IJCSC International Journal of Computer Science and Communication Vol. 3, No. 1, January-June 2012, pp. 17-20

A NEW REGION GROWING SEGMENTATION ALGORITHM FOR THE DETECTION OF BREAST CANCER

B. Senthilkumar[1] and G. Umamaheswari[2]

[1]Department of Electronics and Communication Engineering, Tamilnadu College of Engineering, Coimbatore, India, *E-mail: sen17580@yahoo.co.in.*

[2]Department of Electronics and Communication Engineering, P.S.G College of Technology, Coimbatore, India.

ABSTRACT

As medical images are mostly fuzzy in nature, segmenting regions based intensity is the most challenging task. Segmentation of medical images using seeded region growing technique is increasingly becoming a popular method because of its ability to involve high-level knowledge of anatomical structures in seed selection process. In this paper, we have made improvements in region growing image segmentation for mammogram images to detect the breast cancer. Selective median filter is used for preprocessing, CLAHE (Contrast Limited Adaptive Histogram Equalization) method is used for the enhancement, Harris corner detect theory is used to auto find growing seeds and the seeded region growing rule for the development of regions. This work also includes a new uncertainty theory-Cloud Model to realize automatic and adaptive segmentation threshold selecting, which considers the uncertainty of image and extracts concepts from characteristics of the region to be segmented like human being. We found this method works reliable on homogeneity and region characteristics. Furthermore, the method has been tested for over 30 sample images and the results found were good.

Keywords: Breast cancer, mammogram, segmentation, region growing.

1. INTRODUCTION

In computer vision, segmentation refers to the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics [15]. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic (s). Several general-purpose algorithms and techniques have been developed for image segmentation [16]. Since there is no general solution to the image segmentation problem, these techniques often have to be combined with domain knowledge in order to effectively solve an image segmentation problem for a problem domain. In this paper we have proposed a new region growing segmentation algorithm for the mammogram image segmentation to detect breast cancer. And this consists of noise removal (Selective median filtering), enhancement (CLAHE), seed selection, Harris corner detect theory and the cloud model.

2. PREPROCESSING

The basic need for enhancement in mammography is to increase the contrast, especially for dense breasts. Contrast between malignant tissue and normal dense tissue may be present on a mammogram but below the threshold of human perception. Similarly, microcalcifications in a sufficiently dense mass may not be readily visible because of low contrast [4]. As a result, defining the characteristics of microcalcifications is difficult. Conventional image processing techniques do not perform well on mammographic images. The large variation in feature size and shape reduces the effectiveness of classical fixed neighborhood techniques such as unsharp masking. Fixed neighborhood or global techniques may adapt to local features within a neighborhood, but do not adept the size of the neighborhood to local properties. Alternatively, they modify the image depending on global properties, such as the image spatial-frequency spectrum, which may not be representative of a small region of interest in the image [1]. Many images, including mammograms, have isolated regions, which are the primary feature of interest. These features can vary widely in size and shape, and often cannot be enhanced by fixed neighborhood or global techniques. There are two possible approaches to enhancing mammographic features. One is to increase the contrast of suspicious areas as stated earlier, and the other is to remove background noise [2]. The "region-based image processing" technique which adapts to image features and enhances these features with respect to their surroundings, regardless of the shape and size of the features. In adaptive-neighborhood or region-based image processing, a neighborhood is defined about each pixel in the image, the extent of which is dependent on the characteristics of the image feature in which the given pixel is situated. This neighborhood of similar pixels is called a region [14]. If properly defined, regions should correspond to image features. Then, image-processing procedures can be applied on an image feature basis, rather than pixel-by-pixel. There are two classes of regions: non overlapping regions, which are they disjoint segmentation of an image with subsequent enhancement of the segments would result in noticeable edge artifacts and an inferior enhanced image. This method uses each pixel in the image as a seed to grow a region. The extent and shape of the region adapt to local image gray-level variations, corresponding to an image feature. The contrast of each region is calculated with respect to its individual background [6]. Applying an impirical transformation based on the seed pixel value of each region, its contrast, and its background then enhances contrast. The objective of this scheme is to enhance the quality of "difficult" mammograms to allow the radiologists to make their diagnosis. In order to achieve this objective, a high-resolution digitization is maintained throughout the processing procedures. Removing background noise while preserving the edge information of suspicious areas can enhance a digital mammogram. This approach was investigated by Lai *et al.* [3], who used four selective averaging schemes and a modification of median filtering called selective median filtering. A selective median filter is, Given a window $W(i, j)$, centered at image coordinates (i, j), the output of the selective median filter is given in (1).

$$X_{i,j} = \text{median}\,\{X_{r,s,} : (r, s) \in N(i, j),\ \text{and}\ |X_{r,s} - X_{i,j}| < T\} \quad (1)$$

And where $X_{i,j}$ is the image intensity at (i, j), $N(i, j)$ is the area in the image covered by the window $W(i, j)$, and T is a threshold. In computing the median, the set of pixels is restricted to those with a difference in gray level no greater than some threshold T. Adjusting the parameter T can control the amount of edge smearing. If T is small, the edge-preserving power of the filter is strong, but its smoothing effect is small. If T is large, the filter behaves the other way around. It is used to remove the background noise with MAE (mean average error) of 0.45 and MSE (mean square error) of 11.91, which is less than the other noise reduction techniques. Contrast limited adaptive histogram equalization (CLAHE) is a special class of adaptive histogram equalization. Adaptive histogram equalization maximizes the contrast throughout an image by adaptively enhancing the contrast of each pixel relative to its local neighbourhood [7, 18]. This process produces improved contrast for all levels of contrast in the original image [8]. For adaptive histogram equalization to enhance local contrast, histograms are calculated from small regional areas of pixels, producing local histograms. These local histograms are then equalized or remapped from the often-narrow range of intensity values indicative of a central pixel and its closest neighbours to the full range of intensity values available in the display [17]. The digital mammogram processed with CLAHE, lesions appear obvious to the background and the image detail is very good. This algorithm is useful for radiologists to see subtle edge information, such as spiculation.

3. SEEDS SELECTION BASED ON HARRIS CORNER DETECTOR

The seeds selecting depends on the nature of the problem. The conventional seed selecting method is an interactive way, which means this method is not automatic. Some researchers use edge-based method to select seeds. So, the seed is just beside the edge, but to get more accurate and good segment result, take the center of the region as seed will make the segmentation more effective. For automatic seed selection, the following three criteria must be satisfied. First, the seed pixel must have high similarity to its neighbors. Second, for an expected region, at least one seed must be generated in order to produce this region. Third, seeds for different regions must be disconnected [3]. Based on the automatic seed selection criteria, we use Harris corner detect theory to realize automatic seed selection. The Harris corner detector is a popular interest point detector due to its strong invariance to rotation, scale, illumination variation and image noise. Harris corner detector is based on the local auto-correlation function of a signal; where the local autocorrelation function measures the local changes of the signal with patches shifted by a small amount in different directions [9]. Given a shift $(\Delta x, \Delta y)$ and a point (x, y), the autocorrelation function is defined in (2),

$$s(x, y) = \sum_{w} w(x, y)[I(x_i, y_i) - I(x_i + \Delta x, y_i + \Delta y)]^2 \quad (2)$$

where $I(,)$ denotes the image function and (x_i, y_i) are the points in the Gaussian window W centered on (x, y). The window function $w(x, y)$ is given in (3)

$$w(x, y) = \frac{1}{2\prod\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (3)$$

The shifted image is approximated by a Taylor expansion truncated to the first order terms is given below in (4),

$$I(x_i + \Delta x, y_i + \Delta y) = I(x_i, y_i) + [I_x\, I_y] \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (4)$$

where I_x and I_y denote the partial derivatives in x and y, respectively. Substituting approximation equation (4) into (2) gives (5).

$$s(x, y) = [\Delta x, \Delta y] \sum_w w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$$

$$= [\Delta x, \Delta y]\, S(x, y) \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (5)$$

Where matrix $S(x, y)$ captures the intensity structure of the local neighbourhood. Let λ_1, λ_2 be the eigenvalues of matrix $s(x, y)$. The eigenvalues form a rotationally invariant description. By analyzing the eigenvalues of $S(x, y)$, there are three cases to be considered. First, if the λ_1 and λ_2 are small, so that the local auto-correlation function is flat, that means there is little change in $s(x, y)$ in any direction, the windowed image region is of approximately constant intensity [9]. Second, if one eigenvalue is high and another is low, so the local auto-correlation function is ridge shaped, then only local shifts in one direction cause little change in $s(x, y)$ and significant change in the orthogonal direction and this indicates an edge. Finally, if both eigenvalues are high, so the local auto-correlation function is sharply peaked, then shifts in any direction will result in a significant increase and this indicates a corner. Based on the above analysis, we use it to find the point in the uniform area. In this process, the Gaussian window size is 6*6, $\sigma = 0.8$, and to get the most flat point and not too many points be detected we used a non-minimum inhibition window, the non-minimum inhibition window size is used, the lesser seeds are selected. The threshold of λ_1 and λ_2 is also influenced the number of selected seeds. The lower threshold will generate lesser seeds and bigger region of interest [5]. Here, the threshold is 0.01. The figure shows the original mammogram image and the seed selected image based on Harris corner detector theory.

4. THE CLOUD MODEL

Uncertainty is widely existed in the subjective and objective world. In all kinds of uncertainty, randomness and fuzziness are the most important and fundamental. Cloud model is an effective tool of uncertain transition between qualitative concepts and their quantitative expressions, can express the relationship between randomness and fuzziness [10]. It is in accord with the process of human thinking. It is a simple and effective way to simulate the uncertainty by means of knowledge representation which provides a basis for the automation of both logic and image thinking with uncertainty [5]. We suppose that U is a quantitative domain represented by accurate numerical value, $U = \{x\}$; C is a qualitative concept under U. If the element $x \in U$, and x is a random implement of C, the certainty degree of x to C, $\mu(x) \in [0, 1]$ is a random number with stable tendency [11].

$$\mu : U \to [0, 1] \ \forall x \in U \ x \to \mu(x) \quad (6)$$

From (6) the distribution of x in U is called Cloud, and each x is called a cloud drop. Cloud model has three numerical characteristics, Expected value (Ex), Entropy (En) [2], and hyper-Entropy (He), which are used to reflect the features of the concept [12, 13]. CG (forward Cloud Generator) generates the cloud with the help of given numerical characteristics (Ex, En and He) and CG^{-1} (backward Cloud Generator) generates the numerical characteristics from the cloud drop. From the point of view of cognitive science, concept is the basic cognitive element. It is corresponding to a quantitative data space, and is the nature form of thinking about the object formed in the minds of human. In order to make use of the abstract concept to observe and analyze region, we must in some way to express the region into concepts. From the Cloud Model theory, we know that it's a concept express model. It reflects the homogeneity, fuzziness and randomness of a region. So, we use CG^{-1} to extract the concept of the region around the seed, it can get the concept's connotation and extension [13], which is used here as the segment threshold.

5. RESULTS

The proposed region growing algorithm detects the cancer in an effective way and the resulted images are given in Figure 1 and Figure 2.

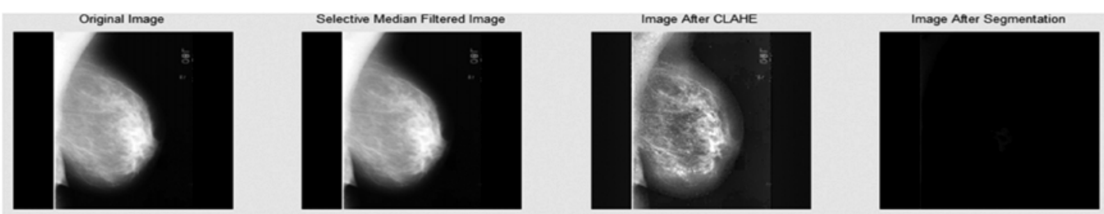

Figure 1: Results of the Mammogram Image (mdb218) Taken from MIAS Database.

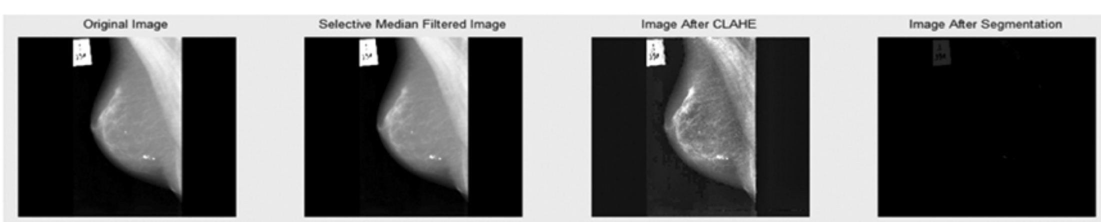

Figure 2: Results of the Mammogram Image (mdb75) Taken from MIAS Database.

In Figure 1, first image from left is the original mammogram, the second is the modified selective median filtered image, third is the image after CLAHE and fourth is the segmented (calcification detected) image. In this mdb218, it has already been proven that it contains benign (fatty – glandular, calcification). Similar to this, mdb75 has been tested and found that a fatty breast with malignancy and shown in Figure 2.

6. CONCLUSION

Region growing method can correctly separate the regions and provide the original images which have clear edges with good segmentation results. The concept is simple and only need small numbers of seed point to represent the property we want. We have determined the seed points and multiple criteria at the same time. Here the Harris corner detector seed selection method and the cloud technique performs well on mammogram images and able to find the cancer area (Calcification) clearly. It also combines with selective median filtering and CLAHE and performs well in breast cancer detection with comparatively small MAE (0.45) and MSE (11.91). The detection accuracy of this new method is 93%. This computer aided detection method can correctly detects the cancer location but not capable of diagnosing between benign and malignancy. Our future work will include the design and development of a new expert system for the real time analysis (detection and diagnosis) of mammogram image.

REFERENCES

[1] Mehul P. Sampat, Mia K. Markey and Alan C. Bovik, "Computer-Aided Detection and Diagnosis in Mammography", Elsevier Academic Press, pp. 1195-1217, 2005.

[2] Sheng-Chih-Yang, Chuin-Mu Wang, Yi-Nung Chung, Giu-Cheng Hsu, San-Kan Lee, Pau-Choo Chung and Chein-1 Chang, "A Computer Aided System for Mass Detection and Classification in Digitized Mammograms", *Biomedical Engineering- Applications, Basis and Communications*, **17**, pp. 215-228, 2005.

[3] S. Lai, X. Li and W. Bischof, "On Techniques for Detecting Circumscribed Masses in Mammograms", *IEEE Trans. Med. Imag*, **8**, pp. 377-386, 1989.

[4] W. Morrow, R. Paranjape, R. Rangayyan, and J. Desautels, "Regionbased Contrast Enhancement of Mammograms", *IEEE Trans. Med. Imag.*, **11**, pp. 392–406, 1992.

[5] Shu-Yen Wan and William E. Higgins, "Symmetric Region Growing", *IEEE Trans. Image Processing*, **12**, pp. 1007-1015, 2003.

[6] Alfonso Rojas Dominguez and Ashok K.Nandi, "Detection of Masses in Mammograms Via Statistically Based Enhancement, Multilevel-Thresholding Segmentation and Region Selection", *Computerized Medical Imaging and Graphics*, **32**, pp. 304-315, 2008.

[7] S. M. Pizer, E. O. P. Amburn and J. D. Austin, "Adaptive Histogram Equalization and its Variations", *Comput. Vision Graphics Image Process*, **39**, pp. 355-368, 1987.

[8] A. Papadopoulos, D. I. Fotiadis and L. Costaridou, "Improvement of Microcalcification Cluster Detection in Mammography Utilizing Image Enhancement Techniques", *Computers in Biology and Medicine*, vol.38, pp. 1045-1055, 2008.

[9] C. Schmid, R, Mohr and C. Bauckhage, "Evaluation of Interest Point Detectors", *International Journal of Computer Vision*, **37**, pp. 151-172, 2000.

[10] Qin Kun, Li Deyi and Xu Kai, "Image Segmentation Based on Cloud Model, Information Mapping and Engineering", **31**, pp. 3-5, 2006.

[11] Li Deyi, Liu and Changyu, "Study on the Universality of the Normal Cloud Model", *Engineering Science*, **6**, pp. 28-34, 2004.

[12] Xu Kai, Qin Kun and Pei Tao, "Interactive Method for Image Segmentation Based on Cloud Model", *Computer Engineering and Application*, **34**, pp. 33-35, 2006.

[13] Liu Changyu, Li Deyi, Du uo and Han Xu, "Some Statistical Analysis of the Normal Cloud Model", *Information and Control*, **34**, pp. 236-239, 2005.

[14] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai and H. N. Du, "Approaches for Automated Detection and Classification of Masses in Mammograms", *Pattern Recognition*, **39**, (2006), pp. 646-668.

[15] S. Singh and R. Al-Mansoori, "Identification of Regions of Interest in Digital Mammograms", *J. Intell. Systems*, **10**, pp. 183-210, 2000.

[16] R. M. Rangayyan, L. Shen, Y. Shen, J. E. L. Desautels, H. Bryant, T. J. Terry, N. Horeczko and M. S. Rose, "Improvement of Sensitivity of Breast Cancer Diagnosis with Adaptive Neighbourhood Contrast Enhancement of Mammograms", *IEEE Trans. Inform. Technol. Biomed.*, **1**, pp. 161-170, 1997.

[17] Y. J. Lee, J. M. Park and H. W. Park, "Mammographic Mass Detection by Adaptive Thresholding and Region Growing", *Int. J. Imaging Systems Technol.*, **11**, pp. 340-346, 2000.

[18] N. Petrick, H. P. Chan, B. Sahiner and M. A. Helvie, "Combined Adaptive Enhancement and Region Growing Segmentation of Breast Masses on Digitized Mammograms", *Med. Phys.*, **26**, pp. 1642-1654, 1999.