JCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693

A Hierarchical Spatial Fuzzy C Means Algorithm for Mammographic Mass Segmentation

Manasi Hazarika[1*,2], Lipi B. Mahanta[3]

[1]Dept. of Computer Science, Gauhati University, Guwahati, India
[2]Dept. of Computer Science & Engineering and IT, Assam Don Bosco University, Guwhati, India
[3]Dept. of Central Computational and Numerical Sciences Division, Institute of Advanced Studies in Science and Technology, Guwahati, India

*Corresponding Author: manasi.hazarika@gmail.com

Available online at: www.ijcseonline.org

Accepted: 17/Jan/2019, Published: 31/Jan/2019

Abstract - Fuzzy C Means is one of the most popular machine learning technique for image segmentation. However, traditional Fuzzy C Means is insensitive to noise as it does not consider spatial information. To solve this issue a wide variety of modified Fuzzy C means techniques, considering spatial information of pixels, are proposed by different researchers. In this paper we propose a hierarchical Fuzzy C Means algorithm considering spatial features of image pixels. Our method aims to overcome the shortcomings of traditional Fuzzy C Means by incorporating spatial feature as well as the issue of misclassification of pixels associated with single level clustering. The proposed method divides the original image pixels into a set of clusters using a spatial fuzzy C means technique in the first level of the hierarchical model. In the second level of the hierarchy, the cluster which contains the candidate mass is further divided into sub clusters using traditional Fuzzy C Means algorithm to yield the final segmentation result. The experimental outputs show improved segmentation result by our proposed method.

Keywords—Clustering, Fuzzy, Spatial, Segmentation, Hierarchical

I. INTRODUCTION

Breast cancer is one of the most common cancers detected amongst women throughout the world. According to the reports of American Institute for Cancer Research in 2018 around 2 million new cases of breast cancer has been registered across the world [1]. Mammography is the most preferred method for early diagnosis of breast cancer. The recent progression of Computer Aided Detection (CAD) system made it a very reliable choice for a radiologist to take it as a second opinion.

Mammographic abnormalities can be categorized as microcalcification or space occupying lesion[2]. Microcalcifications are tiny calcium deposits stays in clusters. Space occupying lesions can be divided into masses, asymmetry and architectural distortions. Masses, microcalcifications and architectural distortions are the typical signs of breast cancer. Depending on the shape and boundaries masses can be further divided into circular, oval, speculated and misc. Masses with smooth boundaries are often found as benign and the masses with ill defined boundaries are mostly malignant. In case of a CAD system segmentation of masses is a very challenging and crucial step. Fully automatic segmentation of masses is very difficult as the masses are often come with irregular boundaries, low contrast, overlapped tissues. In recent years, various studies on mass detection have been made. Segmentation methods can be broadly categorized into region based, contour based, model based, graph based and clustering based etc.

In present time clustering based methods have become a very popular choice for image segmentation for various pattern recognition applications. The goal of clustering is to divide the data into different clusters based on certain similarity measures. Clustering methods can be categorized into a) partitioning methods, b) grid based methods, c) hierarchical, d) density based, e) model based and d) fuzzy set theory based. Amongst these methods fuzzy set theory based clustering methods particularly fuzzy C means is found to be more effective for medical image segmentation as most of the medical images come with lots of fuzzy characteristics. FCM is widely used for image segmentation due to its robust characteristics. However in conventional FCM feature vectors are processed independently without considering their spatial correlation. In case of medical images pixels are highly correlated to their neighboring pixels. In such a scenario a clustering method that incorporates the details of neighboring pixels can produce a much better segmentation result.

Studies show in the last few decades several modifications have been made to improve the image segmentation. An adaptive FCM based image segmentation method was presented in [3]. In this work a novel dissimilarity index was introduced by utilizing the spatial correlation of the pixels. Zhou et al.[4] proposed a faster version of FCM by incorporating a mean shift term to the objective function of the standard FCM. A kernel based FCM to segment mammogram masses was proposed in [5]. The images were preprocessed before the application of the clustering method. Finally a level set function was applied to the cluster with highest mode value to get refined segmentation results.

The rest of the paper is of the paper is organized as follows: Section II gives a brief review of standard FCM and four variants of modified FCM. In section III, we presented our proposed Hierarchical Spatial Fuzzy C Means Algorithm (HSFCM). Section IV presents the experimental results and finally conclusion is presented in section V.

II. RELATED WORK

A. FUZZY C Means

Fuzzy Clustering was introduced in[6]to divide N datapoints into C clusters; Bezdek later improved this method by introducing Fuzzy C Means (FCM) algorithm[7]. In pattern recognition applications FCM is often used due its robust characteristics. FCM is basically a combination of fuzzy clustering and K-means algorithm. K-means is a hard clustering algorithm in case of which a datapoint can belong to only one cluster while FCM is a soft clustering technique in case of which a datapoint can belong to multiple clusters. The FCM algorithm assigns fuzzy membership values to each datapoint; a membership value μ_{ij}indicates the degree of membership of the i^{th} datapoint to the j^{th}cluster. FCM attempts to minimize the following objective function as specified by equation (1).The objective function is minimized when the datapoints are assigned highest degree of memberships to the clusters they are closely associated with.

$$J = \sum_{i=1}^{N}\sum_{k=1}^{C}\mu_{ik}{}^{m}\, d_{ik} (1)$$

Where N is the number of datapoints, C is the number of clusters, μ_{ik} is the degree of membership of i^{th} in the k^{th} cluster, m is the weighing exponent and is a constant, d_{ik} is the Euclidean distance of the i^{th} datapoint and from the k^{th} cluster. The algorithm is presented below.

Step 1. Choose number of clusters $\boldsymbol{C}(2\leq\boldsymbol{C}<\mathrm{N})$,weighing exponent m and sensitive threshold $\mathcal{E}$

Step 2. Initialize fuzzy membership matrix U, U=[μ_{ij}]

Step 3. Calculate cluster center, $v_k$as

$$v_k = \frac{\sum_i^N \mu_{ij}{}^m * y_i}{\sum_i^N \mu_{ij}{}^m} (2)$$

Step 4. Update the fuzzy membership matrix U as

$$\mu_{ik} = \frac{1}{\sum_{l=1}^{c}\left(\frac{d_{ik}}{d_{il}}\right)^{2/(m-1)}} (3)$$

Step 5. Repeat steps 3 to 4 until the change in coefficient is less than and equal to Ɛ

B. Fuzzy clustering with Spatial information(SFCM) [8]

A fuzzy clustering technique for image segmentation was proposed by Xiang by employing a local similarity measure model. In this paper the authors applied a local similarity measure model[9] by incorporating both spatial relationships and intensity levels amongst the image pixels. The local similarity measure model is presented below

$$F_{ij} = \begin{cases} F_{ij}^{S} \times F_{ij}^{G} & if\ i \neq j \\ 0 & if\ i = j \end{cases} (4)$$

Where F_{ij} is the spatial relationship between the centered pixel of a local window and one of its adjacent pixel in that local neighborhood; in this case i is indicating the centered pixel and jis indicating an adjacent pixel, F_{ij}^{S} indicates spatial relationship between pixel i and j while F_{ij}^{G} indicates the gray level relationship between these pixels. F_{ij}^{S} and $F_{ij}^{G}$are defined by the equations (5) and (6) respectively.

$$F_{ij}^{S} = exp\left(\frac{-max(|i_x - j_x|, |i_y - j_y|)}{\alpha_S}\right) (5)$$

$$F_{ij}^{G} = exp\left(\frac{-\|g_i - g_j\|^2}{\alpha_G \times v_i}\right) (6)$$

Where (i_x,i_y) and (j_x,j_y) specify the spatial coordinates of pixel i and pixel j respectively, g_i and g_j indicate the gray level intensities of pixel i and pixel j respectively. α_S and α_G are scale factor of the spread of F_{ij}^{S} and F_{ij}^{G}respectively and v_i is the variance inside the local neighborhood.

C. Fuzzy local information C means clustering algorithm[10]

Another variant of spatial information based FCM was proposed by Stelios Krinidis et al. This technique aims to provide a flexible means of controlling the influence of adjacent pixels based on their distance from the central pixel. The objective function proposed in this technique is defined as

$$J = \sum_{i=1}^{N}\sum_{k=1}^{C}[\mu_{ik}^{m}\|y_i - v_k\|^2 + FF_{ki}] (7)$$

Where y_i denotes the i^{th} datapoint, v_k denotes the k^{th} cluster center and FF_{ki} is the fuzzification factor defined as follows.

$$FF_{ki} = \sum_{l\in\varphi_i}\frac{1}{d_{il}+1}(1-\mu_{ki})^m\|y_l - v_k\|^2 (8)$$

Where φ_i is a nxn neighborhood around the pixel i, k is the referenced cluster, y_j is the intensity value of the lth pixel falling inside φ_i, $d_{il}$is the spatial distance between i^{th} and l^{th} pixel.

D. Enhanced Fuzzy C-Means Clustering[11]

A modified FCM was proposed by Ahmed et al. [12] by considering information of neighboring pixels during the clustering process. The objective function is defined by the equation (9).

$$J = \sum_{i=1}^{N}\sum_{j=1}^{C}\mu_{ij}^{m}\|y_i - v_j\|^2 + \frac{\alpha}{M}\sum_{i=1}^{N}\sum_{j=1}^{C}\mu_{ij}^{m}\sum_{r\epsilon\varphi_i}\|y_r - v_j\|^2 (9)$$

Where φ_i denotes the set of neighbor around the i^{th} pixel with gray level valuey_i, M is the cardinality value of the neighborhood, the parameter α is used to control the influence of the neighbors.

But main drawback of this technique is that the neighboring term has to be computed for each iteration. Thus this technique consumes lots of processing time. In order to improve this Szilagyi et al. [11] presented an enhanced FCM that takes lesser time to complete the clustering process. To accelerate the method proposed by Ahmed et al. a linearly weighted sum image βis computed prior to the iterations from the original image and average of its local neighbors in terms of

$$\beta_i = \frac{1}{1+\alpha}\left(y_i + \frac{\alpha}{M}\sum_{j\in\varphi_i} y_j\right)(10)$$

Where β_i is the gray level value of the image β. The objective function is defined as follows:

$$J = \sum_{i=1}^{N}\sum_{j=1}^{C} p_i\,\mu_{ij}^{m}\left(\beta_i - v_j\right)^2(11)$$

Where p_i indicates the frequency of having gray level valuei, v_j represents the cluster center of the j^{th} cluster.

E. Image Segmentation using Generalized Hierarchical Fuzzy C Means[13]

Hierarchical fuzzy clustering methods have become a very popular option for image segmentation in recent times. A generalized hierarchical FCM was proposed in [13]. Standard FCM suffers badly as it is insufficient to handle noisy images. Moreover the Euclidean distance used in traditional is very sensitive to outliers. To handle these issues the authors have proposed two new algorithms viz., generalized FCM(GFCM) and Hierarchical FCM(HFCM). The GFCM employs generalized local mean on the membership functions. The objective function used in GFCM is defined by equation(12).

$$J = \sum_{i=1}^{N}\sum_{j=1}^{C}\mu_{ij}^{m}\sum_{r\in N_i} wt_r d_{rj}(12)$$

Where N_i is the neihborhood around the pixel i, wt_r is the weighted factor used to control the effect of the local neighborhood based on their distance from the i^{th} pixel. wt_r should be inversly proportional to the inter pixel distance. d_{rj} is the distance between j^{th} pixel and r^{th}(of N_i pixel).

The proposed HFCM is a two level FCM. In this method at first the image is divided into C clusters, later on the clusters are divided further into sub clusters. Both GFCM and HFCM are finally combined into one method to get a new segmentation method.

III. PROPOSED METHOD

It has been observed that the methods discussed in the previous section performed well in their respective fields, but these techniques are not effective enough to accurately segment out masses from mammogram images. In this section we propose hierarchical fuzzy C-means method (HSFCM) incorporating spatial features to segment mammographic masses. The advantage of using a hierarchical FCM is that it helps in separating the merged clusters. Here we apply a two level FCM to perform the segmentation. The main idea is to make C1 clusters in the first level and then to re-apply clustering to one or two selected clusters out of C1 in the second level. In this work we have combined SFCM[8] and standard FCM. In the first level of the hierarchy we have applied spatial FCM followed by standard FCM in the second level. For our application we have made 3 clusters in the first level of clustering. The mammographic masses have higher intensity than the background. Thus the cluster with the highest cluster center possesses the candidate mass. So in the second level of FCM we have applied clustering only to the cluster with the highest cluster center. In this level also we have made 3 clusters and based on the experimental results it is observed that the cluster with the median cluster center represents the pixels in the mass area. The hierarchy is depicted in Figure 1, in which cluster 1 represents the cluster with smallest cluster center, cluster 2 represents the cluster with median cluster center and cluster 3 represents the cluster with highest cluster center. The algorithm is presented in the flowchart shown in Figure 2.

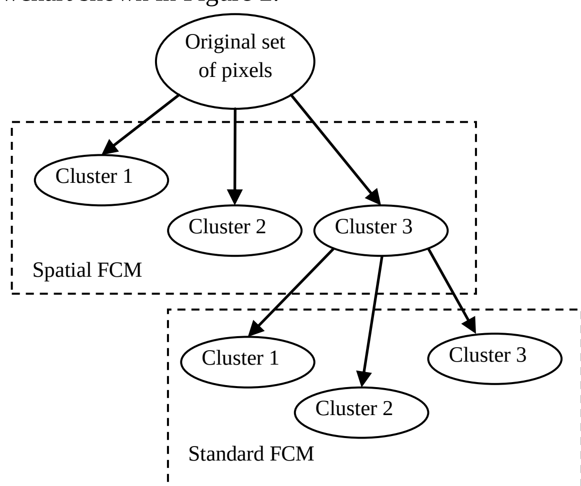


Figure 1: Cluster Hierarchy

IV. RESULTS AND DISCUSSION

In this work we have performed our experiments on the images collected from Mini Mias Dataset[14]. We have chosen a window of size 5x5 for calculating local features to perform clustering using spatial FCM.

The performance of the proposed method is compared with the performance of spatial FCM and standard FCM. Figure 3 shows the segmentation result of standard FCM, spatial FCM and our proposed method. From the figures it is clear that our algorithm performs better than spatial FCM and standard FCM.

To evaluate the performance of our proposed method we have we have calculated the segmentation accuracy of the proposed method, spatial FCM and standard FCM over five randomly selected mammogram images containing masses. The segmentation accuracy, SA is calculated as given by the equation (13).

$$SA = \frac{Total number of correctly classified pixel}{Total number of pixels} \times 100 \qquad (13)$$

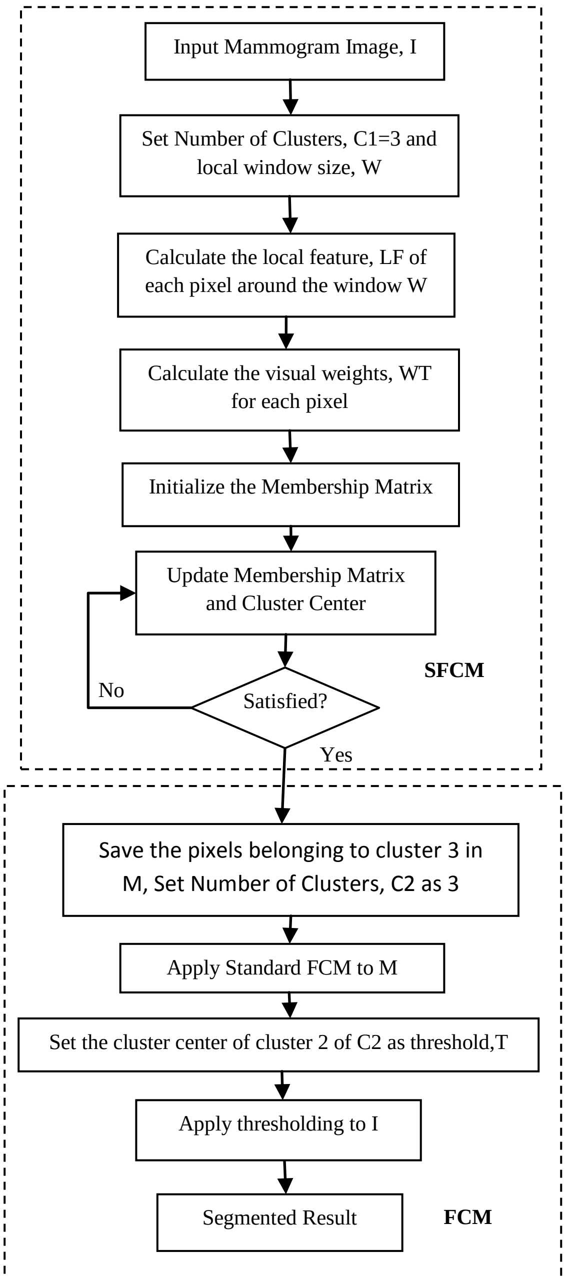
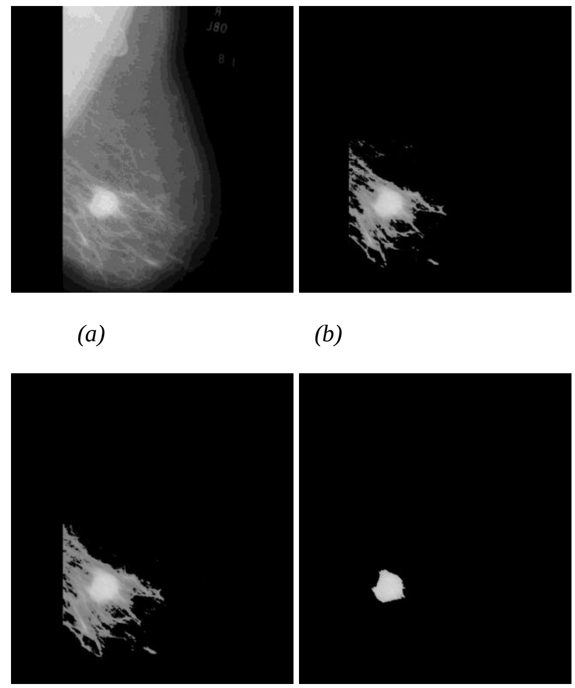


Figure 2: Flowchart of Proposed HSFCM Algorithm

The segmentation accuracies of different techniques are tabulated in table 1. From the table it is clear that our proposed method shows higher segmentation accuracy over spatial FCM and standard FCM when applied on the mammogram images.



(a) *(b)*

(c) *(d)*

Figure 3: (a) Original Image (mdb0025) (b) Segmented Image by FCM (c) Segmented Image by SFCM (d) Segmented Image by HSFCM

Table 1: Segmentation Accuracy of Different Methods

Filename	HSFCM	SFCM	FCM
Mdb005	99.50189590	87.15925216	92.16642379
Mdb025	99.33567047	90.99502566	97.26171493
Mdb028	99.63884353	93.47887039	94.44789886
Mdb0178	98.15261839	91.28917872	95.39863641
Mdb0184	99.71027516	94.21091857	97.65221693

V. CONCLUSION

Computer aided diagnosis system plays a very crucial role in early detection of breast cancer. In this paper, we have presented a hierarchical FCM method incorporating spatial information to segment suspicious masses from mammogram images by utilizing spatial features. Standard FCM does not consider spatial information, which makes it intolerant to noise. Although the spatial FCM discussed here incorporate local information effectively, it is not very effective in segmenting mammogram masses. But when these two methods are combined together in different levels of hierarchical FCM, they produce a very good segmentation results.

REFERENCES

[1] *"Breast Cancer Statistics"*, Worldwide Data, World Cancer Research Fund, USA, 2018

[2] *"American College of Radiology (ACR): ACR Breast Imaging Reporting and Data System"*, Breast Imaging Atlas, edn. 4, USA, 2003.

[3] A. W. C. Liew, H. Yang, H. F. Law, *"Image segmentation based on adaptive cluster prototype estimation"*, IEEE Trans. Fuzzy Syst., Vol.13, Issue4, pp.444–453,2005.

[4] H. Zhou, G. Schaefer, C. Shi, *"A mean shift based fuzzy c-means algorithm for image segmentation"*, In the Proceeding of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Canada, pp. 3091–3094, 2008.

[5] J. Anitha, J. D. Peter, "*Mass segmentation in mammograms using a kernel-based fuzzy level set method*", Int. J. Biomedical Engineering and Technology, Vol. 19, Issue.2, pp. 133-153,2015.

[6] J. C. Dunn, *"A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters"*, Journal ofCybernetics, Vol3, Issue. 3, pp.32–57, 1974

[7] J. Bezdek, *"Pattern Recognition with Fuzzy Objective FunctionAlgorithms"*, Springer, United States, pp. 1981.

[8] X. Y. Wang, J. Bu, *"A fast and robust image segmentation usingFCM with spatial information"*, Digital Signal Processing, Vol. 20, Issue. 4, pp. 1173–1182, 2010.

[9] W. Cai, S. Chen, D. Zheng, *"Fast and Robust Fuzzy C-Means Clustering Algorithms Incorporating Local Information for Image Segmentation"*, Pattern Recognition, Vol. 40, Issue. 3, pp. 825-838, 2007.

[10] S. Krinidis, V. Chatzis, *"A Robust Fuzzy Local Information C Means Clustering Algorithm"*, IEEE Trans Image Process, Vol. 19, Issue. 5, pp. 1328–1337, 2010.

[11] L. Szilagyi, Z. Benyo, S. Szilagyii, and H. Adam, *"MR brain image segmentation using an enhanced fuzzy C-means algorithm"*, In the Proceedings of 25th Annual International Conference of IEEE EMBS, Mexico, pp. 17–21, 2003.

[12] M. Ahmed, S. Yamany, N. Mohamed, A. Farag, and T. Moriarty,*"A Modified Fuzzy C-Means Algorithm for Bias Field Estimation and segmentation of MRI Data"*. IEEE Transactions on Medical Imaging, Vol. 21, Issue. 3,pp. 193–199, 2002.

[13] Z. Yuhui, B. Jeon, Q. M. J. Wu, *"Image segmentation by generalized hierarchical fuzzy C-means algorithm"*, Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, Vol. 28, Issue 2, pp. 961-973, 2015.

[14] J. Suckling et al., *"The mammographic image analysis society digital mammogram database"*, In the Proceeding of 2nd International Workshop on Digital Mammography, Elsevier Science, England, pp 375‑378, 1994.

Authors Profile

Ms. Manasi Hazarika completed Bachelor degree from Institute of Electronics and Telecommunication Engineers, India in 2007 and Master of Technology from Tezpur University, Assam in 2010. She is currently pursuing Ph.D. from Gauhati University and currently working as Assistant Professor (Sr.) in Department of Computer Science & Engineering and IT, Assam Don Bosco University, Assam since 2010. She is a life member of IETE society since 2007. Her research work focuses on Medical Image Processing. She has published research articles in reputed Journals and Conferences. She has a teaching experience of 8.5 years and has 10 years of research experience.

Dr. Lipi B. Mahanta is an Associate Professor working in the Central Computational and Numerical Sciences Division of Institute of Advanced Study in Science and Technology, Guwahati, Assam, India. She obtained her Ph.D. degree in Statistics in 2004 from Gauhati University, Guwahati, Assam. After collecting experience as a teacher of Computer Science for more than 20 years she started research work in different areas of health care like Medical Image processing and Pattern Recognition, Epidemiology and Development of Operations Research techniques with applications to health care, preferring to work specially on cancer. She has published more than 60 papers in national and international journals in these various fields.