$Dim_B(F) = \lim_{\delta \to 0} \inf \frac{\log N_\delta(F)}{-\log \delta} \tag{6}$$

As a result, the minimum number that can cover set F whose diameter is δ is about δ^{-5}.

$$s = Dim_{(B)}(F) \tag{7}$$

Sarkar and Chaudhuri put forward an effective calculation method of fractal dimension, called the differential box method (DBC) which is given by [11]. Consider an $M \times M$ pixel image, it is divided into $s \times s$ pixels grids, so the proportion is $r = s / M$. In each grid, there is a box column whose size is $s \times s \times s'$. Numbers the box in by 1, 2,…, n. If in the grid (i, j), the minimum gray value of the image falls in the box whose number is k, the maximum gray value of the image falls in the box whose number is l, then, nr(i, j) = l - k + 1 which is the part of Nr in the (i, j) grid. The sum of all parts of the grid is:

$$N_r = \sum_{i,j} n_r(i, j) \tag{8}$$

For different values of r and s, Nr should be calculated separately. And then calculate log(Nr) and log(1/r), using the least square method to fit them can acquire box dimension.

According to the digital image box counting dimension calculation method, compute box counting dimension of various forms of tumour and calcification. According to the principles that window should include some basic parts of textures, select images whose size is 16×16. Table 2 shows the average fractal dimension of various forms of image segmentation.

TABLE 2 FRACTAL DIMENSION FOR 10 FORMS OF LESIONS

Tumor	**Conglome -ration**	**Star-Shaped Spikes**	**Cloud Sheet**	**Hemisph-erical**	**Comet**
FD	2.2134	2.3225	2.2724	2.2578	2.3864

TABLE 3 FRACTAL DIMENSION FOR 10 FORMS OF LESIONS

Calcificatin	**Round**	**Oval**	**Irregular Polygon**	**Linear**	**Bifurcation**
FD	2.1823	2.1764	2.0562	1.4513	1.6238

As can be seen from Table 2 and Table 3, fractal dimension of each form of lesions is different. But the box counting dimension from different mammography images which has the same kind of lesions is different also. Thus, box counting dimension can not be used to distinguish various forms of lesions alone. Fractal dimension must be combined with other features. After a great deal of analysis and experiment six other parameters is obtained which has great influence on identification of lesions.

(1) Shape parameters F: F=B2/4 π A. B is the circumference of the lesions.

(2) Ratio of orthogonal long axis C: The ratio between the longest axis of segmentation region and the longest axis in vertical direction.

(3) Roundness S: Take the region centroid as the centre of a circle, to do the inscribed circle and circumcircle of the region, ratio between radius of inscribed circle ri and the circumcircle radius rc, that is S=ri/rc.

(4) Energy E: is a measure of consistency of image texture. The gray level co-occurrence matrix of the region is p (i, j), so:

$$E(d,\theta) = \sum_{i,j} \{p(i,j) \mid d,\theta\}^2 \tag{9}$$

(5) Entropy H: it measures the disorder of an image and it achieves its largest value when all elements in normalized entry of the co-occurrence matrix are equal. When the image is not texturally uniform many GLCM elements have very small values, which implies that entropy is very large. Therefore, entropy is inversely proportional to GLCM energy.

$$H(d,\theta) = -\sum_{i,j} \{p(i,j \mid d,\theta\} - \log\{p(i,j \mid d,\theta\} \tag{10}$$

(6) Moment of inertia I: Measure the homogeneity of an image.

$$I(d,\theta) = \sum (i-j)^2 P(i,j \mid d,\theta) \tag{11}$$

5. Design of neural network classifiers

Artificial neural network is topology that can organically link the input and output through neurons. There is a nonlinear mapping between the input unit and output units so it has a function of a classifier. Compared with the traditional classifier such as statistical classification, it has many advantages including fault-tolerant ability, large-scale parallel processing capabilities, adaptive self-learning ability, nonlinear dynamic characteristics etc. which is given by [12].

The main function of classification is to identify tumour and calcification of the 10 lesions forms. Take a 3-layer BP neural network with one hidden layer. Number of neurons in input layer is decided by the size of the dimension feature parameters. Parameters are the fractal dimension D, shape parameters F, ratio of orthogonal long axis C, roundness S, energy E, entropy H and moment of inertia I, which constitute the feature vector. Output layer has 10 nodes, corresponding to the 10 forms of lesions. For example, for some input vector, the output is tumour lesions with conglomeration shape, so, the first node of output is 1 and others are 0.

For the decision of the number of hidden nodes can not be determined by theoretical method so far, can only be estimated by experience or experiment. Based on the past experience, take the 14 hidden nodes in our research.

S-function is the best choice for the activation function Φ (.),such as logisic() function, hyperbolic tangent function. S-function generally has smooth, differentiable, nonlinear, and saturation characteristics, and the derivative Φ '(.)is easy to use φ(•)itself to express. In this study, the hidden layer neuron activation function select logsic() function, the output layer activation function select pureline() function.

After several tests, the system training step is 1000, every 10 steps display 1 time, the target value of network training is 0.001, learning rate is 0.01, trainlm() functions as the training network.

6. Training and results

Experimental images are provided by a hospital in Daqing. The image samples selected covers the 10 kinds of lesions forms of tumour and calcification. There are 84 training sample images. After segmentation by fuzzy clustering, 244 segmented regions have been produced. There are 80 validation image samples, after segmentation by fuzzy clustering, 238 segmented regions have been produced. Then, calculate the 7 characteristic parameters of each partition region as the input of neural network. And then, medical experts diagnose each division parts and give descriptions of lesions forms. Thus corresponding output vector can be obtained.

Training and identification results are shown in Table 4 and Table 5.

TABLE 4 TRAINING RESULTS OF BP NEUTRAL NETWORK

Number of Input Nodes	Number of Hidden Nodes	Error Target	Learning Rate	Iteration Times
7	14	0.002	0.01	143

TABLE 5 IDENTIFICATION RESULTS OF BP NEURAL NETWORK

Forms of Lesions	Number of Samples	Accurate Rate
Conglomeration Shape	23	100%
Star-Shaped Spikes	27	96.296%
Cloud sheet	28	96.429%
Hemispherical shape	24	95.833%
Comet Shape	32	96.875%
Round	24	95.833%
Oval	28	96.429%
Irregular Polygon Shape	18	94.444%
Linear	16	93.750%
Bifurcation Shape	18	100%

7. Conclusion

Due to the complexity texture of breast molybdenum palladium ray image and diversity of forms of breast lesions, the identification of breast cancer has in great trouble. In this paper, a new image recognition method of breast cancer is realized by combining fuzzy image enhancement, fuzzy clustering image segmentation and the fractal texture analysis method and technique. The merit of this method is in that it thins the form of lesions and extracts effective parameters. The defect is that the fuzzy image segmentation algorithm's efficiency is not high, sometimes has leakage of segmentation, this will affect the overall effect, so still need to improve.

8. References

[1] LI Xiao-feng, SHEN Yi, WANG Qiang. Computer-aided diagnosis system of breast tumor ultrasound images[J].Journal of Jilin University (Engineering and Technology Edition), 2009,39(3):700-702.

[2] Yang Wei. Research on Aided Diagnosis of Microcalcification Based on Hidden Markov Tree Model in Wavelet Domain:[dissertation]. Xiangtan: Xiangtan University,2009.

[3] Hao Xin, Cao Ying, Xia Shun-Ren. Computer-aided Diagnosis Technique on Mammograms Using Content-based Medical Image Retrieval[J].Chinese Journal of Biomedical Engineering, 2009,28(6):922-924.

[4] Lu Wen, Zhang Guang-yu. Effect of Fuzzy Genetic Artificial Neural Networks Agorithm Extracting Microcalcification[J]. JOURNAL OF SUN YAT-SEN UNIVERSITY(MEDICAL SCIENCES), 2008,29(2):216-219.

[5] Antonie M L, Zaiane O R, Coman A. Application of data mining techniques for medical image classification [C]. In: Proc. of Second Intl Workshop on Multimedia Data Mining (MDM/KDD’2001)in Conjunction with Seventh ACM SIGKDD, San Francisco, USA, 2001.94~101.

[6] A.E.Hassanien. Classification and feature selection of breast cancer data based on decision tree algorithm[J], International Journal of Studies in Informatics and Control,2003,12(1):33-39.

[7] J.C. Bezdek, R. Ehrlich andW. Full, FCM: The Fuzzy c-Means Clustering Algorithm, Computers and Geosciences 10 (1984),191–203.

[8] J.C. Bezdek and S.K. Pal, Fuzzy Models for Pattern Recognition,IEEE Press, 1992.

[9] R.M. Haralick, Statistical and structural approaches to texture,Proceedings of the IEEE 67(5) (1997).

[10] SARKAR N, CHAUDHUR IB B. An efficient app roach to estimate fractal dimension of textural images[J]. Pat2tern Recognition, 1992, 25 (9) : 1035 - 1041.

[11] Xu Dong, Wu Zheng, System Analysis and Design Based on MATLA6.x: Neural Networks, Xi An:Xidian University Press,2002.