Comparitive Analysis Of K Means And Fuzzy C Means Algorithm

Poonam fauzdar and Sujata Kindri

ABSTRACT

Clustering or grouping a set of objects is a key procedure for image processing .It is an unsupervised technique that is used to arrange pattern data into clusters .This research work deals with two of the most representative clustering algorithms namely centroid and crisp values based K-Means and represent object based on calculation of membership function. Fuzzy C-Means are described and analyzed for a standard database images and also for coloured images. Based on experimental results the algorithms are compared regarding their clustering quality and their performance, which depends on the time complexity between the various numbers of clusters chosen by the end user. The total elapsed time to cluster all the images and Clustering time for each cluster are also calculated in seconds and the results compared with one another.

Key Words: *K-Means Algorithm, Fuzzy C-Means Algorithm, Image segmentation, cluster Analysis , fuzzy logic , Unsupervised learning.*

INTRODUCTION

Segmentation is a process that partition a single image into multiple clusters. In recent scenario, growing attention is towards data clustering as robust technique for data analysis applied both for natural and biomedical images[1][2]. Clustering or data grouping describes important technique of unsupervised classification that arranges pattern data (multidimensional space vectors) in the clusters (or regions). A number of data clustering methodologies have been proposed in number of papers and a dense literature is available for extracting information from an image and to partition into multiple parts .These algorithms have their advantages and disadvantages in terms of cluster processing , time complexity and accuracy[1] .When data is divided into a number of distinct clusters where each data element belong to exactly each cluster is called hard clustering whereas a process of assigning the membership levels and then utilize these levels to assign data to one or more clusters with a degree of some membership values is called soft clustering.

Patterns or vectors in the same cluster are analyzed as related to some predefined criteria, in comparison to distinct patterns belongs to different clusters . Application areas of image clustering comprises of different areas that include data mining, statistical data analysis, compression, vector quantization and pattern recognition . The procedure for Image analysis moves in a direction where data is grouped into meaningful regions also known as image

segmentation which is the presents the first step into clustering of images from basic level to more detailed routines and procedures in computer vision and image understanding[4][5].

This paper in Section 2 briefly reviews clustering techniques. K-means algorithms described in section 3. Fuzzy logic is outlined in Section 4. The fuzzy k-means algorithms briefly described in section 5 and section 6 describe performed experiments and obtained results. Section 7 concludes the paper.

2.CLUSTERING

Clustering is an unsupervised learning task where the primary condition is to identify a finite set of categories of clusters to describe the data .As compare to image classification that analyses labeled instances of classes , image clustering is done when there is no training stage, and is usually used when the classes are not known in advance. The data items are defined by a similarity metric , and then similar regions are grouped together to form clusters. Often, the attributes providing the best clustering should be identified as well. The grouping of data into clusters is based on the principle of maximizing the intra class similarity and minimizing the inter class similarity. To distinguish one cluster from another few Properties about the clusters can be analyzed to determine cluster profiles. A new data instance is classified by assignment to the closest matching cluster, and is assumed to have similar characteristics to the data of other region in the image[1][3][5].

High quality clusters with high intra-class similarity are used to cluster image and that method is termed as good clustering method as compare to low inter class similarity (Dissimilar to the objects in other

clusters). The quality of a clustering result depends on both the similarity measure used by the method and its implementation. The quality of a clustering method is also measured by its ability to discover some or all of the hidden patterns.
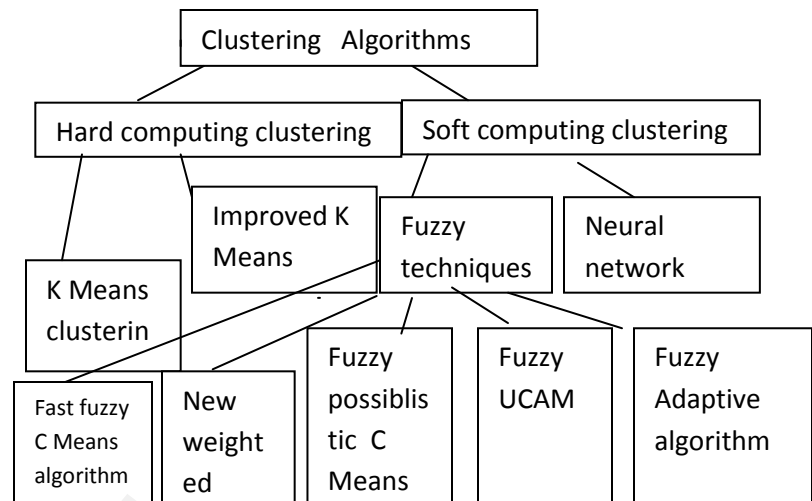
3. K-MEANS ALGORITHM

K-means is one of the simplest unsupervised hard learning algorithm and iterative technique that is used for clustering to partition the image into K clusters .It solve the well known clustering problem .Since it follows iterative algorithm to find minimized sum of distances from each each object to its cluster centroid, over all clusters .The procedure of k means algorithm is a simple and easy way to classify a given data set through priori fixed specified number of clusters [5][8]. The main idea is to find k number of clusters and centroids should be placed in cunning way because different distance clusters produce different results .These centroids should be chosen in a greedy way so that of different location causes different result. So, the better choice is to place them as much as possible far away from each other. Now in next step is each point belonging to a given data set is taken and associated with the nearest centroid. When no point is found further

pending, the first step is completed and an early grouping is done. At this point we again need to recalculate k new centroids of the clusters resulting from the previous step. After all this we have k new centroids, then a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice the end of clustering when the k centroids change their location step by step until no more changes are done. In other words centroids do not move anymore and the main aim of K means is to discover clusters of data within data is reached the stopping criteria[1][2].

The algorithm can be summarized as follows:

- Initialize the number of clusters.
- Randomly set the clusters centroids.
- For all data or image :
    a) Calculate the distance between clusters centroids and the data.
    b) Move the data closer to the cluster that has the less distance as compare to others .
- Again calculate the new centroids of the remaining clusters so that data can be moved.
- Compare new centroids with old centroids , if they are different then go to step 3, otherwise stop.

4. FUZZY-LOGIC

Fuzzy logic is an organized method for dealing with imprecise data .The input data used for clustering are termed as fuzzy set .Fuzzy Logic is a problem-solving control system methodology not a control methodology that provides a way of processing data by allowing partial set membership rather than crisp set . It uses membership values lends itself to implementation of systems that in turn used for systems that can be simple, small, multi-channel PC or workstation-based data acquisition and control systems, embedded microcontrollers , networked systems etc[1...8]. The fuzzy logic approach was not applicable to control systems till the era of 1970's due to insufficient computer capability that is why earlier systems were designed to accept precise and accurate data. This approach can be used for implementation of systems that comprises either of both hardware, software, or a combination of both. It can be defined also as a tool that provide a simple way even for vague, ambiguous, imprecise, noisy, or missing input information to arrive at a definite conclusion. Soft computing paradigm approach are also related to fuzzy logic for controlling problem solutions as it mimic the way how a person would make decisions, for faster computing.

5. FUZZY C MEAN ALGORITHM

The most widely used approach for fuzzy clustering is Fuzzy C means that assign membership levels and use them to assign data elements to one or more clusters .Traditional clustering approaches like K Means clustering algorithms generate partitions, in a partition, each pattern belongs to one and only one cluster(a set comprises of crisp and accurate values) but Fuzzy clustering extends this approach further to assign or associate each pattern present in a image to every cluster of data of image using a membership function .Fuzzy C Means is a "soft clustering" technique and provide a more precise computation of the cluster membership and has been used successfully for image clustering applications like medical, geological and satellite images[1][3][5].Sometimes it is not partition of regions ,the output is just

clustering .This approach is a widely applied method for obtaining fuzzy models for images that comprises of data. It has been applied successfully in numerous fields including geographical surveying, finance or marketing and pattern recognition. The main difference with K means algorithm is that instead of making a hard decision about clusters such that which cluster belong to which region of data, it assigns a value between 0 and 1 which describes every cluster "how many clusters belong to each pixel " but as per fuzzy logic rule it states that the sum of the membership value related to every pixel to all clusters must be 1 . When the membership function calculated is highest among all the more possibility is there that the pixel belongs to the particular cluster. It is based on basically minimization of the following objective function[3][4][5].

$$J = \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik}^{m} |p_i - v_k|^2$$

- J is the objective function.
- number of pixels in the image E is n .
- c is the number of clusters.
- μ is the fuzzy membership value from table.
- m is a fuzziness factor (a value > 1).
- pi is the ith pixel in E.
- vk is the kth cluster centroid.
- |pi − vk| is the Euclidean distance between pi and vk .

The algorithm can be summarized as:
- Initialize the fuzzy parameter (a constant >1) , the number of cluster, and the stopping condition.
- Set initial values in matrix after fuzzy partition .
- Set the initial loop counter variable k = 0.
- Calculate the cluster centroids and objective value J.

- For every pixel associated with cluster, compute the membership values and fill the matrix.
- If the value of objective function between consecutive iterations is less than the stopping condition, then stop otherwise set k=k+1 and go to step 4.
- Defuzzification of membership values .

6. EXPERIMENTAL RESULTS

So far after the introduction of the two clustering techniques and their basic mathematical foundations , now it's turn for the discussion of these techniques on the basis of a practical approach. First step is to collect data for image clustering , as if there should be some standard collection of data on which the clustering of image can be differentiated easily by various algorithm .When benchmarking an algorithm it is recommendable to use a standard test data set for researchers to be able to directly compare the results. "MIAS" is a database of static images of mammogram images[26]. Images were taken in uncontrolled indoor environment using five video surveillance cameras of various qualities .The experiment is done on database contains static medical images and some real time coloured environment images .

Both approaches work well for medical images and real time environmental images .After collection of data , the next step is to design a GUI interface by using C# language by the help of various classes and methods . K means algorithm is simply implemented by help of classes and methods but Fuzzy C Means includes special classes and methods for fuzzification and defuzzification. This project compares K-means and Fuzzy C-

Means clustering image segmentation algorithms. The algorithms are developed in C# programming language for analysis and comparison. K-means clustering produces fairly higher accuracy and requires less computation. C means clustering produces close results to K-means clustering, yet it requires more computation time than K-means because of the fuzzy measures calculations involved in the algorithm.

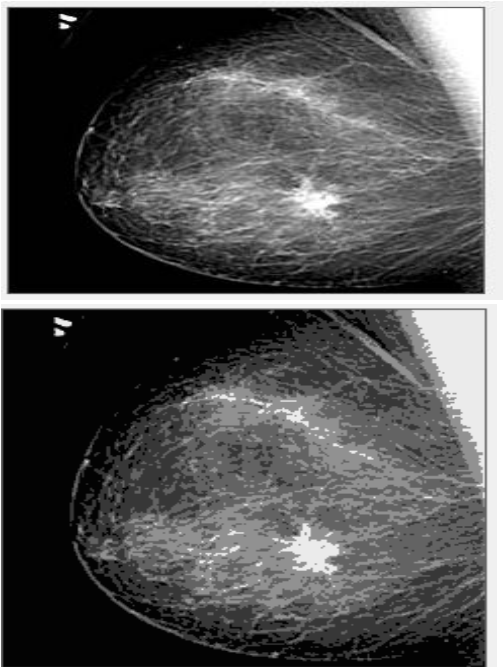|	No. of clusters	Time Taken	Accuracy
K Means	5	75 seconds	Less accurate
Fuzzy C Means	5	1 min 30 seconds	More accurate

Figure 3- K Means Clustered Image
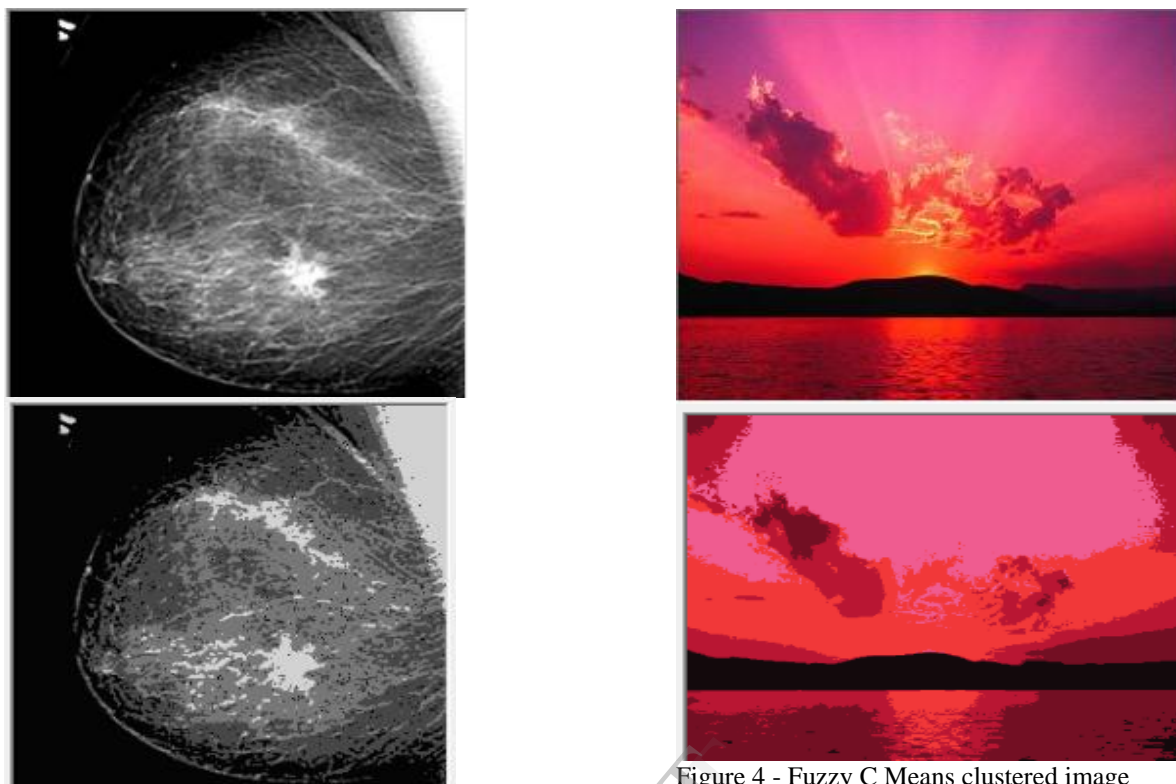
Figure 4 - Fuzzy C Means clustered image

	Initial Cluster	Concept	Centroids	Cluster Value	Cluster Error
K-Means	K	Initial cluster value	Initial Seeds	Depend on Seeds	Yes , If Wrong Seeds
Fuzzy C Means	C	Membership values of clusters	Initial Seeds	Depend on Seeds	Yes , If Wrong Seeds

6. CONCLUSION

The intension of this project is to do comparative analysis between K means algorithm and Fuzzy C means clustering algorithm of image clustering K-means clustering produces fairly higher accuracy and requires less computation. C means clustering produces close results to K-means clustering, yet it requires more computation time than K-means because of the fuzzy measures calculations involved in the algorithm.

Some algorithms has less advance in image which is high proportion of intersected segmentation object and less background. So the future scope is to develop clustering techniques that is better for noisy images , MRI images , Remote Sensing Images ,

mammogram images, background images and other applications.

REFERENCES

[1] Chih-Cheng Hung, Sameer Kulkarni, and Bor-Chen Kuo, "A New Weighted Fuzzy C-Means Clustering Algorithm for Remotely Sensed Image Classification" IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING January2010.

[2]. Nor Ashidi Mat Isa " Adaptive Fuzzy Moving K-means Clustering Algorithm for Image Segmentation" IEEE Trans. On Systems, Man. and Cybernetics, vol. 30 September 08, 2009.

[3] R. L. Cannon, J. V. Dave, and J. C. Bezdek, "Efficient implementation of the fuzzy c-means clustering algorithm" IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, pp. 248-255, 2009.

[4] P. Ganesan,V. Rajini, " A Method to Segment Color Images based on Modified Fuzzy-Possibilistic-C-Means Clustering Algorithm " IEEE Transactions on Volume 34 2010 IEEE.

[5] A. Banumathi A. Pethalakshm "Refinement of K-Means and Fuzzy C-Means " IEEE Trans.Med. ImageVolume 39– No.17, February 2012.

[6] N. A. Mat-Isa, M. Y. Mashor, and N. H. Othman, "Comparison of segmentation performance of clustering algorithms for Pap smear images". Proceedings of International Conference on Robotics, Vision, Information and Signal processing (ROVISP2003). pp. 118-125, 2003.

[7] Chun-yan and Yu1 Ying Li2 "A Novel Modified Kernel Fuzzy C-Means Clustering Algorithm on Image Segmentation" The 14th IEEE International Conference on Computational Science and Engineering August 2011 .

[8] Mrs. Bharati R.Jipkate "A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms" International Journal Of Computational Engineering Research / ISSN: 2250–3005 May-June 2012.

[9] Zhi-bing Wang Rui-hua Lu " A New Algorithm for Image Segmentation Based on Fast Fuzzy C-Means Clustering" June 2008 International Conference on Computer Science and Software Engineering.

[10] R. Zhang, G. Zhao, and L. Su, "A new edge detection method in image processing," IEEE International Symposium on Communications and Information Technology, vol. 1, pp. 445 – 448, 2005.

[11] P. Ganesan and V. Rajini, "A Method to Segment Color Images based on Modified Fuzzy-Possibilistic-C-Means Clustering Algorithm" ACM Computer Surveys 978-1-4244 2010 IEEE.

[12] Hongbao Cao1 and Yu-Ping Wang1,"Segmentation of M-FISH images for improved Classification of chromosome with an Adaptive Fuzzy C-Means clustering algorithm"Advanced Digital Imaging Research, LLC, League City,2011 IEEE.

[13] Jie Yu, Peihuang Guo, Pinxiang Chen, Zhongshan Zhang and Wenbin Ruan, "Remote sensing image classification based on improved fuzzy c-means", Geo-Spatial Information Science, vo1.11, no.2, pp:90-94, 2008.

[14] Stelios Krinidis and Vassilios Chatzis, " A Robust Fuzzy Local Information C-Means

Clustering Algorithm",IEEE Trans on image processing,vol 19,May 2010.

[15] F. Cui, L. Zou, and B. Song, "Edge feature extraction based on digital image processing techniques," IEEE International Conference on Automation and Logistics, pp. 2320-2324, 2008.

[16] Brinkmann, B.H., Manduca, A., Robb, R.A, "Optimized homo morphicun sharp masking for MR grayscale in homogeneity correction", IEEE Transactions Medical Imaging 17, 161–171, 1998.

[17] Metin Kaya, "Image Clustering and Compression Using an Annealed Fuzzy Hopfield Neural Network", International Journal of Signal Processing, 2005, pp.80-88.

[18] J.P. Fan, D.K.Y Yau, and A.K. Elmagarmid. Automatic image segmentation by integrating color edge extraction and seeded region growing.IEEE Transactions on Image Processing, 10(10):1454–1466,2001.

[19] Sudhavani and Sathyaprasad, "Segmentation of Lip Images by Modified Fuzzy C-means Clustering Algorithm",

IJCSNS International Journal of Computer Science and Network Security, Vol.9 No.4, pp.187-191, April 2009.

[20]http://public.fhwolfenbuettel.de/~hoepp nef/clustering.html

[21] www.reference.wolfram.com

[22] J. Leski, "Toward a robust fuzzy clustering," Fuzzy Sets Syst., vol.137,no. 2, pp. 215–233, 2003.

[23] K.Wu and M. Yang, "Alternative C-means clustering algorithms," Pattern Recognition., vol. 35, no. 10, pp. 2267–2278, 2002.

[24] F. Masulli and A. Schenone, "A fuzzy clustering based segmentationas support to diagnosis in medical imaging," Artif. Intell. Med.,vol. 16, no. 2, pp. 129–147, 1999.

[25] S.R.Kannan, S. Ramathilagam, R.Pandiyarajan and A.Sathya, " Fuzzy Clustering Approach in Segmentation of T1-T2 brain MRI," Int. Journal of Recent Trends in Engineering, Vol2, No.1, pp. 157-160, Nov.2009.

[26] http://peipa.essex.ac.uk/info/mias.html.