

# Gray Code Formulated Split-Merge Algorithm for Determining Number of Clusters in FCM

V.Royna Daisy<sup>1</sup> and Dr.S.Nirmala<sup>2</sup>

AP, Department of EEE, Paavai Engineering College, Namakkal, India

Principal, Muthayammal Engineering College, Namakkal, India

**ABSTRACT:** Fuzzy C- Means (FCM) clustering is a potential method, mainly for data clustering and image segmentation. But the major limitation of Fuzzy C Means and its derived methods is that, the number of clusters is either randomly initialized or prior knowledge of the image is needed. Though there are numerous works done in this context, most of them are distance based leading to increase in the computational time and complexity. A fully automated algorithm developed for determining the number of clusters based on Gray code is presented in this paper. Experimental are done on brain Magnetic Resonance Images and results are produced. Performance evaluation based on segmentation accuracy and computational time is done and compared with other methods.

**KEYWORDS:** Fuzzy C-Means, FCM based splitting algorithm, Gray Code,

## I. INTRODUCTION

Image Denoising, Segmentation, Feature Extraction and Classification are the important steps in image processing. Enormous works have been done on these areas Magnetic Resonance Imaging (MRI) is a vital diagnostic procedure in practice. The most crucial step in the process of image processing is image segmentation. Different image segmentation algorithms have been developed, yet each method has its own advantages and limitations. Fuzzy is an important area which can be used in image processing. In Denoising Fuzzy rules exploits the fuzzy connectivity of the pixel intensity as its tuning parameter to distinguish edge and background region in image [1]. The Fuzzy C- Means (FCM) clustering, one of the powerful clustering techniques shown in literature is a popular unsupervised classification method and has found many applications in pattern classification and image segmentation [2-5]. The clustering algorithms highly depend on the parameter setting and initialization. Although FCM is prominent for image segmentation, the results are mainly affected by the number of clusters and the initialization of the centroids. If the number of clusters is not set to the number of natural clusters or the centroids are initialized randomly, the clustering results would be unreliable and inconsistent. Several techniques have been developed for automatic segmentation using FCM [6-9].

There are small numbers of algorithms that determine the cluster number in the literature. A variety of self splitting-merging clustering algorithms have been developed for determining the number of clusters. Ya-Jun Zhang and Zhi-Qiang Liu proposed a self splitting algorithm for clustering [10]. Sun et.al proposed a model based selection algorithm which involves high computational complexity as the splitting involves calculation of distance of all data vectors in the cluster with the remaining cluster centers [11]. Cheng-Ru Lin and Ming-Syan Chen has proposed a combinatorial splitting algorithm with cohesion merging [12]. Splitting one cluster at a time will lead to high computational time. Hence a multiple splitting algorithm was proposed by Liu.et.al [13]. Miin-ShenYang et.al and RuiFa et.al have proposed automatic detection of cluster numbers. But these methods involve distance calculation which increases the computational time of FCM algorithm [14,15]. Nirmala et.al proposed the concept of binary connectedness for computing the similarities of a pixel to all the other in the neighborhood to distinguish the edges and clusters which can be used for segmentation using FCM clustering. [16].

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

The proposed method makes use of Gray code (a reflected binary code) assignment techniques which concentrates on avoiding the distance calculation which is commonly done in most of the proposed methods. The most fascinating part of Gray code is that two successive values differ by a single bit [17]. This interesting concept is made use of for splitting and merging of clusters. Experiments are done on the proposed method with brain MRI and the results are compared with computational time.

## II. BASIC FCM ALGORITHM

FCM has its own meritorious place in the field of image segmentation and pattern recognition. FCM, each image pixel has certain membership degree associated with each cluster centroid. These membership degrees have values in the range, indicating the strength of the association between that pixel and a particular cluster centroid. The FCM algorithm attempts to partition every image pixel into a collection of the fuzzy cluster centroids by minimizing the weighted sum of squared error objective function. The Objective function of FCM is

$$\text{Minimize } J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c u_{ki}^m \|x_k - v_i\|^2 \quad (1)$$

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, updating of membership  $u_{ij}$  and the cluster centers  $c_i$

$$\text{The membership function } u_{ki} = \frac{1}{\sum_{j=1}^c \left( \frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{m-1}}} \quad (2)$$

$$\text{The cluster centers } v_i = \frac{\sum_{k=1}^n u_{ki}^m x_k}{\sum_{k=1}^n u_{ki}^m} \quad (3)$$

The basic FCM algorithm is as follows

- (1) Input the number of clusters  $c$ , the fuzzifier  $m$  and the distance function
- (2) Initialize the cluster centers  $v_i^0$  ( $i = 1; 2; \dots; c$ ).
- (3) Calculate  $u_{ki}$  ( $k = 1; 2; \dots; n; i = 1; 2; \dots; c$ ) using Eq.(2).
- (4) Calculate  $v_i^1$  ( $i = 1; 2; \dots; c$ ) using Eq. (3).
- (5) If  $\max_{1 \leq i \leq c} \left( \frac{\|v_i^0 - v_i^1\|}{v_i^1} \right) \leq \epsilon$ , then go to Step 6;  
else let  $v_i^0 = v_i^1$  ( $i = 1; 2; \dots; c$ ) and go to Step 3.
- (6) Output the results

## III. PROPOSED FCM BASED SPLIT AND MERGE ALGORITHM

FCM based splitting algorithm [11] operates by splitting the worst cluster at each stage for clusters from  $C_{\min}$  to  $C_{\max}$ . The choice of  $C_{\min}$  and  $C_{\max}$  is predefined arbitrarily, and the choice of initial cluster centre is done randomly. The proposed work primarily focus to make the method fully automatic, by dividing the image or data into  $2^n$  triangular regions, where  $n=1,2,3\dots$ . These triangular regions are the initial clusters with their respective centroids as cluster centers. The centroid of a triangle is the point at which the medians meet. The initial cluster centers are,  $[(x_1+x_2+x_3)/3, (y_1+y_2+y_3)/3]$  where  $(x_1, y_1), (x_2, y_2)$  and  $(x_3, y_3)$  are the vertices of the triangle. By assigning cluster centre in this way it is sure that the cluster centers are far apart from each other. The value of 'n' is chosen based on the complexity of the data set.

The data set is divided into  $2^n$  triangular clusters with their medians as centers. The number of clusters is increased and /or decreased at each phase of Split Merge algorithm. The flowchart of the proposed work is shown in Fig.1. The worst cluster with sparse distribution is identified and splitted whereas the nearest clusters with close proximity are merged into a single cluster. The cluster centers are modified and the new cluster centers are used to continue the FCM algorithm.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

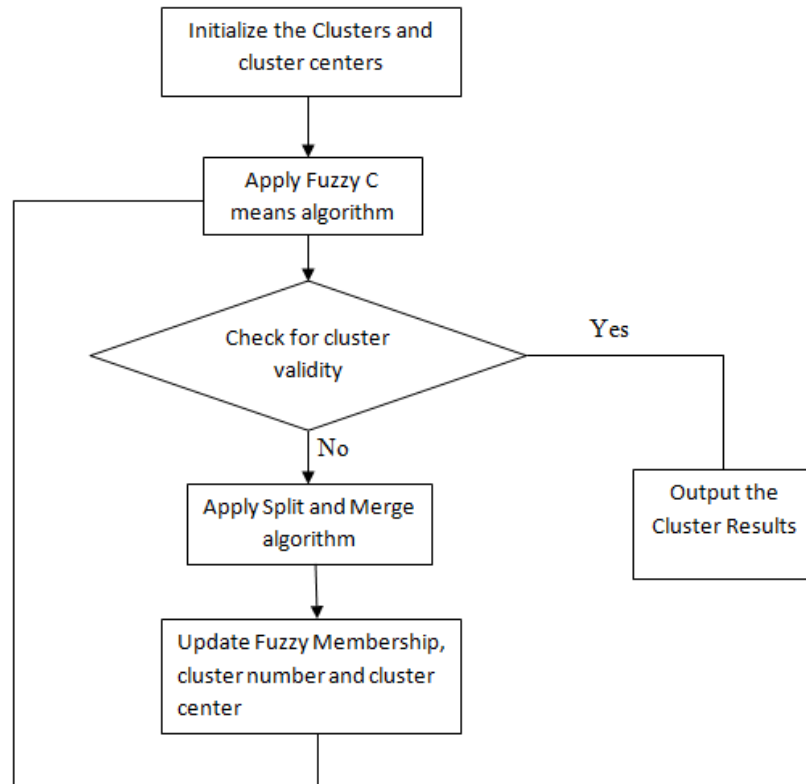


Fig 1. Flowchart of the proposed work

### 3.1 FCM Based Modified Split and Merge Algorithm

1. Choose cluster  $C=2^n$ , where  $n=1,2,3 \dots$
2. Initialize the C cluster center by finding the median of each triangle cluster
3. For C update the membership matrix(U) and cluster centers (V) by applying the basic FCM algorithm
  - 3.1. Test for Convergence. If convergence is reached then go to step 6 else go to step 4
4. Compute the Score function S(i)

$$S(i) = \frac{\sum_{k=1}^n u_{ki}}{\text{number of data vectors in cluster } i}$$

5. Apply the split and merge algorithm for the clusters with minimum and maximum S(i) respectively
  - 5.1. Find the cluster  $C\{\min(S(i))\}$ . Apply the split algorithm if  $0.9 \geq \frac{|C_1|}{|C_2|} \geq 1.1$  and update  $C=C+1$  else C is unchanged.
  - 5.2. Find the cluster  $C\{\max(S(i))\}$ . Apply the Merge algorithm to merge the cluster with near largest S(i) cluster if  $0.9 \leq \frac{|C_1|}{|C_2|} \leq 1.1$  and update  $C=C-1$ , else C is unchanged.
6. The number of clusters, cluster centers and membership functions are given as output

### 3.2 Gray Code

Gray code is a binary number system, where two successive values differ in only one bit (binary digit). The gray code was originally designed to prevent spurious output from electromechanical switches. Nowadays, Gray codes are widely used to facilitate error correction in digital communications. In the proposed system gray codes are made used for splitting and merging of image clusters. The images are clustered into regions based on the pixel intensity. To implement this, the pixel intensities of images are divided into 8 regions and assigned with their

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

respective Gray codes. The Most Significant Bit (MSB) of the gray code indicates maximum pixel intensity (0- black dominates, 1- white dominates). The Least Significant Bit (LSB) indicates the minimum pixel intensity. Table 1 shows the Gray code assignment of pixel intensity. Since the Gray codes vary by one bit in the successive numbers, this interesting concept is made to split clusters when they differ in their MSB and merge two clusters when their MSB remain same with their LSB varying or not. The pixels are considered as straight kernels [18] and slid over the entire image.

**Table.1. Gray code assignment for pixel intensity**

Gray Code	Range of Pixel Intensity
<b>000</b>	<b>0-0.125</b>
<b>001</b>	<b>0.126-0.25</b>
<b>011</b>	<b>0.26-0.375</b>
<b>010</b>	<b>0.376-0.5</b>
<b>110</b>	<b>0.51-0.625</b>
<b>111</b>	<b>0.626-0.75</b>
<b>101</b>	<b>0.76-0.875</b>
<b>100</b>	<b>0.875-1</b>

### 3.3 Split Algorithm

The score function  $S(i)$  elucidate the distribution of data vector of the cluster 'i' and the membership value. Low  $S(i)$  means that large number of data vectors with low membership values are available in the cluster 'i' and this cluster is identified as the cluster to be split. Assign gray code for all the pixels in the cluster. Group of data ( $D_i$ ) which differs in the MSB of the gray code is searched. This group D will be taken as the splitted region. If there is more than one group that differs in the MSB the largest group,  $\max(D_i)$  is taken as the splitted cluster.

### 3.4 Merge Algorithm

The cluster with high  $S(i)$  is identified as the cluster to be merged. Identify the near largest  $S(i)$  cluster and check whether the cluster has the same code assignment or a gray code with respect to the LSB. The merging is done with respect to LSB because LSB indicates minimum pixel intensity. If the two clusters are of the same code or a gray with respect to LSB, then merge the two clusters.

## III. RESULTS AND DISCUSSION

The proposed Gray code formulated cluster determination technique has been tested over brain MRI images. Here the Gray code assignment is done considering the image divided into eight distance regions. In cases of more accuracy or more complex images, it can be divided into 16 or 32 regions. It is seen that proper segmentation has been done. Fig 2 shows Gray assignment of a region which is splitted based on the split algorithm. It is noted that the pixels which vary in their MSB are grouped separately whereas the pixels varying in the LSB are kept in the same cluster. Fig 3 shows the Gray code assignment of two regions which has high  $S(i)$ . The two nearest clusters which have the same gray code are merged to form a single cluster.

000	000	000	000	110	110	110	100	000	000	000	000	110	110	110	100
000	000	000	110	110	110	110	110	000	000	000	110	110	110	110	110
000	001	001	110	110	110	110	000	000	001	001	110	110	110	110	000
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000

**Fig.2 Gray code assignment and splitting of regions**

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

000	000	000	001
000	000	000	000
000	000	110	000
000	001	001	110

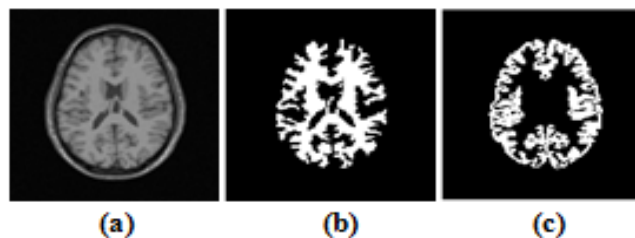
000	111	000	001
000	000	111	000
000	000	000	000
000	001	001	000

**Fig.3 Gray code assignment of two regions to be merged**

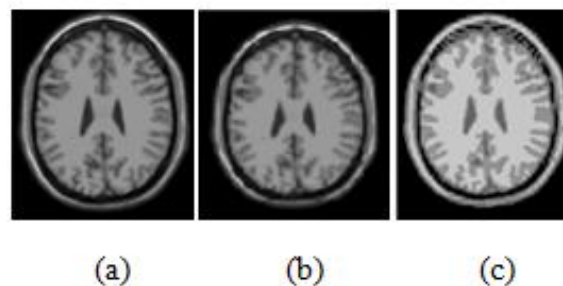
000	000	000	001	000	111	000	001
000	000	000	000	000	000	111	000
000	000	110	000	000	000	000	000
000	001	001	110	000	001	001	000

**Fig.4 Regions merged with gray code assignment**

In case there are noisy pixels in the image as shown in Fig.4 it is seen that during splitting or merging of clusters, the presence of noisy pixels (110 and 111) are neglected as they do not form any cluster and appear as individuals. Thus the segmentation accuracy is improved by Gray code splitting and merging. The segmentation result of the proposed method that separates the white matter and Cerebro Spinal Fluid (CSF) is shown in Fig 5. The segmentation is done unsupervised without any prior knowledge of the image and the results obtained are same as that obtained using FCM.



**Fig.5 a) Original Image, b) Segmented White matter and c) CSF**



**Fig.6 a) Original Image, b) Segmented by FBSA and c) Segmented by proposed method**

Segmentation of brain MRI by FCM-Based Model Selection Algorithms (FBSA) as proposed in [11] and proposed method is shown in Fig.6. It is seen that good segmentation is provided in the proposed method.

The segmentation results of different algorithms are measured based on a comparison score [19] defined as

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

$$S = \frac{|A \cap A_{ref}|}{|A \cup A_{ref}|} \tag{17}$$

where A is the set of all pixels belonging to a class based on the algorithm performed and A<sub>ref</sub> is the reference cluster pixels. The Comparison scores of the three methods for different noise level is shown in table 2

**Table 2. Computational time for various methods**

	Noise	FCM	FBSA	Proposed Method
Brain data	0%	94.23	93.03	94.10
	5%	92.04	90.56	93.18
	10%	91.43	89.29	92.28

Comparison scores show that FCM gives higher segmentation accuracy in case of high Signal to Noise Ratio (SNR) signals. But the prior information of the number of clusters is required for proper segmentation. As the SNR decreases the performance of FCM segmentation decreases. However in the proposed method noisy pixels are also taken into consideration during splitting and merging and so the segmentation accuracy increases.

It is seen that the number of iterations of the proposed method is far less than that required for FBSA. The computational overhead is reduced as the distance calculation of all the pixels in the cluster from the remaining centroids is reduced at each iteration level. This shows the efficiency of the proposed system in the context of computational time.

**Table 3. Computational time for various methods**

	FCM	FBSA	Proposed
Computational time (seconds)	2.56	3.23	2.78

The computational time for image segmentation by various methods is tabulated in Table 3. FCM shows less computational time, but it does not work well for noisy images. Also it requires an initial estimate of number of clusters and cluster centers. If any of the above said goes wrong it will affect the segmentation results. FBSA does not require initial cluster number as it starts the clustering considering minimum two clusters. But the computation of actual number of clusters is based on a distance calculation approach thus increasing the computational time. This limitation is reduced in the proposed method thereby reducing the computational time. Also by using Gray code some reduction of noise has been shown thereby increasing the segmentation accuracy. Though the time consumption is more when compared with FCM, it requires the number of clusters to be known in advance which has been solved in the proposed work.

## IV. CONCLUSION

Automatic determination of number of clusters is proposed based on Gray code formulation. Experiments are performed on brain MRI images and the results are produced. Gray code formulation avoids the computation of distance from all the data vectors to the remaining cluster centers during splitting or merging of clusters as done in most of the previous methods, thereby reducing the computation time.

## REFERENCES

[1] S Nirmala, S Sridevi, "Modified rayleigh maximum likelihood despeckling filter using fuzzy rules", *International Conference on Information Communication and Embedded Systems (ICICES)*, 755-760, 2013

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

- [2] AWC. Liew, Leung SH, Lau WH., "Fuzzy image clustering incorporating spatial continuity", *IEEE Proceedings-Vision, Image and Signal Processing*, Vol. 147, pp.185-192, 2000.
- [3] LK. Szeto, AWC. Liew, H. Yan, SS.Tang, "Gene expression data clustering and visualization based on a binary hierarchical clustering framework", Special issue on "Biomedical Visualization for Bioinformatics", *Journal of Visual Languages and Computing*, Vol. 14, pp.341-362, 2003.
- [4] S. H. Wu, A. W.-C. Liew, H. Yan, and M. S. Yang, "Cluster analysis of gene expression data based on self-splitting and merging competitive learning,"*IEEE Trans. on Information Technology in Biomedicine*, vol. 8, no. 1, pp. 5-15, 2004
- [5] J.C. Bezdek, Pattern Recognition in Handbook of Fuzzy Computation. IOP Publishing Ltd., Boston, Ny, 1998 (Chapter F6).
- [6] DL. Pham, JL.Prince, "Adaptive fuzzy segmentation of magnetic resonance images", *IEEE Trans. Medical Imaging*, Vol.18, pp.737-751, 1999.
- [7] E. A. Zanaty, S. Aljahdali, N. Debnath, "A kernelized fuzzy c-means algorithm for automatic magnetic resonance image segmentation", *Journal of Computational Methods in Science and engineering (JCMSE)*, pp. 123-136, 2009.
- [8] Z. Yu, O. C. Au, R. Zou, W. Yu and J.Tian, "An adaptive unsupervised approach toward pixel clustering and color image segmentation", *Pattern Recognition*, vol. 43, no. 5, (2010), pp. 1889-1906
- [9] Maoguo Gong, Yan Liang, Jiao Shi, Wenping Ma, and Jingjing Ma," Fuzzy C-Means Clustering With LocalInformationandKernel MetricforImageSegmentation",*IEEETransactionsOnImageProcessing*,Vol.22,No.2,2013
- [10] Ya-Jun Zhang and Zhi-Qiang Liu, "Self-splitting competitive learning: a new on-line clustering paradigm,"*IEEE Trans. Neural Networks*, Vol.13, no. 2, pp. 369-380, 2002.
- [11] HaojunSun ,Shengrui Wang, Qingshan Jiang , 'FCM-Based Model Selection Algorithms for Determining the Number of Clusters', *Pattern Recognition* 37, 2027 – 2037,2004
- [12] Cheng-Ru Lin and Ming-Syan Chen, "Combining partitional and hier-archical algorithms for robust and efficient data clustering with cohe-sion self-merging,"*IEEE Trans. Know. and Data Eng.*, vol. 17, no. 2, pp. 145-159, 2005
- [13] Jun Liu, KotagiriRamamohanarao , "Multiple Self-Splitting and Merging Competitive Learning Algorithm", *Advances in Knowledge Discovery and Data Mining , Lecture Notes in Computer Science Volume 4426*, 2007, pp 704-711
- [14] Miin-ShenYang ,Chien-Yo Lai," A Robust Automatic Merging Possibilistic Clustering Method", *IEEE Transactions on Fuzzy Systems (TFS)* (Volume:19, Issue: 1 ) Feb. 2011, 26 – 41
- [15] RuiFa, Asoke K Nandi,'SMART: Novel Self Splitting-Merging Clustering Algorithm', *20th European Signal Processing Conference (EUSIPCO 2012)*, Bucharest, Romania, August 27 - 31, 2012
- [16]P. John Vivek ,C.Ganesh, S.Sridevi, S.Nirmala, "Binary Connectedness Based RML Filter for Speckle Reduction in Ultrasound Images", *IJCA Proceedings on International Conference on Innovations In Intelligent Instrumentation, Optimization and Electrical Sciences*, ICIIIOES, pp6-12, 2013
- [17] Morris Mano M., "Digital Design", (Pearson Education, India), 2002 (Third edition)
- [18]S. Rajalaxmi, S. Nirmala, Entropy-Based Straight Kernel Filter for Echocardiography Image Denoising, *Journal of Digital Imaging*, Volume 27, Issue 5, pp 610-624, 2014
- [19]Masulli F, Schenone A. A fuzzy clustering based segmentation system as support to diagnosis in medical imaging. *ArtifIntell Med* 1999;16(2):29-47.