Mass Segmentation in Mammograms Using Texture Features and Fuzzy C-means Algorithm

Moustapha Mohamed Saleck*, Abdelmajide El Moutaouakkil
LAROSERI, Dept. of Computer Science,
Chouaib Doukkali University,
El Jadida, Morocco
E-mail* : saleck.moustapha@gmail.com
E-mail : elmsn@hotmaill.fr

Abstract— **The Fuzzy C-means (FCM) is one of the most efficient algorithms used in various studies, which aims at segmenting the masses in mammogram images, thus to build a Computer Aided Diagnosis (CAD) system capable of helping the physicians for an early diagnosis of the breast cancer. In this paper, we will introduce a new approach using FCM algorithm, in order to extract the mass from Region-of-Interested (ROI). The proposed method aims at avoiding the limitations of cluster number estimation in FCM by selecting as input data, the set of pixels, which are able to provide us the information required to perform the mass segmentation by fixing two clusters only. The Gray Level Occurrence Matrix (GLCM) is used to extract the texture features for getting the optimal threshold, which separate between selected set and the other sets of the pixels that influence on the mass boundary accuracy. The performance of the proposed method is evaluated by specificity, sensitivity, accuracy and overlap. The results obtained from experiments show a good efficiency at the different measures used in this study.**

Keywords- Mammograms; Mass; Fuzzy C-Means; Segmentation; Texture features.

I. INTRODUCTION

Breast cancer is one of the most common dangerous diseases in women. According to the World Health Organization, every year, breast cancer kills more than 500,000 women around the world [1]. Although several imaging techniques, such as sonography and magnetic resonance imaging plays important role in breast cancer diagnosis, X-ray mammography is still the most effective screening technique for the early detection of breast cancer [2].

In the last few decades a large amount of researches has been conducted to detect and segment breast cancer in mammograms images. Soares et al. [3] used taxonomic indexes to describe the texture of the regions of interest. Then, a Support Vector Machine (SVM) was proposed to classify the regions as mass and non-mass. Younesi and al. [4] developed a segmentation method for detection of masses in mammogram images using adaptive thresholding method and fuzzy entropy feature. Anitha et al. [5] proposed an automatic method to identify the suspicious mass region by using a Dual Stage Adaptive Thresholding (DuSAT). Cordeiro and al. [4] proposed an adaptive semi-supervised version of the Grow-Cut algorithm to perform mammographic image segmentation. This method is achieved based on automatic selection of internal points using the differential evolution optimization algorithm and modification of cellular automata evolution rules by introducing Gaussian fuzzy membership functions. Nija et al. [5] investigated combining several image enhancement algorithms in order to enhance the performance of masses segmentation. The results of this investigation showed that a particular combination of image enhancing algorithms that includes Contrast - Limited Adaptive Histogram Equalization (CLAHE) and Median Filtering is an effective way to enhance the appearance of the breast region in mammogram images to be further segmented and classified.

Extracting the mass from Region of Interest (ROI) is a difficult task in mammography Computer-Aided Diagnosis (CAD) due to several factors, such as low contrast, density, indistinct borders and ill-defined shapes of the mass. Fuzzy clustering is more efficient than traditional techniques to handle the fuzziness of mammograms. On the other hand, choosing an optimal number of clusters is a big challenge for automating segmentation of masses by clustering.

In this paper, we aim to develop an automated system for mass segmentation in mammograms using FCM algorithm setting two cluster only (C=2). In order to achieve this goal, we select as FCM-input the set of pixels that enable us to get meaningful information from Region of Interest (ROI). This set of pixels is limited by maximum intensity of ROI and an optimal threshold given by a decrement operator, we get this optimal threshold when big changes happen in texture features.

The remainder of this paper is organized as follows : Section 2 presents the materials and methods used in this work. Section 3 describes the experimental results obtained from the evaluation of the proposed methods. Finally, Section 4 presents the conclusion.

II. MATERIAL AND METHODS

The proposed method consists of three main stages: Firstly, median filter and contrast limited adaptive histogram equalization are applied for enhancing the contrast and quality of images. Secondly, a decrement operator L is used to find an appropriate threshold by increasing the number of pixels (FCM-input). At every decrement process, we monitor

 ISBN: 978-1-61208-638-5

the changes in texture features levels on the area of mass cluster after applying fuzzy c-means clustering with a fixed number of clusters c = 2, where, the set of input data that is subjected to process of clustering should be within a range limited by maximum of intensity M and a threshold test Ti initialized by $T_0 = M - L$. Finally, we choose the value of threshold test, which makes big changes in texture features levels on the mass area as a suitable threshold Top. Figure 1 describes the evolution of input-data and texture features selection in each step of threshold test decrementation.

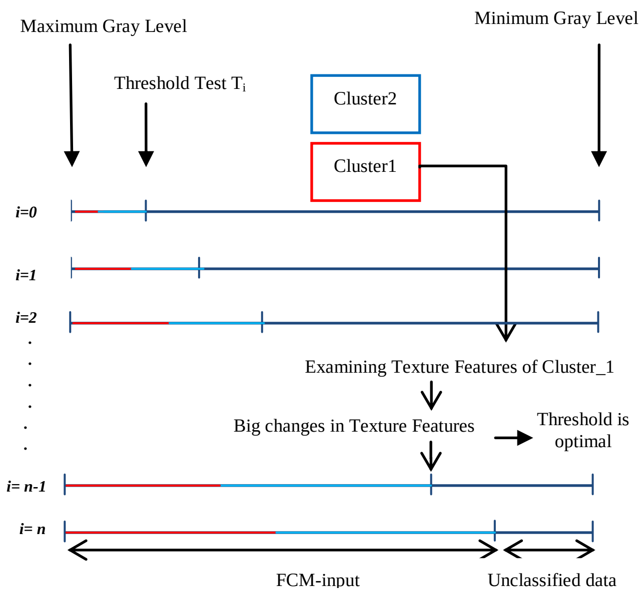


Figure 1. Process of choosing an optimal threshold

A. *Contrast-limited adaptive histogram equalization*

The Contrast Limited Adaptive Histogram Equalization (CLAHE) technique [6] is a special case of Adaptive Histogram Equalization (AHE) [7]. This technique operates on regional areas of pixels in the image called tiles rather than on the entire image. Histograms are calculated from these tiles, producing local histograms. These local histograms are then equalized for getting a uniform distribution. Then, the neighboring tiles are combined based on bilinear interpolation to remove artificially induced boundaries. In this study, we use CLAHE technique to enhance the contrast of image and reduce edge-shadowing effects produced in homogeneous regions and, consequently, improve the accuracy of texture features results.

B. *Fuzzy C-Means (FCM)*

The Fuzzy C-means algorithm was first introduced by Dunn [8] and improved by Bezdek [9]. This algorithm partitions a set of object data X = { $x1$, $x2$, ….., $x3$} into a number of c classes so that items in the same class are as similar as possible, based on minimizing the following quadratic objective function :

$$J = \sum_{k=1}^{c} \sum_{i=1}^{n} u_{ik}^{m} d^{2}(x_i, v_k) \tag{1}$$

where the fuzzy index m>1 determines the amount of fuzziness of the resulting classification, the membership u_{ik} represents the degree of pixel x_k belonging to cluster k ($1 \leq i \leq n$), and $d_{ik} = \| x_k - c_i \|$ is the distance between pixel x_k and the centroid v

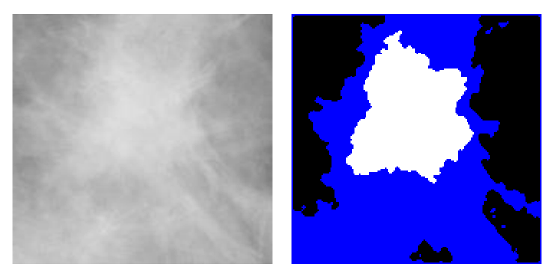
$$u_{ik} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d(x_i, v_k)}{d(x_i, v_i)} \right)^{2(m-1)}} \tag{2}$$

$$v_k = \frac{\sum_{i=1}^{n} u_{ik}^{m} x_i}{\sum_{i=1}^{n} u_{ik}^{m}} \tag{3}$$

The memberships satisfy the constraint that :

$$\sum_{k=1}^{c} u_{ik} = 1, \quad u_{ik} \in [0,1], \quad 0 \prec \sum_{i=1}^{n} u_{ik} \prec n, \tag{4}$$

With this constraint, the objective function reaches a local minimum by updating the fuzzy membership function (2) and cluster centers (3).



☐ Cluster_1 ■ Cluster_2 ■ unclassified pixels

Figure 2. The result of segmentation on image Mdb10.

C. *Texture features*

Statistical texture features have been proven to be useful in classifying masses and normal breast tissues [10][11]. In this work, we examined a set of three texture features (energy, contrast, and homogeneity) from mass area (cluster_1) to identify the most suitable threshold that enables us to extract the boundaries of mass by clustering. The process of texture features selection is achieved using Gray Level Co-occurrence Matrix (GLCM).

D. *Gray-level co-occurrence matrix*

The Gray Level Co-occurrence Matrix (GLCM) is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image with the change of distance d (distance to the neighbor pixel: 1, 2, ..

 ISBN: 978-1-61208-638-5

etc) and by varying directions (rotation angle of an offset: 0º, 45º, 90º, 135º) [12]. A number of studies have compared texture feature extraction schemes based on the second-order gray-level statistics, the co-occurrence statistics, gray-level run-length statistics, and Fourier power spectrum. The co-occurrence features were found to be the best of these methods. [13]. In our approach, the GLCM is computed at a distance of d = 1 and for the direction of θ = 0°, 45°, 90 °, 135 ° on cluster_1 area. Based on the authors' database, the smallest masses area is within a 8x8 window. For this reason we use a tile of size 8x8.

III. EXPERIMENTAL RESULTS

For the experiments, the MiniMIAS database was used. It contains 322 mammographic images from left and right breast of 161 patients. The mammograms are digitized at a resolution of 1024 X 1024 pixels and at 8-bit grey scale level.

After many tests, it was found that the best features for distinguishing between masses and normal breast tissues are the features of GLCM constructed at the direction of 0°. At this direction, the GLCM gives a homogeneity value for the masses area in the range [0.9–1], contrast in [0–0.2], and energy in [0.6–1]. Similarly, for normal breast tissue area, the GLCM gives the homogeneity in the range [0.61–0.8], contrast in [0.26–0.71], and energy in [0.14–0.39]. It is observed that the texture features for normal breast tissues and masses area are highly discriminated. In this study, we examined the texture features after the process of clustering in order to find a threshold that allows us to extract the boundaries of tumor with high accuracy. Figure 2 shows the values of texture features during the process of threshold decrementing for image Mdb10.

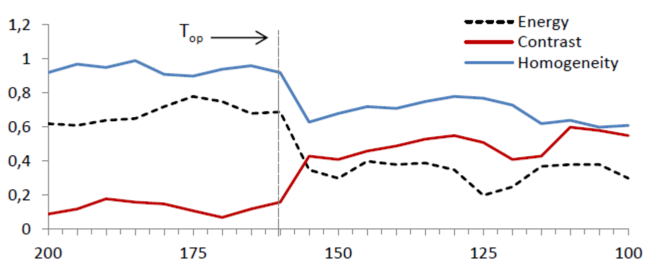


Figure 3. Contrast, Energy and homogeneity values using block size 8 × 8

To evaluate the performance of segmentation all masses are manually marked by radiologist based on the visual criteria.

Assuming that A is the area marked by radiologist and B is the area marked by the system, the area overlap metric (AOM) is given by:

$$AOM(A, B) = Area(A \cap B) / Area(A \cup B) \quad (5)$$

In this work, the test by the measure of overlap (AOM) yielded a mean of 81% for the segmented images.

IV. CONCLUSIONS

In this paper, we developed a novel method for automated detection of masses in mammogram images. The proposed work utilizes fuzzy c-means algorithm to extract the tumor from region of interest, where the FCM input data are verified by using the GLCM feature texture in order to automate the process of segmentation.

REFERENCES

[1] World Health Organization, "WHO Position Paper on Mammography Screening," *WHO Position Pap. Mammogr. Screen.*, p. 78, 2014.

[2] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances.," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 236–251, 2009.

[3] F. Soares Sérvulo de Oliveira, A. Oseas de Carvalho Filho, A. Corrêa Silva, A. Cardoso de Paiva, and M. Gattass, "Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM," *Comput. Biol. Med.*, vol. 57, pp. 42–53, 2014.

[4] F. R. Cordeiro, W. P. Santos, and A. G. Silva-Filho, "A semi-supervised fuzzy GrowCut algorithm to segment and classify regions of interest of mammographic images," *Expert Syst. Appl.*, vol. 65, pp. 116–126, 2016.

[5] N. Al-Najdawi, M. Biltawi, and S. Tedmori, "Mammogram image visual enhancement, mass segmentation and classification," *Appl. Soft Comput. J.*, vol. 35, pp. 175–185, 2015.

[6] K. Zuiderveld, "VIII.5. - Contrast Limited Adaptive Histogram Equalization A2 - Heckbert, Paul S. BT - Graphics Gems," Academic Press, 1994, pp. 474–485.

[7] S. M. Pizer *et al.*, "Adaptive histogram equalization and its variations.," *Comput. vision, Graph. image Process.*, vol. 39, no. 3, pp. 355–368, 1987.

[8] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, Jan. 1973.

[9] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, vol. 25, no. 3. 1981.

[10] N. Youssry, F. E. Z. Abou-Chadi, and A. M. El-Sayad, "Early detection of masses in digitized mammograms using texture features and neuro-fuzzy model," in *National Radio Science Conference, NRSC, Proceedings*, 2003, vol. 2003–Janua, pp. K21–K29.

[11] A. M. Khuzi, R. Besar, and W. M. D. W. Zaki, "Texture features selection for masses detection in digital mammogram," in *IFMBE Proceedings*, 2008, vol. 21 IFMBE, no. 1, pp. 629–632.

[12] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Syst. Man. Cybern.*, vol. 3, no. 6, pp. 610–621, 1973.

[13] P. Hiremath and S. Shivashankar, "Texture classification using wavelet packet decomposition," *ICGSTs GVIP J.*, 2006.