ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
*A Peer Reviewed Open Access International Journal*

# Modified FCM Algorithm for Abnormal Image Segmentation

**Mr.M V Ramanaih, Ph.D**
Professor,
MLR Institute of Technology.

**Mr.S.V.S Prasad, Ph.D**
Associate Professor,
MLR Institute of Technology.

**Mutyala Sudha Rani**
M.Tech Student,
MLR Institute of Technology.

## ABSTRACT:

Modified FCM algorithm is based on the concept of data compression where the dimension of the input is highly reduced. The data compression includes two steps quantization and aggregation. The modified FCM algorithm is efficient when compared with the conventional algorithm. The modified FCM algorithm yields superior convergence rate besides yielding nominal segmentation efficiency. Clustering approach is widely used in biomedical applications particularly for brain tumour detection in abnormal magnetic resonance (MR) images. Fuzzy clustering using fuzzy C-means (FCM) algorithm proved to be superior over the other clustering approaches in terms of segmentation efficiency. But the major drawback of the FCM algorithm is the huge computational rate is improved by modifying the cluster centre and membership value updating criterion. In this paper, the application of modified FCM algorithm for MR brain tumour detection is explored. Abnormal brain images from four tumour classes namely metastases, meningioma, glioma and astrocytoma are used in this work. A comprehensive feature vector space is used for the segmentation technique. Comparative analysis in terms of segmentation efficiency and convergence rate is performed between the conventional FCM and the modified FCM. Experimental results show superior results for the modified FCM algorithm in terms of the performance measures.

## Keywords:

Clustering, Modified Fuzzy C- means, Segmentation efficiency, Convergence rate.

## INTRODUCTION:

Image segmentation plays a major role in the field of biomedical applications. The segmentation technique is widely used by the radiologists to segment the input medical image into meaningful regions. The specific application of this technique is to detect the tumor region by segmenting the abnormal MR input image.

The size of the tumor region can be tracked using these techniques which aid the radiologists in treatment planning[1]. The primitive techniques are based on manual segmentation which is a time consuming process segmentation which is a time consuming process besides being susceptible to human errors. Several automated techniques have been developed which removes the drawbacks of manual segmentation. Clustering is one of the widely used image segmentation techniques which classify patterns in such a way that samples of the same group are more similar to one another than samples belonging to different groups [2]. There has been considerable interest recently in the use of fuzzy clustering methods, which retain more information from the original image than hard clustering methods. Fuzzy C-means algorithm is widely preferred because of its additional flexibility which allows pixels to belong to multiple classes with varying degrees of membership. But the major operational complaint is that the FCM technique is time consuming [3]. Several modifications have been done on the existing network to improve the performance.A hierarchical FCM algorithm based on template matching is proposed by Kwon & Han [4]. But it suffers from the drawback of the requirement for an accurate template. The segmentation efficiency of the FCM algorithm is improved by silhouette method based cluster center initialization instead of random initialization [5]. The FCM algorithm has been also implemented using the concept of parallel processing [6]. Even though it promises high speed processing, the hardware implementation is not effective. Cheng and Goldgof [7] proposed the fast clustering algorithm based on random sampling which yields a speed-up factor of 2-3 times when compared with the conventional FCM algorithm. The vector quantization based FCM algorithm has been implemented and a nominal speed-up factor is achieved [8]. Fast fuzzy clustering for web documentation which is highly robust is proposed in the literature [9]. In [10], the authors utilized visualization in conjunction with automated clustering to speed up the process of partitioning the data A scalable, parallel approach to clustering has been explored for a shared architecture. Eschrich and Ke implemented the clustering algorithm through quantization and aggregation which includes a weight factor for cluster centerupdation.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
*A Peer Reviewed Open Access International Journal*

In fuzzy clustering, every point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be in the cluster to a lesser degree than points in the center of cluster. An overview and comparison of different fuzzy clustering algorithms is available.[1] Any point x has a set of coefficients giving the degree of being in the kth cluster wk(x). With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

$$c_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}.$$

The degree of belonging, wk(x), is related inversely to the distance from x to the cluster center as calculated on the previous pass. It also depends on a parameter m that controls how much weight is given to the closest center. The fuzzy c-means algorithm is very similar to the k-means algorithm:

## PROPOSED METHOD:

Clustering can also be thought of as a form of data compression, where a large number of samples are converted into a small number of representative prototypes or clusters.

High dimensional feature space based image segmentation is time intensive than in one dimensional feature spaces. The modified FCM algorithm is based on the concept of data compression where the dimensionality of the input is highly reduced. The data compression includes two steps: quantization and aggregation. The quantization of the feature space is performed by masking the lower ‘m’ bits of the feature value.

The quantized output will result in the common intensity values for more than one feature vector. In process of aggregation, feature vector which shares common intensity values are grouped together. A representative feature vector is chosen from each group and they are given as input for the conventional FCM algorithm.

Once the clustering is complete, the representative feature vector membership values are distributed identically to all members of the quantization level. Since the modified FCM algorithm uses a reduced dataset, the convergence rate is highly improved when compared with the conventional FCM. A sample operation for the quantization and aggregation techniques with m=2 is given in Table 1 and Table 2.

**Table 1: Quantization Technique**

| Feature vector | Feature value | Binary equivalent | Quantized binary value | Quantized decimal value |
|---|---|---|---|---|
| [A1,B1,C1] | [252 100 60] | [1 1 1 1 1 1 0 0<br>0 1 1 0 0 1 0 0<br>0 0 1 1 1 1 0 0] | [1 1 1 1 1 1 0 0<br>0 1 1 0 0 1 0 0<br>0 0 1 1 1 1 0 0] | [252 100 60] |
| [A2,B2,C2] | [253 101 63] | [1 1 1 1 1 1 0 1<br>0 1 1 0 0 1 0 1<br>0 0 1 1 1 1 1 1] | [1 1 1 1 1 1 0 0<br>0 1 1 0 0 1 0 0<br>0 0 1 1 1 1 0 0 | [252 100 60] |
| [A3,B3,C3] | [192 89 12] | [1 1 0 0 0 0 0 0<br>0 1 0 1 1 0 0 1<br>0 0 0 0 1 1 0 0] | [1 1 0 0 0 0 0 0<br>0 1 0 1 1 0 0 0<br>0 0 0 0 1 1 0 0] | [192 88 12] |
| [A4,B4,C4] | [195 91 15] | [1 1 0 0 0 0 1 1<br>0 1 0 1 1 0 1 1<br>0 0 0 0 1 1 1 1] | [1 1 0 0 0 0 0 0<br>0 1 0 1 1 0 0 0<br>0 0 0 0 1 1 0 0] | [192 88 12] |

In the above table, A, B and C represents the features contrast, correlation and entropy respectively. For a 256*256 image, there are 65,536 feature vectors. For simplicity, the operation is shown in the above table with four feature vectors. Each vector consists of three feature values. Initially the binary equivalent is found out and the bit mask (11111100) is used to quantize the data.

The last column represents the quantized data where some feature vectors share common values which aid in data reduction. Table 2 illustrates the aggregation process.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

**Table 2: Aggregation Technique**

| Feature vector | Mean value |
|---|---|
| [A1A2, B1B2, C1C2] | [252.5 100.5 61.5] |
| [A3A4, B3B4, C3C4] | [193.5 90 13.5] |

In the aggregation process, the feature vectors sharing the common values are grouped together and their mean value is calculated. These mean values form a new feature vector which is the representative for the group. Similarly, representatives are taken from each group which forms a new dataset Y. This reduced dataset Y is used instead of the original dataset X in the conventional FCM algorithm. Once the clustering is complete, the representative feature vector membership values are distributed identically to all members of the quantization level. Since the dimensionality of the input dataset is reduced, the convergence rate is highly improved in the modified FCM algorithm.An improved fuzzy c-means algorithm is put forward and applied to deal with meteorological data on top of the traditional fuzzy c-means algorithm. The proposed algorithm improves the classical fuzzy c-means algorithm (FCM) by adopting a novel strategy for selecting the initial cluster centers, to solve the problem that the traditional fuzzy c-means (FCM) clustering algorithm has difficulty in selecting the initial cluster centers.

Introduction clustering analysis plays an important role in the data mining field, it is a method of clustering objects or patterns into several groups. It attempts to organize unlabeled input objects into clusters or "natural groups" such that data points within a cluster are more similar to each other than those belonging to different clusters, i.e., to maximize the intra-cluster similarity while minimizing the inter-cluster similarity. In the field of clustering analysis, a number of methods have been put forward and many successful applications have been reported. Clustering algorithms can be looselycategorized into the following categories: hierarchical, partition-based, density-based, grid-based and model-based clustering algorithms. Among them, partition-based algorithms which partition objects with some membership matrices are most widely studied. Traditional partition-based clustering methods usually are deterministic clustering methods which usually obtain the specific group which objects belong to, i.e., membership functions of these methods take on a value of 0 or 1.

We can accurately know which group that the observation object pertains to. The probability of the observation object being a part of different groups, which reduces the effectiveness of hard clustering methods in many real situations. For this purpose, fuzzy clustering methodswhich incorporate fuzzy set theory have emerged. Fuzzy clustering methods quantitatively determine the affinities of different objects with mathematical methods, described by a member function, to divide types objectively. Among the fuzzy clustering method, the fuzzy c-means (FCM) algorithm is the most well-known method because it hasthe advantage of robustness for ambiguity and maintains much moreinformation than any hard clustering methods. The algorithm is an extension of the classical and the crisp k-meansclustering method infuzzyset domain. It is widely studied and applied in pattern recognition, image segmentation and image clustering, data mining, wireless sensor network and so on.The modified FCM algorithm uses the same steps of the conventional FCM except for the change in the cluster updating and membership value updating criterions. The modified criterions are shown below:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} y_j}{\sum_{j=1}^{n} u_{ij}^{m}} \quad ; \quad u_{ij} = \frac{1}{\sum_{k=1}\left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}$$

where $d_{ij} = y_j - c_i$

$y$ = reduced dataset

The Improved FCM Algorithm for Meteorological Data According to the discussion about the traditional FCM algorithm in Section 2, the initial condition of cluster centers influences the performance of the algorithm. The best choice of the original cluster centers needs to consider the features of the data set. In this paper, meteorological data is chosen as our experimental data. Meteorological data is different from other experiment data. If we just use the traditional FCM algorithm to deal with the meteorological data, there will be a large error when clustering a certain object. To solve the initialization problem, we put forward an improved FCM algorithm in term of selecting the initial cluster centers. Nowadays, there are several methods to select the original cluster centers. In the following section we will go through some commonly used methods.

(1) Randomly the traditional FCM algorithm determines initial cluster centers randomly. This method is simple and generally applicable to all data but usually causes local minima.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

(2) user-specified International Journal of Database Theory and Application Vol.6, No.6 (2013) 10 Copyright 2013 SERSC Normally, users decide original cluster centers by some priori knowledge. According to the understanding of the data, users always can obtain logical cluster centers to achieve the purpose of the global optimum.
(3) Randomly classify objects into several clusters, compute the center of each cluster and determine them as cluster centers more time consumption is spent when randomly classifying objects in this method. When the number of objects in data sets is very small, the cost of time can be ignored. Nevertheless, as the number of objects increases, the speed of the increasing cost of time can be largely rapid.

(4) Select the farthest points as cluster centers Generally speaking, this method selects initial cluster centers following to the maximum distance principle. It can achieve high efficiency if there are no outliers or noisy points in data sets. But if the data sets contain some outliers, outliers are easier to be chosen as the cluster centers.

(5) Select points with the maximum density the number of objects whose distance is less than the given radius r from the observed object is defined as the density of the observed object. After computing the density of each object, the object whose density is the largest is chosen as the cluster center. Then compute densities of objects whose distances are larger than the given distance d from the selected centercenters, also choose the object whose density is the largest as the second cluster center. And so forth until the number of cluster centers reaches the given number. This method ensures that cluster centers are far away from each other to avoid the objective function into local minima. In our paper, we adopt a new method to determine cluster centers which is based on the fifth method as mentioned above. In our method, we first randomly select the observed object and compute the density of the observed object. If the density of the observed object is not less than the given density parameter, the observed object can be seen as the cluster center. Secondly we keep selecting the second cluster center satisfying the above constraints in the data set which excludes the objects which are cluster centers or objectswhose distances are less than the given distance parameter. Finally we obtain the given number of cluster centers after repeating the above process. The distance parameter and the density parameter are decided by users according to the characteristics of the data sets and the priori knowledge.

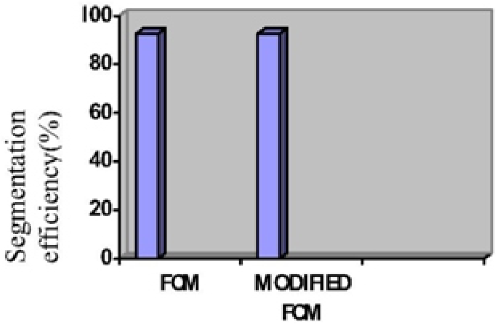



**Fig 5.1 segmentation efficiency**

A better compression can be achieved if more than two bits are changed in the bit mask which would further improve the convergence rate. Even though high dimensional feature space accounts for high efficiency, it significantly reduces the compression ratio which ultimately slows down the training process.

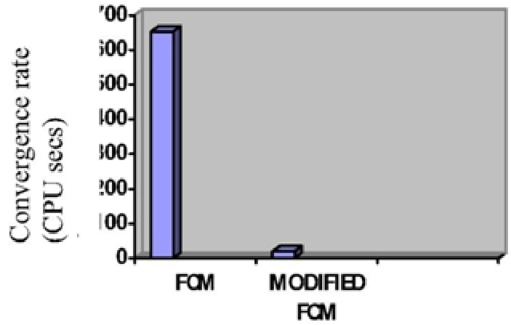



**Fig 5.2 Bar chart representation of performance measures.**

From the bar chart representation, it is evident that modified FCM algorithm is efficient when compared with the conventional algorithm. The modified FCM algorithm yields superior convergence rate besides yielding.

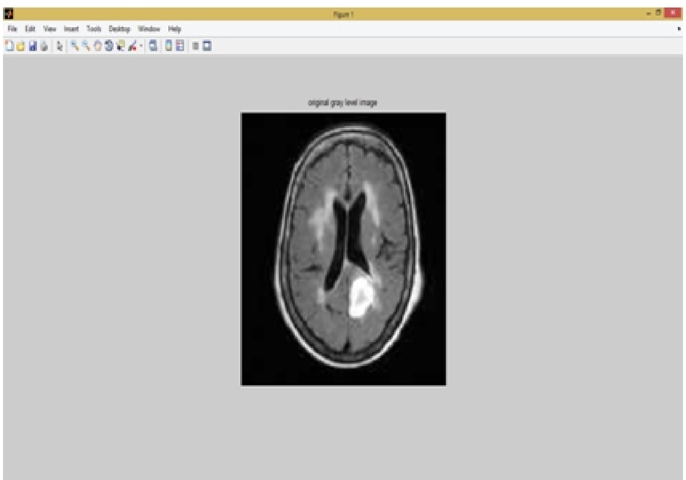
ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

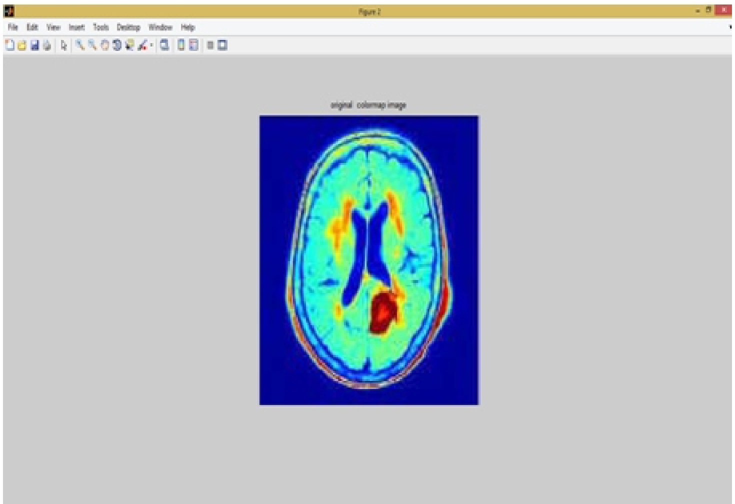
### Performance measures of FCM and modified FCM techniques:

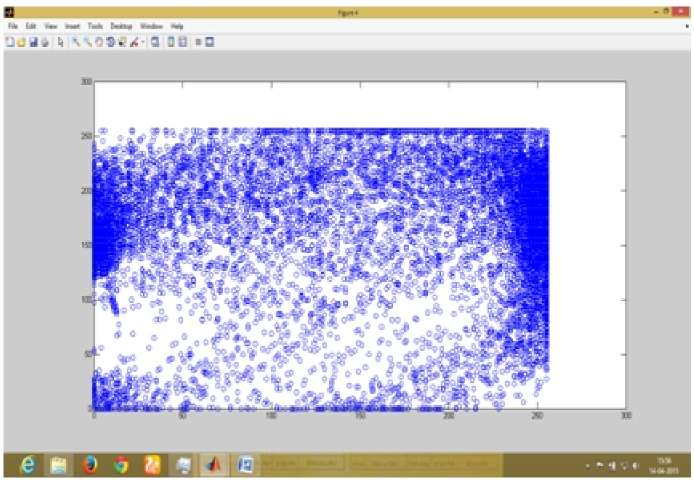
| Segmentation Technique | Number of iterations. |
|---|---|
| Conventional FCM algorithm | 54 |
| Modified FCM algorithm | 40 |

Table outlines the segmentation efficiency for both the techniques which does not reveal any significant improvement in the modified FCM algorithm. Table also reveals the superior nature of the modified FCM technique over the conventional FCM algorithm in terms of convergence rate. Since the reduced dataset is used in the modified FCM algorithm, significant improvement is achieved over the conventional algorithm.

A better compression can be achieved if more than two bits are changed in the bit mask which would further improve the convergence rate. Even though high dimensional feature space accounts for high efficiency, it significantly reduces the compression the compression ratio which ultimately slows down the training process.

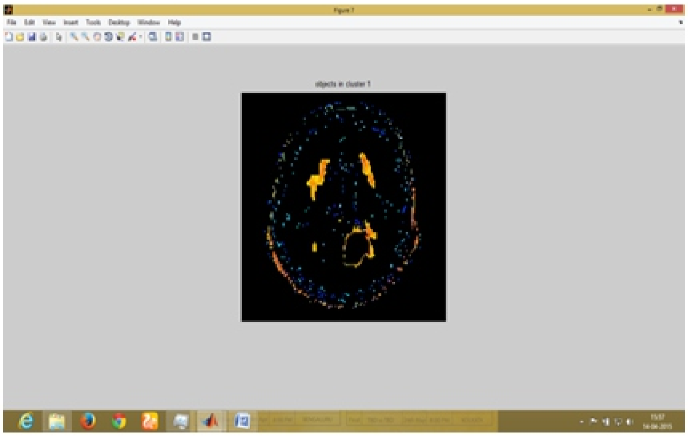


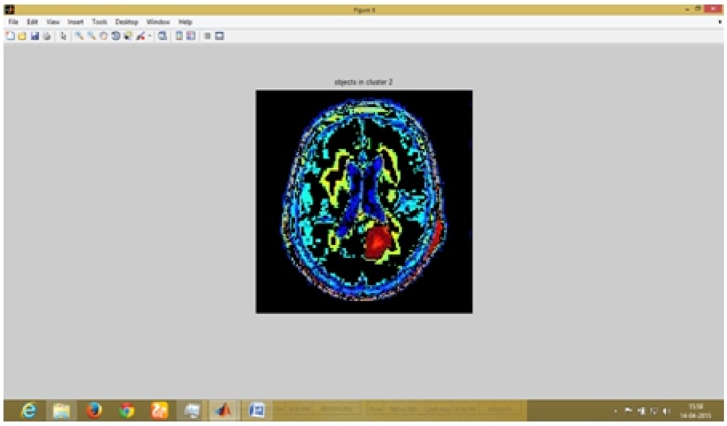


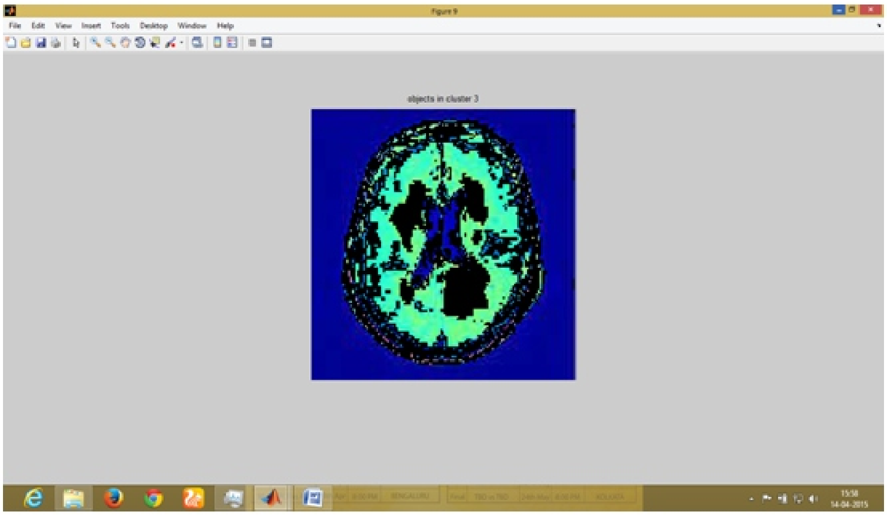


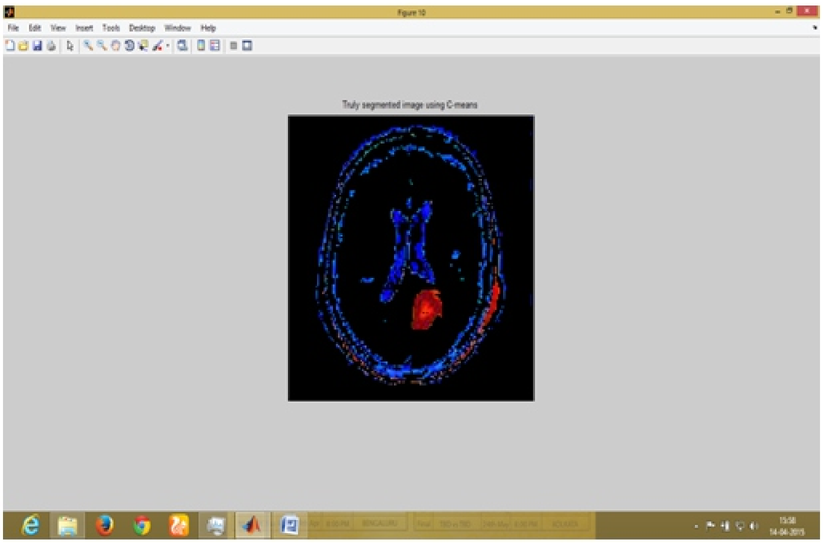
ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
*A Peer Reviewed Open Access International Journal*

## CORRELATED IMAGE



ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

## CONCLUSION AND FUTURE WORK:

Average speed-ups of as much as 80 times a traditional implementation of FCM are obtained using the modified FCM algorithm, while yielding segmentation efficiency that are equivalent to those produced by the conventional technique. Thus, the modified FCM algorithm is a fast alternative to the traditional FCM technique. As a future work, the network performance can be analysed with different bit mask and 'm' value along with different textural features.Thus we have successfully applied Modified fuzzy logic algorithm to enhance the medical images, color photographs and finger prints. Better results are obtained from the proposed method. So in future work there is scope of applying the algorithm to enhance video images.

## REFERENCES:

1. .Norbert Rohrl , Jose Iglesias and GaliaWeidl,"computer Assisted Classification of brain tumors", Springer berlin Heidelberg.

2.Wen , P, Zherg, Zhou,J,"spatialcredibilistic clustering algorithm I noise image segmentation". IEEE International conference on Industrial Engineering and Engineering Management.

3.JieYu,PeihuangGuo, Pinxiang Chen, Zhongshan Zhang and WenbinRuan, "remote sensing image classification based on improved fuzzy c-means", Geo-spatial information science.

4.S R kannan,"segmentation of MRI using new unsupervised fuzzy C-means algorithm".

5.DzungPhan and Jerry Prince,"An adaptive fuzzy C-means algorithm for image segmenatation in the presence of intensity inhomogenetics"

6. R.M Haarlick," statistical and structural approaches to texture", Proceedings of the IEEE.

7. J Ke,"Fast accurate fuzzy clustering through reduced precision", Master's Thesis. University of south Floride, 2000.

8. S R Kannan, "segmentation of MRI using new unsupervised fuzzy C-means algorithm", ICGST-GVIP journel, 2005

9. "J K Han andH M Kim. "

10. Aggarwal, Charu C., Han, Jiawei, Wang, Jianyong, Yu, Philip S., 2003. A framework for clustering evolving data streams. In: Proc. 29th Internat. Conf. on Very Large Data Bases, pp. 81–92.

11. Agrawal, Rakesh, Gehrke, Johannes, Gunopulos, Dimitrios, Raghavan, Prabhakar, 1998. Automatic subspace clustering of high dimensional data for data mining applications. In: Proc. ACM SIGMOD, pp. 94–105.

12. Anderberg, M.R., 1973. Cluster Analysis for Applications. Academic Press.

13. Andrews, Nicholas, O., Fox, Edward A., 2007. Recent developments in documentclustering.Technical report TR-07-35.Deparment of Computer Science, Virginia Tech.

14. Arabie, P., Hubert, L., 1994. Cluster analysis in marketing research. In: Advanced Methods in Marketing Research. Blackwell, Oxford, pp. 160–189.

15. Backer, Eric, 1978. Cluster Analysis by Optimal Decomposition of Induced Fuzzy Sets. Delft University Press.

16. Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X., 2006. Group formation in large social networks: Membership, growth, and evolution. In: Proc. 12th KDD.

17. Baldi, P., Hatfield, G., 2002. DNA Microarrays and Gene Expression.Cambridge University Press.

18. Ball, G., Hall, D., 1965. ISODATA, a novel method of data anlysis and pattern classification. Technical report NTIS AD 699616. Stanford Research Institute, Stanford, CA.

19. Banerjee, Arindam, Merugu, Srujana, Dhillon, Inderjit, Ghosh, Joydeep. 2004.

20. Clustering with bregman divergences. J. Machine Learn. Res., 234–245.

21. Banerjee, Arindam, Basu, Sugato, Merugu, Srujana, 2007a. Multi-way clustering on relation graphs. In: Proc. 7th SIAM Internat. Conf. on Data Mining.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
*A Peer Reviewed Open Access International Journal*

22. Banerjee, S., Ramanathan, K., Gupta, A., 2007b. Clustering short texts using Wikipedia. In: Proc. SIGIR.

23. Bar-Hillel, Aaron, Hertz, T., Shental, Noam, Weinshall, Daphna, 2003. Learning distance functions using equivalence relations. In: Proc. 20th Internat. Conf. on Machine Learning, pp. 11–18.

24. Basu, Sugato, Banerjee, Arindam, Mooney, Raymond, 2002. Semi-supervised clustering by seeding. In: Proc. 19th Internat. Conf. on Machine Learning.

25. Basu, Sugato, Bilenko, Mikhail, Mooney, Raymond J., 2004. A probabilistic framework for semi-supervised clustering. In: Proc. 10th KDD, pp. 59–68.

26. Basu, Sugato, Davidson, Ian, Wagstaff, Kiri (Eds.), 2008. Constrained Clustering: Advances in Algorithms, Theory and Applications. Data Mining and Knowledge Discovery, vol. 3, Chapman & Hall/CRC.

27. Bekkerman, Ron, El-Yaniv, Ran, McCallum, Andrew, 2005. Multi-way distributional clustering via pairwise interactions. In: Proc. 22nd Internat. Conf. Machine Learning, pp. 41–48.

28.Belkin, Mikhail, Niyogi, Partha, 2002. Laplacianeigenmaps and spectral techniques for embedding and clustering.Advances in Neural Information Processing Systems, vol. 14. pp. 585–591.

29. Ben-David, S., Ackerman, M., 2008. Measures of clustering quality: A working set of axioms for clustering. Advances in Neural Information Processing Systems.

30. Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms.Plenum Press.

31. Bhatia, S., Deogun, J., 1998. Conceputal clustering in information retrieval.IEEE Trans. Systems Man Cybernet. 28 (B), 427–436.

32.Bishop, Christopher M., 2006. Pattern Recognition and Machine Learning.Springer. Blake, Merz C.J., 1998. UCI repository of machine learning databases.

33. Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Machine Learn. Res. 3, 993–1022.

34. Bradley, P.S., Fayyad, U., Reina, C., 1998. Scaling clustering algorithms to large databases. In: Proc. 4th KDD.

35. Buhler, J., 2001. Efficient large-scale sequence comparison by locality-sensitive hashing. Bioinformatics 17 (5), 419–428.

36. Busse, Ludwig M., Orbanz, Peter, Buhmann, Joachim M., 2007. Cluster analysis of heterogeneous rank data. In: Proc. 24th Internat. Conf. on Machine Learning, pp. 113–120.

37. Cao, F., Ester, M., Qian, W., Zhou, A., 2006. Density-based clustering over an evolving data stream with noise. In: Proc. SIAM Conf. Data Mining.

38. Chapelle, O., Schölkopf, B., Zien, A. (Eds.), 2006. Semi-Supervised Learning.MIT Press, Cambridge, MA.

39. Cheng, Yizong, Church, George M., 2000. Biclustering of expression data. In: Proc. Eighth Internat. Conf. on Intelligent Systems for Molecular Biology, AAAI Press, pp. 93–103.

40. Connell, S.D., Jain, A.K., 2002. Writer adaptation for online handwriting recognition.IEEE Trans. Pattern Anal. Machine Intell. 24 (3), 329–346.

41. Critchlow, D., 1985. Metric Methods for Analyzing Partially Ranked Data. Springer.Datta, Ritendra, Joshi, Dhiraj, Li, Jia, Wang, James Z., 2008. Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys 40 (2) (Article