MRI Brain Image Segmentation for Spotting Tumors Using Improved Mountain Clustering Approach

Nishchal K. Verma[1] , Payal Gupta[2], Pooja Agrawal[1] and Yan Cui[3]

[1] Department of Electrical Engineering, Indian Institute of Technology, Kanpur, India.
[2]Department of Computer Science and Engineering, KEC Ghaziabad, U.P Tech. Univ, Luknow, India.
[3]Department of Molecular Sciences, Center for Integrative and Translational Genomics, University of Tennessee, Memphis, Memphis, TN 38163, USA.

***Abstract:* This paper presents improved mountain clustering technique based MRI (magnetic resonance imaging) brain image segmentation for spotting tumors. The proposed technique is compared with some existing techniques such as K-Means and FCM, clustering. The performance of all these clustering techniques is compared in terms of cluster entropy as a measure of information and also is visually compared for image segmentation of various brain tumor MRI images .The cluster entropy is heuristically determined, but is found to be effective in forming correct clusters as verified by visual assessment.**

Keywords: *Clustering, image segmentation, Magnetic Resonance Imaging, modified mountain clustering, Expectation Maximization, fuzzy clustering, validity function Cluster Entropy.*

I. INTRODUCTION

Image segmentation is widely used in exploratory pattern-analysis, grouping, decision-making, and machine-learning situations, including data mining, document retrieval and pattern classification. However, in many such problems, there is little a priori information (e.g., statistical models) available about the data, and the decision-maker must make as few assumptions about the data as possible. It is under these restrictions that clustering methodology is particularly appropriate for the exploration of interrelationships among the data points to make an assessment (perhaps preliminary) of their structure.

Segmentation is a tool that has been widely used in medical image processing and computer vision for a variety of reasons. The goal is to segment the medical images with respect to their certain properties such as color and texture. Segmentation is an [11] important tool in medical image processing and it has been useful in many applications. The applications include detection of the coronary border in angiograms, multiple sclerosis lesion quantification, surgery simulations, surgical planning measuring tumor volume and its response to therapy, functional mapping, automated classification of blood cells, studying brain development, detection of microcalcifications on mammograms, image registration, atlas-matching, heart image extraction from cardiac cineangiograms, detection of tumors, etc.

Clustering is basically the (automatic) grouping of similar objects such that:

- Each group or cluster is homogeneous; examples that belong to the same group are similar to each other.
- Each group or cluster should be different from other clusters, that is, examples that belong to one cluster should be different from the examples of other clusters.

The Clustering technique can be hard or fuzzy. A hard clustering algorithm allocates each pattern to a single cluster during its operation and in its output. A fuzzy clustering method assigns a degree of membership to each input pattern depending on its association with several clusters. A fuzzy clustering can be converted to a hard clustering by associating each pattern to the cluster with the largest membership.

All clustering techniques cannot uncover all the clusters present in the data with equal facility, because clustering algorithms often contain implicit assumptions about cluster shape or multiple-cluster configurations based on the similarity measures and grouping criteria used. Humans perform competitively with automatic clustering procedures in two dimensions, but most real problems involve clustering in higher dimensions. It is difficult for humans to obtain an intuitive interpretation of data embedded in a high-dimensional space. In addition, data hardly follow the “ideal” structures (e.g., hyper spherical, linear). This explains the reason behind the development of a large number of clustering algorithms in the literature.

In this paper we have present improved Mountain Clustering method and compared it with some existing techniques. It will be shown that the proposed technique is better in terms of optimum number of clusters with low computational complexity and high validity. Fuzzy C-Means gives good results in terms of time complexity and cluster validity, but the only constraint is that the number of clusters must be known apriori, which is difficult to obtain at times. In the proposed technique, after determining the first potential cluster center we will make a cluster around that point and remove the rest of the points from the dataset, thereby maintaining the potential of points. This improved mountain clustering technique gives better results in terms of cluster validity and time complexity. We are able to get all-relevant clusters by reducing the number of redundant clusters. All the clusters are demarcated with good performance with this technique.

This paper is organized into 4 sections. An overview of the existing clustering algorithms is given in Section 2. The proposed technique is explained in Section 3. In Section 4, we present the results of all the clustering techniques including the results of comparison of different techniques on the basis of cluster validity and computational complexity. Finally we conclude in the last section.

II. An Overview of Some Clustering Techniques

Here, we will briefly discuss some of the commonly used clustering techniques, such as K-Means Clustering, FCM Clustering [4], Gath-Geva Clustering [2], EM Clustering [8], Mountain Clustering [10] and Modified Mountain Clustering [9].

K-means clustering algorithm is proposed by Forgy and MacQueen in 1967. It is one of the simplest unsupervised clustering algorithms that has been widely used in various applications. Using K-means, data is clustered into a fixed number of clusters and centroid of one the clusters is placed as far away as possible from another. Each data point is associated to the nearest centroid. Fuzzy C-Means (FCM) clustering developed by Dunn in 1973 and improved by J. C. Bezdek in 1981, allows one piece of data to be in two or more clusters. FCM is frequently used in pattern recognition. In this technique, data points are bound to each cluster by means of membership functions, which give degree of association to the cluster. Gath-Geva Clustering is based on the minimization of the sum of weighted squared distances between the data points and the cluster centers. In Gath-Geva Clustering, the weighing component in the range [1, 0) determines the fuzziness of the resulting clusters. The Expectation-Maximization (EM) clustering algorithm is based on the statistical model that makes use of the finite Gaussian mixtures model. This clustering algorithm utilizes the K-means initialization, i.e. initially the fixed number of clusters is obtained through K-Means clustering. The cluster parameters (weights, means and co-variances) are re-computed until a desired convergence value is achieved. The finite mixture model assumes all attributes to be independent random variables. The Probabilistic Clustering technique aims at fitting k-component Gaussian mixture model (GMM) to the data set. It evaluates the probability of the components of a GMM such that all the points in the data set can be categorized into components with high probability. This clustering algorithm employs the basic Expectation-Maximization Algorithm [8]. Here, we explain the P-value in a mixture distribution. P-value parameter determines whether sample values lie in the region of low or high probability density. We will make use of the concept of significance level for D-variate normal distribution, in the probabilistic clustering algorithm for Gaussian mixtures. Yager and Filev [10] proposed a simple and easy to implement, Mountain Clustering algorithm for estimating the number and location of cluster centers. This algorithm is a grid based three step procedure. In the first step, the hyperspace is discretized with a certain resolution in each dimension so that grid points are obtained. The second step uses the data set to construct the mountain function around all grid points. The third step generates the cluster centers by an iterative destruction of the mountain function. Though this method is simple, the computation grows exponentially with the dimension of hyperspace. In the n-dimensional hyperspace with m number of grid lines in each dimension, the number of grid points that must be evaluated is m^n. To overcome the computational complexity of this clustering technique, Azeem et al. [9], presented the Modified Mountain Clustering technique which determines cluster centers by an iterative destruction of the mountain function. Destruction of the mountain function implies reduction of potential values of the data points, which are nearer to the cluster center than a threshold. This iterative reduction in potential of all the data points with respect to the cluster center leads to loss of certain potential clusters. Because of iterative reduction, potential of some of the data points which can become a potential cluster center reduces to a degree that they lose the potential to become cluster center and hence we loose the corresponding cluster. The useful feature of the Modified Mountain Clustering is that its computational complexity is independent of the dimension and eliminates the need to specify a grid resolution.

III Improved Mountain Clustering

A. The IMC Algorithm

The improvement in computational complexity of modified mountain clustering is achieved up to a greater extent since as we go on making the clusters from the dataset, we remove these from the original data set and consequently we achieve successively reduced computation time for each successive cluster. This we can realize by looking at Step 6 in the algorithm given below:

Step 1: Normalize each dimension of hyper-space, so that the data points are bounded by the unit hypercube.

We define the j^{th} data in $\mathbf{x}$ hyperspace as:

$$\mathbf{x}^j = \left\{ x_1^j, x_2^j, \ldots, x_D^j \right\}$$

the normalized data points $\bar{\mathbf{x}}^j$, are defined as:

$$\bar{\mathbf{x}}^j = \frac{\left\langle \mathbf{x}^j - (\mathbf{x}^j)_{\min} \right\rangle}{\left\langle (\mathbf{x}^j)_{\max} - (\mathbf{x}^j)_{\min} \right\rangle} \qquad (1)$$
$$\forall j = 1, 2, ...n$$

Where

$$(\mathbf{x}^j)_{\min} = \left\{ \min_{j=1}^{n} x_1, \min_{j=1}^{n} x_2, ..., \min_{j=1}^{n} x_D, \right\} \qquad (2)$$

$$(\mathbf{x}^j)_{\max} = \left\{ \max_{j=1}^{n} x_1, \max_{j=1}^{n} x_2, ..., \max_{j=1}^{n} x_D, \right\} \qquad (3)$$

and n is the total number of data points.

Step 2: Determine the potential of each data d_1 and d_2, for each window. d_1 and d_2 are the positive constants defining the neighborhood of the data point. We compute these from the heuristic:

$$d_1 = d_2 = \frac{1}{2n} \sum_{j}^{n} \left(\frac{\min(\mathbf{x}^j)}{\sum_{i=1}^{D} x_i^j} \right) \qquad (4)$$

Step 3: Calculate the potential value of each point using mountain function, which is a function of distance $d^2\left(\bar{\mathbf{x}}^r, \bar{\mathbf{x}}^j\right) = \left(\bar{\mathbf{x}}^r - \bar{\mathbf{x}}^j\right) Q \left(\bar{\mathbf{x}}^r - \bar{\mathbf{x}}^j\right)'$ between $\bar{\mathbf{x}}^r$ and all other data points.

$$P_1^r = \sum_{j=1}^{n} \exp\left[-\left(\frac{d^2\left(\bar{\mathbf{x}}^r, \bar{\mathbf{x}}^j\right)}{d_1^2} \right) \right] \qquad (5)$$
$$\forall r = 1, 2, ...n$$

Step 4: Select the first cluster center according to the highest value of P_1^r as:

$$\bar{c}_1 = \bar{\mathbf{x}}^{1^*} \Leftarrow P_1^* = \max_{r=1}^{n}\left(P_1^r\right) \qquad (6)$$

Step 5: **A**ssign those data points to the first cluster whose Euclidean distance from the first cluster center is less than a threshold, d_2 i.e.

If

$$d^2\left(\bar{\mathbf{x}}^r, \bar{c}_1\right) \le d_2; \quad \forall r = 1, 2, ...n \qquad (7)$$

Then $\bar{\mathbf{x}}^r$ is assigned to the first cluster.

Step 6: Remove all those data points from the dataset which are assigned to the first cluster.

Step 7: Repeat Steps 2 to Step 5 for the remaining or reduced dataset to make successive clusters. Similarly for selection of m^{th} cluster center, revision of potential value is done for the reduced dataset and m^{th} cluster center is selected with the highest value of P_m^r as under:

$$\bar{c}_m = \bar{\mathbf{x}}^{m^*} \Leftarrow P_m{}^* = \max_{r=1}^{n}\left(P_m^r\right)$$

Step 8: Determine the optimum number of clusters using the validity function S defined by,

$$S = \frac{\sum_{m=1}^{M} \sum_{r=1}^{n} \left(\bar{\mu}_m^r\right)^2 . \| \bar{\mathbf{x}}^r - \bar{c}_m \|^2}{n . \min_{i \ne j} \| \bar{c}_i - \bar{c}_j \|^2} \qquad (8)$$

where $\bar{\mu}_m^r = \frac{\mu_m^r}{\sum_{m=1}^{M} \mu_m^r}$

and the membership function μ_m^r represents the degree of association of the r^{th} data point to the m^{th} cluster center and is defined as

$$\mu_m^r = \exp\left[-\left(\frac{d^2\left(\bar{\mathbf{x}}^r, \bar{c}_m\right)}{d_2^2} \right) \right] \qquad (9)$$

The optimum number of clusters for a given dataset is decided by the validity function S, which is the ratio of compactness to separation. The compactness is related to the closeness of all other points with respect to a cluster center whereas the separation indicates the distance between the cluster centers.

Step 9: Form an optimum number of clusters M, using the Steps 2 to7 and then separate out these clusters from the whole dataset, rest of the data points are distributed among the formed clusters depending upon their Euclidean distance, i.e. nearness to the respective clusters.

B. Entropy as Quality Measure

Here we employ entropy as the quality measure of a cluster. Let E_m be the entropy of m^{th} cluster having N_m as the number of data point in cluster. The expression for m^{th} cluster is:

$$E_m = -\bar{p}_m \log(\bar{p}_m) \qquad (10)$$
$$\text{For } m = 1, 2, .., M$$

Taking, $\bar{p}_m = 1 - p_m$ is found to be consistent vis-à-vis visual assessment of the clusters. Here p_m constructed heuristically as follows

$$p_m = \left(\frac{\rho_m}{N_m^2} \right)$$

Where,

$$\rho_m = \sum_{j=1}^{N_m} \left(\frac{\max\limits_{d=1}^{3}(x_d^j)}{\sum\limits_{d=1}^{3} x_d^j} \right)$$

represents the cumulative response dominant feature in a vector. As will be demonstrated in the result section, the proposed entropy measure is found to distinguish the clusters based on the shades of their colors.

IV RESULTS AND DISCUSSIONS

A. Examples

Example.1: In this example, we consider segmentation of a brain tumor MRI image with resolution 40x30 using K-means, FCM and the Improved mountain clustering techniques. The performance of these techniques is shown in Table-I. In this example IMC scores over all others followed by the FCM clustering, in terms of the average entropy measure. Table II shows the segments of the image by all the above mentioned clustering techniques

Example.2: Here, we deal with a MRI image of brain tumor again, with resolution 40x30. The Improved mountain clustering yields least average entropy, Followed by the FCM its close rival and the K-means clustering technique. Table III gives the performance and Table IV shows the clusters.

Example.3: In this example, we consider the segmentation of a brain tumor MRI image with resolution 40x30. Table V shows the performance where we can see FCM, K-means and the Improved mountain clustering all have almost the same average entropy value. Clusters are shown in Table VI.

Example.4: Here, we have a MRI brain tumor image for segmentation with resolution 40x30. Table VII shows the performance and Table VIII gives the clusters. Here we can see FCM, K-means and the Improved mountain clustering all have almost the same average entropy value.

Example 5: Here also we deal with an MRI image of brain tumor with resolution 40x30. The Improved mountain clustering yields least average entropy followed by the FCM and the K-means clustering technique. Table IX gives the performance of all the clustering techniques and Table X shows the clusters obtained from all the clustering techniques.

B. Comparison of Performance

The widely used clustering algorithms discussed above were tested on various MRI (magnetic resonance imaging) images. Simulations were done using Matlab on a Core 2 Duo machine with 1.50 GHz Processor and 0.99 GB RAM. Prior to the experiments, no pre-processing was done on these images. RGB features were used in clustering. The effectiveness of clusters is compared in terms of Entropy of each cluster and all clusters.

The average Cluster entropy values for improved mountain clustering are least in most cases. The time complexity of the improved technique is also less. Results show that the Improved Clustering technique is able to retrieve all the clusters. The cluster centers in the case of Fuzzy C-Means are widely separated. It is able to retrieve all basic clusters, giving less redundant clusters. The disadvantage with Fuzzy C-Means is that it is sensitive to selection of initial partitions and may converge to a local minimum of the criterion function as can be seen from the results of various images. The proposed technique gives best results with the tested images and has less time complexity, since it reduces the number of patterns examined during execution.

TABLE I

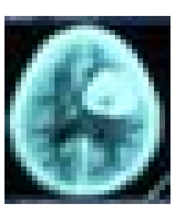
Original Image / Optimum no. of Clusters=3	Clustering Method	1st Cluster	2nd Cluster	3rd Cluster	Average Entropy
	K-means	0.001248	0.000825	0.001010	0.001028
	FCM	0.001251	0.000832	0.000997	0.001027
	IMC	0.001191	0.000818	0.001063	0.001024

TABLE II

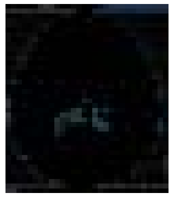
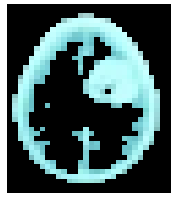
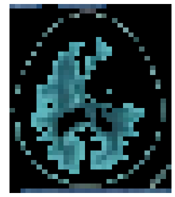
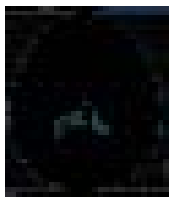
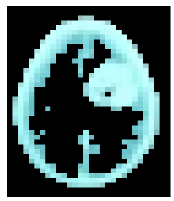
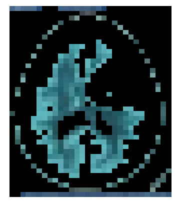
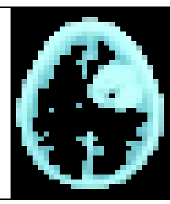
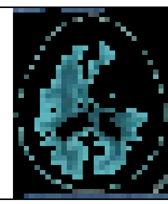
Clustering Method	1st Cluster	2nd Cluster	3rd Cluster
K-means			
FCM			
IMC			

TABLE III

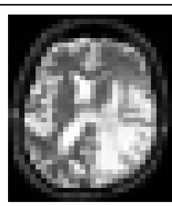
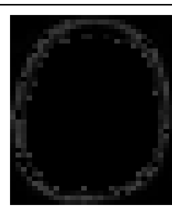
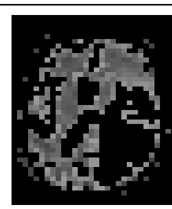
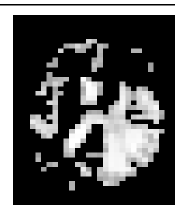
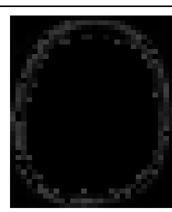
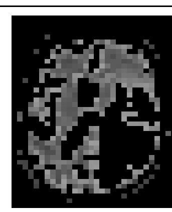
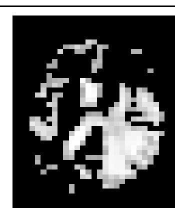
Original Image / Optimum no. of Clusters=3	Clustering Method	1st Cluster	2nd Cluster	3rd Cluster	Average Entropy
	K-means	0.000493	0.000898	0.001129	0.000840
	FCM	0.000497	0.000875	0.001145	0.000839
	IMC	0.000504	0.000863	0.001125	0.000831

TABLE IV

Clustering Method	1st Cluster	2nd Cluster	3rd Cluster
K-means			
FCM			

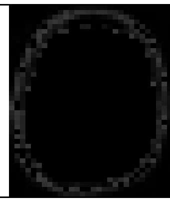
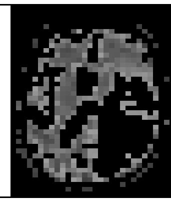
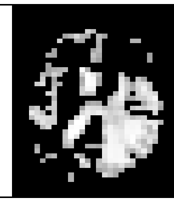
IMC			

TABLE V

Original Image / Optimum no. of Clusters=2	Clustering Method	1st Cluster	2nd Cluster	Average Entropy
	K-means	0.000623	0.000607	0.000615
	FCM	0.000625	0.000605	0.000615
	IMC	0.000632	0.000599	0.000615

TABLE VI

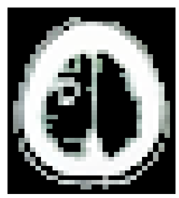
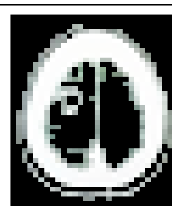
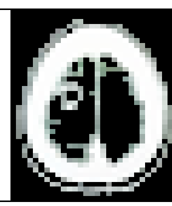
Clustering Method	1st Cluster	2nd Cluster
K-means		
FCM		
IMC		

TABLE VII

Original Image / Optimum no. of Clusters=2	Clustering Method	1st Cluster	2nd Cluster	Average Entropy
	K-means	0.000622	0.000676	0.000649
	FCM	0.000620	0.000678	0.000649
	IMC	0.000620	0.000678	0.000649

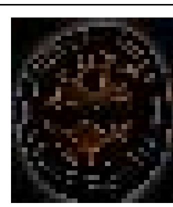
TABLE VIII

Clustering Method	1st Cluster	2nd Cluster
K-means		
FCM		
IMC		

TABLE IX

Original Image	Clustering Method	1st Cluster	2nd Cluster	Average Entropy
Optimum no. of Clusters=2				
	K-means	0.000611	0.000741	0.000676
	FCM	0.000613	0.000738	0.000675
	IMC	0.000621	0.000727	0.000674

TABLE X

Clustering Method	1st Cluster	2nd Cluster
K-means		
FCM		

IMC		

V. Conclusions

This paper compares the performance of a few clustering techniques in terms of entropy measure, viz, K-Means, Fuzzy C-Means, for the segmentation of the MRI images. During implementation, we have found that the the Improved mountain clustering fares over others in view of less computational complexity and speed of operation. It is found that FCM and K-means trails behinds IMC performance wise though they are equally competitive at times. The performance of this Improved Mountain Clustering is found to be the best. Visual assessment has also confirmed our findings through experiments on MRI images.

References

[1] Mohammad Fazle Azeem, Madasu Hanmandlu and Nesar Ahmad,”Structure Identification of Generalized Adaptive Neuro-Fuzzy Inference Systems”, *IEEE Trans. on Fuzzy Systems*, Vol II, No 5,Oct 2003

[2] Janos abonyi, Robert Babuska, Ferene Szeifert, ”Modified Gath-Geva Clustering for Identification of Takagi-Sugeno Fuzzy Models”, *IEEE Trans. on Systems, Man, and Cybernetics-partB: cybernetics*, Vol.32, no5, October 2002

[3] de Oliveira, M.C., Kitney, R.I.*”* 7 Texture analysis for discrimination of tissues in MRI data**”,** *IEE Conf. On Computers in Cardiology 1991*. Proceedings. , 23-26 Sept. 1991 Page(s): 481 –484.

[4] Razaz M, “A fuzzy C-means clustering placement algorithm”, .Circuits and Systems, 1993., ISCAS '93, 1993 *IEEE International Symposium*, 3-6 May 1993 Page(s): 2051 -2054 vol.3.

[5] A K Jain, M.N.Murthy, P.J.Flynn,”Data Clustering: A Review”, *ACM Computing Surveys*, Vol. 31,No. 3,September 1999.

[6] Gan Ming Tao, A .Tan, M.Hanmandlu, “The Model-Selection Problem”, submitted to *IEEE Trans. on Fuzzy Systems.*

[7] Richard O.Duda, Peter E.Hart, David G.Stork ,*Pattern Classification*, Second Edition ,John Wiley & Sons ,Inc.

[8] A.P. Dempster, N.M. Laird, and D.B. Rubin,” Maximum likelihood from incomplete data via the EM algorithm,”*J.Roy Stat.Soc. B*, Vo. 39, pp.1-38, 1977.

[9] Azeem, M. F., Hanmandlu, M., and N. Ahmad , ”Modified mountain clustering and dynamic fuzzy modeling”, *2^{nd} Int. Conf. Inform. Techno.* , Bhubaneshwar, India, 63-68.

[10] Yager R. R., Filev D. P. , ”Approximate clustering VIA the mountain method”, *IEEE Trans. on System, Man and Cybernatics,* Vol 24, No 8, Aug. 1994 page(s):1279-1284.

[11] Isaac N. Bankman, “Handbook of medical imaging processing and analysis”, *Academic press*, United States.