Tumor Detection in Brain MRI Image Using Template based K-means and Fuzzy C-means Clustering Algorithm

Rasel Ahmmed

Department of Electronics and Communication Engineering Khulna University of Engineering & Technology Khulna-9203, Bangladesh E-mail: rasel kuet2k8@yahoo.com

Abstract—This paper presents a robust segmentation method which is the integration of Template based K-means and modified Fuzzy C-means (TKFCM) clustering algorithm that, reduces operators and equipment error. In this method, the template is selected based on convolution between gray level intensity in small portion of brain image, and brain tumor image. K-means algorithm is to emphasized initial segmentation through the proper selection of template. Updated membership is obtained through distances from cluster centroid to cluster data points, until it reaches to its best. This Euclidian distance depends upon the different features i.e. intensity, entropy, contrast, dissimilarity and homogeneity of coarse image, which was depended only on similarity in conventional FCM. Then, on the basis of updated membership and automatic cluster selection, a sharp segmented image is obtained with red marked tumor from modified FCM technique. The small deviation of gray level intensity of normal and abnormal tissue is detected through TKFCM. The performances of TKFCM method is analyzed through neural network provide a better regression and least error. The performance parameters show relevant results which are effective in detecting tumor in multiple intensity based brain MRI image.

Keywords-Magnetic resonance imaging (MRI); template based k-means and modified fuzzy c-means clustering (TKFCM); gray level intensity; coarse image; features selection; artificial neural network (ANN). Introduction (Heading 1)

I. INTRODUCTION

In recent decades, efficient detection of brain tumor is being stupendous challenge for medical science. Especially for Magnetic Resonance Imaging (MRI), it is quite concerning content since, MRI image rarely color image. MRI imaging technique has good contrast value over different technique. However, an appropriate segmentation of brain MRI image is apparent for detecting abnormality in brain. As brain comprehends complicated structure so segmentation of MRI image obliges good care and should be precise [1]. Segmentation describes salient image regions to procure region(s) of interest (ROI's) such as legions, tumors, edema, and necrotic tissues in brain image [2]. Md. Foisal Hossain

Department of Electronics and Communication Engineering Khulna University of Engineering & Technology Khulna-9203, Bangladesh E-mail: foisalkuet@yahoo.com

For brain image segmentation numerous image processing techniques have been proposed, for example- region growing, thresholding, classifiers, Artificial Neural Networking (ANN), clustering, and so on.

Segmentation of scalar images by creating a binary partition of the image intensities is done based on thresholding. Due to structure complexity of brain tissue, proper threshold value is very hard to achieve. The main drawback of thresholding is that, it cannot be applicable for multiple channel images. In addition, it does not provide spatial characteristics, which causes it to be sensitive to noise as well as inhomogeneity intensity [3]. On the other hand, foundation of restricting threshold is also used collectively [4-5] with other methods such as classifier, ANN, clustering etc. Based on some predefined criteria *i.e.* intensity information and/or edges, the connected region of an image is extracted in region growing. [6]. Besides, the precise anatomical information is needed to locate single or multiple seed pixels for each region and together with their associated homogeneity refers as region growing [7–9]. The primary limitation in the region growing is that its seed point is found through manual interaction. In the Classifier method, it needs a perfect pixel classifier for training data. Inefficient training data and classifier are aimed to time consumption and hilarious results [3]. In the clustering technique, Fuzzy c-means (FCM) clustering and expectationmaximization (EM) algorithms are being the most widely used methods for clustering [10]. The applications of the EM algorithm to brain MR image segmentation and a common disadvantage of EM algorithms are reported Wells et al. [11]. There is a FCM algorithm which contacts with a knowledgebased classification and tissue labeling used Li et al. [12]. Firstly, this FCM method segments MR brain images, and then introduces an expert system to locate a landmark tissue by matching them with a prior model. An ANN is used Hall et al. [13], and compared the performance with FCM for segmenting brain MR images. Conventional FCM Pham et al. [14] has limitation of noise sensitivity and imperfection to the abnormality of brain e.g. tumor, edema, and cyst. Although kmeans segmentation is noise immune, but it is prerequisite of this method that there should be perfect thresholding [15], which is quite hard for complex brain structure.



From the above discussion, it can be said that, there is no such technique which is perfect. But to get optimum results in this field we have to minimize the limitation of the individual methods describes here.

In this paper, we have proposed a different technique for tumor detection from brain MRI image, based on the combination of template based K-means and modified FCM (TKFCM). It will reduce the problem of template or gray level selection and noise sensitivity accustomed to FCM and, Region growing. In this paper, template based k-means extension has applied through pixel intensity positioning. Very little bit of pixel intensity is not avoided as the template is selected based on the number of grav level to be used and the coarse image. FCM algorithm is modified on the basis of updated membership and number of clusters for the filtered image. The membership values are updated based on the features such as intensity, entropy, contrast, and homogeneity of the MRI brain image. By choosing right membership input & output variables along with adjusted number of feature and cluster the detection of tumor is completed nicely. The whole performance is analyzed through neural network, which is highly accepted in the segmentation field. It is done by preparing train and target data for the relevant image through the network with feed forward back-propagation algorithm.

This paper is organized as follows. In section II conventional k-means and fuzzy c-means algorithm is explained. The proposed algorithm is described step by step in section III. The results and discussion is shown throughout in section IV and finally the whole work is concluded in section V.

II. K-MEANS AND FUZZY C-MEANS ALGORITHM

A. K-means algorithm

Historically, the K-mean clustering is normally introduced to group a set of data points $\{x_1, x_2, \ldots, x_N\}$ into K clusters [15]. It has high computational efficiency and can support multidimensional vectors. So it reduces the distortion measurement by minimizing a cost function as:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} b_{nk} || x_n - c_k ||^2$$
(1)
$$b_{nk} = \begin{cases} 1 & \text{if } k = \arg \min_a || x_n - c_a ||^2, a = 1, \dots, k \\ 0 & \text{Otherwise} \end{cases}$$

$$c_k = \frac{1}{N_k} \sum_{x \in C_k} x$$

Where, $|| \cdot ||$ measures distance from the center. The center and number of data points in the cluster C_k is represented by the variables c_k and N_k , respectively.

B. Fuzzy C-means Algorithm

The fuzzy c-means (FCM) clustering algorithm, introduced by Bezdek, is an improvement of earlier clustering methods [17]. It is based on minimizing an objective function, with respect to fuzzy membership set U of cluster centroids V:

$$J_m(U,V) = \sum_{j=1}^N \sum_{i=1}^C u_{ij}^m d^2(x_j, v_i)$$
(2)

In (2), $X = \{x_i, x_2, ..., x_j, ..., x_N\}$ is a $P \times N$ data matrix, where P represents the dimension of each x_j 'feature' vectors and N represents the number of feature vectors (pixel numbers in the image). C is the number of clusters. $U_{ij} \subseteq U(P \times N \times C)$ is the membership function of vector x_j to the *i*th cluster, which satisfies $U_{ij} \in [0 \ 1]$ and $\sum_{i=1}^{c} U_{ij} = 1$, (j=1,2,...,N). The membership function is expressed as:

$$U_{ij} = \sum_{k=1}^{C} \left(\frac{d(x_j, v_i)}{d(x_j, v_k)} \right)^{-\left(\frac{2}{m-1}\right)}$$
(3)

 $V = \{v_1, v_2, \dots, v_i, \dots, v_c\}$ which is a $P \times C$ matrix which denotes the cluster feature center is like:

$$V_{i} = \frac{\sum_{j=1}^{N} (U_{ij})^{m} \times j}{\sum_{j=1}^{N} (U_{ij})^{m}} \quad (i=1,2,...C)$$
(4)

 $m \in (1, \alpha)$ is a weighting exponent on each fuzzy membership, which controls the degree of fuzziness $d^2(x_j, v_i)$ and that is a measurement of similarity between x_i and v_i :

$$d^{2}(x_{j}, v_{i}) = || x_{j} - v_{i} ||^{2}$$
(5)

 $\|.\|$ can be defined as either a straightforward Euclidean distance or its generalization such as Mahalanobis distance [16]. The feature vector X in MR image represents the pixel intensity P=l. The FCM algorithm iteratively optimizes $J_m(U,V)$ with the continuous update of U and V, until $\|U_{ij}^{(l)} - U_{ij}^{(l+1)}\| \le \varepsilon, \varepsilon = \{0 \text{ to } 1\}$, where l is the number of iterations.

III. PROPOSED ALGORITHM

The proposed algorithm is integration of the k-means and fuzzy c-means with some modification. The template is added along with the conventional k-means, which is identified by the temper or gray level intensity in the brain image. Besides the fuzzy c-means membership and Euclidian distance is modified by the image features. Template based k-means and modified fuzzy c-means clustering algorithm for segmentation can be written in equation as below:

$$J = \sum_{i=i+1}^{M} \sum_{j=j+1}^{N} B(x_i, y_j) \times \sum_{i=1}^{K} \sum_{j=1}^{C} P_{ij} \parallel x_i - c_j \parallel^2 \times \sum_{j=1}^{K} \sum_{i=1}^{C} (U_{ij})^m d^2(x_j, v_i)$$
(6)

Where, M and N are the row and column of P_{ij} , a binary image matrix. The centroid of the cluster, number of data points in clusters and number of cluster is defined by R, K and C respectively. In eqn. (6) the last portion is defined as modified fuzzy c-means whose Euclidian distance is depended on the image features. The middle portion is used as the conventional k-means algorithm, which is defined by the distance from each point to cluster center. Here, $B(x_i, y_i)$ is the coarse image

which is marked in describing desired template could be found through below eqn. (7):

$$B(x_{i}, y_{j}) = \sum_{i=i+1}^{M} \sum_{j=j+1}^{N} P(x_{i}, y_{j}) \times T_{mn}$$
(7)

Template based window is selected by T_{mn} which is given as:

$$T_{mn} = \sum_{i=1}^{M} \sum_{j=1}^{N} P(x_i, y_j) \oplus \sum_{k=1}^{G} \sum_{l=1}^{S} P(x_k, y_l) \quad k \in M, l \in N$$
(8)

In this eqn. (8), there is an temper based image matrix with number of gray level intensity, G and number of bins, S which is used to detect the temper of the image $P(x_i, y_j)$. Here the convolution of temper based image matrix and image cause of obtaining the template for the k-means algorithm.

In the fuzzy C-means, U_{ij}^{m} is the membership function whose value updated with Euclidian distance d(x, v) which relies on the image features $F= \{F_1, F_2, ..., F_C\}$, degree of fuzziness *m*, and feature center $V=\{v_1, v_2, ..., v_L\}$ is expressed as:

$$U_{ij} = \left[\sum_{k=1}^{C} \left\{ \frac{d(x_j, v_i)}{d(x_j, v_k)} \right\}^{\frac{2}{m-1}} \right]^{-1}$$
(9)

In the previous research works, this Euclidian distance was used based on only one features for example similarity [18], but in our proposed method this relies on features like contrast, homogeneity, entropy etc.

Clusters center from where the clusters position and tumor are detected can be defined as:

$$V_{i} = \frac{\sum_{j=1}^{N} (U_{ij})^{m} \times j}{\sum_{i=1}^{N} (U_{ij})^{m}} \quad where, \ (i = 1, 2, ..., C)$$
(10)

We have extracted the following five features for the classifier as:

- a. Energy, $F_1 = \sum_{i=1}^{G} \sum_{j=1}^{G} |P(x_i, y_i)|^2$ Here, G is the gray level co-occurrence matrix.
- b. Contrast, $F_2 = \sum_{n=1}^{G} n^2 \sum_{i=1}^{G} \sum_{j=1}^{G} |P(x_i, y_i)|$ and |i j| = n

c. Homogeneity,
$$F_3 = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(i+j)(j+j)}{1+|i+j|}$$

d. Entropy,
$$F_4 = \sum_{i=1}^{G} \sum_{j=1}^{G} P(x_i, y_j) (-\ln(P(x_i, y_j)))$$

e. Dissimilarity,
$$F_5 = \sum_{i=1}^{O} \sum_{j=1}^{O} P(x_i, y_i) |i - j|$$

The TKFCM algorithm can be summarized as follow:

1. Define number of gray level and determine square matrix, $A = \sum \sum P(x_k, y_l)$ and image matrix,

$$P = \sum \sum P(x_i, y_j)$$

2. Initialize template, $T_{mn} = P \oplus A$

3. Determine coarse image, $B(x_i, y_j)$ from template, T_{mn}

4. Reshape template based k-means segmented image
$$\frac{K}{K} = \frac{C}{C}$$

$$P_{1} = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij} || x_{i} - C_{j} ||^{2} \times \sum_{i=i+1}^{n} \sum_{j=j+1}^{n} B(x_{i}, y_{j})$$

5. Repeat step 2 to 4 until
$$T_{mn} \le \sum_{k=1}^{\infty} \sum_{l=1}^{3} [T_{mn}(k) - T_{mn}(l)]$$

- $6. \qquad Post \ process \ the \ P_1$
- 7. Determine cluster centroid, C and degree of fuzziness, m.
- 8. Initialize membership $U_{ii}^{(0)}$ of FCM

9. Calculate cluster center,
$$v_i^{(l)} \Leftrightarrow U_{ij}^{(l)}, (i = 1, 2,C) \text{ and } (l = 0, 1, 2,)$$

10. Determine image features,
$$F(x_i, v_i^{(l)}) \Leftrightarrow v_i^{(l)}$$

11. Update
$$U_{ij}^{(l)}$$
 with $d(x_j, v_i^{(l)})$ until
 $\||U_{ii}^{(l)} - U_{ij}^{(l+1)}|| \le \varepsilon, \varepsilon = \{0 \text{ to } 1\}$

The whole method that has been proposed for the detection of tumor in brain MRI image using template based K-means and modified fuzzy C-means is described by the following flow chart. In this flow chart firstly there is manipulated acquisition of brain MRI image, and then it is processed and given at the input of template based k-means segmentation method. Finally from the modified fuzzy C-means with updated membership the detected tumor with red line marked is obtained. This is done through the clustered image which is automatically selected from the image features.



Figure 1. The flowchart of proposed TKFCM algorithm

IV. RESULTS AND PERFORMANCE ANALYSIS

There is a database of 30 brain tumor images shown in Fig. 2. We have made the database through different complex brain tumor image. These images are collected from [17-19] and processed for the betterment of application in our algorithm. Then, we have processed these images alongwith Matlab and get the database for final use shown in Fig. 2. The tumor in these image is so critical that it is difficult for the common people to identify it so easily.

In this TKFCM algorithm we have used 3, 5, 10, 27 number images of database to detect the tumor position which is shown in Fig. 3. A small portion of the brain tumor are not avoided. The input image is firstly processed through some filter then in Fig. 3(b) there is first segmentation of the image using template k-means (TK) which is segmented based on there gray level intensity and temper of color. After that the tumor is detected and marked it as red line in Fig. 3(c) using modified fuzzy c-means (FCM) algorithm on the basis of euclidian distance from cluster centre to each data point which mostly depends on the different features. This could be significant to understand the importance of this modified and incorporated method.

On the basis of gray level intensity the modified FCM is performed for 15 clusters. Clustered image is referred as the image with its smallest gray level and separated from each other with their consecutive color intensity. For example, several clustered images for input image no. 3 are shown in Fig. 4. Here, the tumor portion with other portion of the image are shown in separate image, from this on the basis of the feature the tumor is selected. Here the index no. 13 is selected automatically and marked as red line in Fig. 3.



Figure 2. The database of 30 brain tumor MRI image



Figure 3. (a) Input images for TKFCM, (b) Segmented images from the 1st segmented algorithm (TK-means) with no. of gray level=10, (c) Detected brain tumor images from the TKFCM.



Figure 4. Several clustered images for input image no. 3.



Figure 5. The Neural Network architecture

The criterion for choosing network was based on the performance of this technique which is shown in Fig. 5. Here the network consists of 30 input vector layers, one hidden layer with 17 neurons, and one output layer. Usually the

network is consists of three layers like input layer, hidden layer, and output layer. The input and output layer is user defined nevertheless hidden layer is selected on the basis of performance. Hidden layer can be one or more for the analysis, but it is required that, in hidden layer the no. of layer should be least numbered.

This is the performance curve shown in Fig. 6 of the trained data, for number of iterations 60, increment no. 0.05; performance goal 0.5e-02, minimum gradient 1e-10, and validation check 6. The performance goal is achieved between 10 and 12 iterations. Best validation performance is 1.5483 at iteration 11. The test data meets along with validation at its best values. The performance is depended on the error value or goal of least error. In this analysis we have used error value of 0.005 for the betterment of our designed network. It is required that the no. of iteration is least and for our network it has taken 11 iterations to meet the least error.

This Fig. 7 shows how the gradient decreasing and reached descent value at 12 iterations. Also Levenberg-Marquardt optimization parameter (MU) increases and decreases at its given values. Besides, there is two validation check failure. The value of gradient is reached at 6.6352 for the 12 epochs. Again the MU value is increased and decreased then proceeds to the value of 1e-09 at the 12 iterations. Finally there are 1 validation checks at the 12 epochs.



Figure 6. The performance curve for this method through neural network.



Figure 7. Validation check, minimum gradient, and Mu curve for the network.



Figure 8. Error histogram curve for the network.



Figure 9. The regression curve for the network.

The error histogram curve is described in Fig. 8. The horizontal axis is belongs to the errors and vertical lines describes the number of instances for this errors. Where, the training, validation, test, and zero error are present. The yellow line shows the zero error. The train error is belonging from - 0.7043 to 1.032 values. That means the data between this values are appropriately trained and there are these errors. The green values indicate validation errors and red data indicate the test errors.

The regression process of the training, validation, test and all data are preceded in the Fig. 9. This shows that the regression to training is 0.99983, to validation is 0.99342, to test is 0.99542, and throughout all is 0.99848 which is so good. The curve describes the fitting of data to the ideal regression line. The small circular values are our data. The dotted line is the ideal regression line. And the full straight line with RGB color is our regression curve.

There is some error rate on the basis of identifying or not identifying any abnormal tissue in all brain tumor MRI image. This can be calculated depending on the value of true positive, false positive, true negative, and false negative [20]. The performance parameters can be calculated for the 30 brain tumor images through following mathematical expression:

Sensitivity =
$$TP/(TP+FN)*100$$

Specificity =
$$TN / (TN + FP) * 100$$

$$Accuracy = (TP + TN) / (TP + TN + FN + FP) * 100$$

TP=True Positive, *TN*=True Negative, *FP*= False Positive, *FN*= False Negative

In TABLE I a comparison chart is prepared among the conventional brain tumor detection processes and our proposed method where it can be observed that TKFCM performs better than them. Some of them have accuracy better than us but the other parameters are not so good. Apart from other our method sensibly pursue the all parameters to a distinguished value, which is required for the betterment of the experts.

Again, the time required for detecting each tumor image is 40-50sec, in the Matlab2014a with Core2duo processor.

 TABLE I.
 : COMPARISON AMONG THE CONVENTIONAL METHODS & TKFCM METHOD

Algorithms	Sensitivity (%)	Specificity (%)	Accuracy (%)
Thresh-holding	84	80	83.3
Region Growing	88.46	75	86.7
Second order + ANN	91.42	90.1	92.22
Texture Combined +ANN	95.4	96.1	97.22
FCM	96	93.3	86.6
K-Mean	80	93.12	83.3
Proposed TKFCM	96.67	100	97.1

V. CONCLUSION

In this paper, we have proposed a new approach namely Template based K-means and modified fuzzy C-means clustering algorithm. It is used to remove the limitation of conventional K-means and conventional FCM algorithm for brain tumor MRI image. The template is selected based on convolution between gray level intensity in small portion of brain image and brain tumor image. K-means algorithm is to emphasized initial segmentation through the proper selection of template. Updated membership is obtained from the distance measurement from centroid to clusters, until it reaches to its best. On the basis of updated membership and automatic selected cluster, a sharp segmented image is obtained with tumor from modified FCM technique. The segmented tumor is shown as red marked with their proper detected position. The performance is analyzed through neural network, which shows better accuracy and least error. The accuracy, sensitivity, and specificity show that it is better than other previous conventional methods. Though it is less noise sensitive, but for some images where the gray level intensity difference is very small causes trouble to select perfect template.

REFERENCES

- S. Shen, W. Sandham, M. Granat, and A. Sterr, "MRI fuzzy segmentation of brain tissue using neighborhood attraction with neuralnetwork optimization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 3, pp. 459 – 467, September 2005.
- [2] N. Noreen, K. Hayat, and S. A.Madani, "MRI Segmentation through wavelets and Fuzzy C_Means,"World Applied Sciences Journal 13(Special Issue of Applied Math): pp.34-39,2011.
- [3] .D. L. Pham, C. Xu, and J. L. Prince, "Current Methods in Medical Image Segmentation," *Annual review on Biomedical E ngineering*, vol.2, pp.315-337, 2000.
- [4] L. Lemieux, G. Hagemann, K. Krakow, and F. G. Woermann, "Fast, accurate, and reproducible automatic segmentation of the brain in T1weighted volume MRI data," *Magnetic Resonance Medical*, vol. 42, pp. 127–135, 1999.
- [5] H. Suzuki and J. Toriwaki, "Automatic segmentation of head MRI images by knowledge guided thresholding," *Computed Medical Image Graphics*, vol. 15, no. 4, pp. 233–240, 1991
- [6] R.M. Haralick and L.G. Shapiro, "Image segmentation techniques" Computer Vision & Graphics Image Processing, vol.29, pp.100–132, 1985
- [7] A. R. Robb, Biomedical Imaging, Visualization, and Analysis. New York: Wiley, 2000.
- [8] R. Pohle and K. D. Toennies, "Segmentation of medical images using adaptive region growing," *Proceeding SPIE— Medical Imaging*, vol. 4322, pp. 1337–1346, 2001.
- [9] T. Y. Law and P. A. Heng, "Automated extraction of bronchus from 3D CT images of lung based on genetic algorithm and 3D region growing," *Proceeding SPIE — Medical Imaging*, vol. 3979, pp. 906–916, 2000.
- [10] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Transaction Medical Imaging*, vol. 18, no. 10, pp. 897–908, Oct. 1999.
- [11] W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Transaction Medical Imaging*, vol. 15, no. 4, pp. 429–442, Aug. 1996.
- [12] C. L. Li, D. B. Goldgof, and L. O. Hall, "Knowledge-based classification and tissue labeling of MR images of human brain," *IEEE Transaction Medical Imaging*, vol. 12, no. 4, pp. 740–750, Apr. 1993.
- [13] L. O. Hall, A. M. Bensaid, L. P. Clarke, R. P. Velthuizen, M. S. Silbiger, and J. C. Bezdek, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," *IEEE Transaction Neural Network.*, vol. 3, no. 5, pp. 672–682, Sep. 1992.
- [14] D. L. Pham and J. L. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Transaction Medical Imaging*, vol. 18, no. 9, pp. 737–752, Sep. 1999.
- [15] W. Narkbuakaew, H. Nagahashi, K. Aoki, and Y. Kubota, "Integration of Modified K-Means Clustering and Morphological Operations for Multi-Organ Segmentation in CT Liver-Images," *Recent Advances in Biomedical & Chemical Engineering and Materials Science*,pp.34-39,March.2014.
- [16] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, USA, 1981.
- [17] (2015) PE- Brain tumour website. [Online]. Available: http:// www.mayfieldclinic.com /
- [18] University of Maryland Medical Center (2015) homepage. Primary brain tumor [Online]. Available:

http://umm.edu/health/medical/reports/articles/brain-tumors-primary

- [19] Oncolgy (2015) homepage. Tumor type [Online]. Available: http://www.oncolink.org/types/article.cfm?c=52&aid=247&id=9534
- [20] X. Descombes, F. Kruggel, G. Wollny, and H.J. Gertz, "An object-based approach for detecting small brain lesions: Application to Virchow-robin spaces," *IEEE Transaction Medical Imaging*, vol.23, no.2, pp.246–255, 2004.