"A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms"

Mrs. Bharati R.Jipkate and Dr.Mrs.V.V.Gohokar
SSGMCE, Shegaon, Maharashtra-443101 (India)

Abstract— Segmentation of an image entails the division or separation of the image into regions of similar attribute. The most basic attribute for segmentation of an image is its
luminance amplitude for a monochrome image and color components for a color image. Clustering is one of the methods used for segmentation. The objective of this paper is to compare the performance of various segmentation techniques for color images. K-means clustering and Fuzzy C-Means clustering techniques are compared for their performance in segmentation of color images.

Keywords— K-Means clustering, Fuzzy C- Means Clustering.

I. Introduction

Segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful easier to analyze. Clustering is one of the methods used for segmentation. Clustering can be considered the most important *unsupervised learning* problem. it deals with finding a *structure* in a collection of unlabeled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way".A *cluster* is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters.

Clustering algorithms may be classified as listed below:

- Exclusive Clustering
- Overlapping Clustering

In the first case data are grouped in an exclusive way, so that if a certain datum belongs to a definite cluster then it could not be included in another cluster. On the contrary the second type, the overlapping clustering, uses fuzzy sets to cluster data, so that each point may belong to two or more clusters with different degrees of membership. K-means is an *exclusive clustering* algorithm, Fuzzy C-means is an *overlapping clustering* algorithm.

The specific criterion to be used for segmentation depends on the application. Pixels may belong together on the basis of the same color and/or the same texture and/or distance.

For segmentation Input image is transformed into feature space. The clustering task separates the data into number of partitions, which are volumes in the n-dimensional feature space. These partitions define a hard limit between the different groups and depend on the functions used to model the data distribution Clustering is the search for distinct groups in the feature space [7].
It is expected that these groups have different structures and that can be clearly differentiated.'

II K-means clustering

It is an iterative technique that is used to partition an image into *K* clusters.K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids,[7] one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, 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 that the k centroids change their location step by step until no more changes are done. In other words centroids do not move anymore. K-Means clustering generates a specific number of disjoint, flat (non-hierarchical) clusters. It is well suited to generating global clusters. The K-Means method is numerical, unsupervised, non-deterministic and iterative.

This algorithm aims at minimizing an *objective function*, e.g. a squared error function. The objective function

$$\sum_{j=1}^{k} \sum_{i=1}^{n} \|x_i^{(j)} - c_j\|^2$$

where$\|x_i^{(j)}-c_j\|$is a chosen distance measure between a $x_i^{(j)}$ data point and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centers[1-2].

III Fuzzy c-means clustering

In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In **fuzzy clustering** (also referred to as **soft clustering**), data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters.

One of the most widely used fuzzy clustering algorithms is the Fuzzy C-Means (FCM) Algorithm (Bezdek 1981). The FCM algorithm attempts to partition a finite collection of n elements $X = \{x_1,...,x_n\}$ into a collection of c fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of c cluster centres $C = \{c_1,...,c_c\}$ and a partition matrix

$$J=U = u_{i,j} \in [0,1],\ i = 1,\ldots,n,\ j = 1,\ldots,c$$

where each element u_{ij} tells the degree to which element x_i belongs to cluster c_j . Like the k-means algorithm, the FCM aims to minimize an objective function. The standard function is:

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(\mathrm{center}_k,x)}{d(\mathrm{center}_j,x)}\right)^{2/(m-1)}}.$$

which differs from the k-means objective function by the addition of the membership values u_{ij} and the fuzzifier m. The fuzzifier m determines the level of cluster fuzziness. A large m results in smaller memberships u_{ij} and hence, fuzzier clusters. In the limit m = 1, the memberships u_{ij} converge to 0 or 1, which implies a crisp partitioning. In the absence of experimentation or domain knowledge, m is commonly set to 2.

In **fuzzy clustering**, each 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.

Any point *x* has a set of coefficients giving the degree of being in the *k*th cluster $w_k(x)$. With fuzzy *c*-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

$$C_k = \sum_x w_k(x)x / \sum_x w_k(x)$$

The degree of belonging, $w_k(x)$, 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[8].

IV The algorithm of k-means clustering

Step1.Choose the number K of clusters either manually, randomly or based on some heuristic.

Step2.Generate K clusters and determine the cluster's center.

Step3**.** Assign each pixel in the image to the cluster that minimizes the variance between the pixel and the cluster center

Step4**.**Re-compute cluster centers by averaging all of the pixels in the cluster.
Step5**.**Repeat steps 3 and 4 until some convergence criterion is met

Figure1 shows segmentation in 2 and 3 and 5 clusters using this method

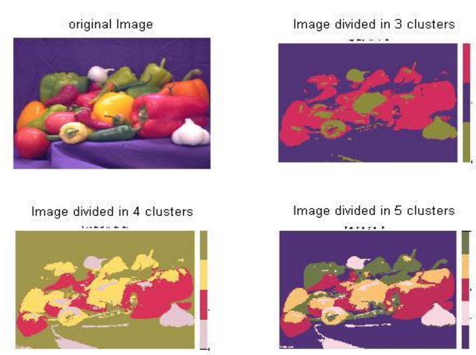

Fig:1 Kmeans Segmentations

V The algorithm of fuzzy c-means clustering

Step1.Choose a number of clusters in a given image.

Step2.Assign randomly to each point coefficients for being in a cluster.

Step3**.** Repeat until convergence criterion is met.

Step4. Compute the center of each cluster.

Step5**.**For each point, compute its coefficients of being in the cluster[4-5].

Fig2 shows the image segmented by c means algorithm

Fig :2. C means

3.Result and Conclusion

The paper compares k-means and fuzzy c-means clustering image segmentation algorithms .The algorithms are developed in MATLAB for analysis and comparison. K-means clustering produces fairly higher accuracy and requires less computation. C means clusyering 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.

References

[1] ^ a b MacQueen, J. B. (1967). "Some Methods for classification and Analysis of Multivariate Observations". **1**. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp. 281–297. MR0214227. Zbl 0214.46201. http://projecteuclid.org/euclid.bsmsp/1200512992. Retrieved 2009-04-07..

[2] Dr.Teknomo, Kardi. K-Means Clustering Tutorials. http:\\people.revoledu.com\kardi\ tutorial\kMean\July 2007.

[3] S. Adaekalavan,C.Chandrasekar'A Comparative Analysis of Fuzzy C-Means Clustering and Harmonic K Means Clustering Algorithms'European Journal of Scientific Research ISSN 1450-216X Vol.65 No.1 (2011), pp. 45-49© EuroJournals Publishing, Inc. 201 http://www.europeanjournalofscientificresearch.com .

[4] J. C. Dunn (1973): "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *Journal of Cybernetics* 3: 32-57

[5] J. C. Bezdek (1981): "Pattern Recognition with Fuzzy Objective Function Algoritms", Plenum Press, New York

[6] Tariq Rashid: "Clustering" http://www.cs.bris.ac.uk/home/tr1690/documentation/fuzzy_clustering_initial_report/node11.html

[7] A Tutorial on Clustering Algorithmshome.dei.polimi.it/matteucc/Clustering/tutorial_html/

[8] **Fuzzy** clustering - Wikipedia, the free encyclopedia en.wikipedia.org/wiki/Fuzzy_clustering -