Analysis of Image Clustering Techniques

Agrawal Avni Rajeev, Dr. Saroj Hiranwal, Vijay Kumar Sharma

Abstract—In field of image processing, image segmentation plays a vital role. Segmentation can be useful in various applications namely, medical diagnosis, forensics, object recognition. At present several segmentation techniques are available such as edge detection, watershed technique, region growing techniques, and clustering. The clustering technique is emerging as the most prominent technique. Clustering is grouping technique that groups the data in clusters such that image can provide the significant data. Thus this paper evaluates some important clustering techniques according to latest researches. We here study fuzzy c-means (FCM), conventional k means (KM), and moving k means (MKM), adaptive fuzzy k-means (AFKM) techniques. The paper thoroughly studies the procedure of those clustering techniques. This paper discusses the concept of soft and hard clustering i.e. FCM technique supports the concept of soft clustering whereas k means supports the concept of hard. Moreover the concept of fitness and belongingness are also described. Adaptive fuzzy K means (AFKM) combines the concepts of FCM and conventional KM techniques in order to provide better segmentation. We here analyze the improved method adaptive fuzzy k means (AFKM) by comparing it with other clustering techniques. The review consists of some valuable facts as consequence of vital analysis. The analysis is carried out in terms of the execution speed, sensitivity to noise and their visual quality. We discuss the pros and cons of these clustering techniques. However those clustering technique can be appropriate for specific application.

Index Terms—Image segmentation, K means (KM), Fuzzy C means (FCM), Adaptive fuzzy Moving K means (AFMKM), Adaptive fuzzy K means (AFKM).

I. INTRODUCTION

Image segmentation is the technique to classify digital images into multiple significant regions or sets of pixels. Typically, image segmentation is characterized as partitioning an image into distinct segments that are identical with respect to some properties [1]. As image segmentation can detect the regions of concern in a scene or annotate the data. Image segmentation plays a vital role in automated object recognition systems, medical science and many more.

Segmentation algorithms can be classified into different categories based on the features such as threshold, template matching, and region based technique and clustering. Those techniques have their own restrictions and benefits in terms of suitability, performance and computational cost. For instance, the thresholding technique produces a good quality and rapidity segmentation, but it remains a fact that it is sensitive to noise if threshold value chosen is the appropriate Clustering is an unsupervised classification designed to group a set of data samples in a manner where data with similar characteristics are grouped together into larger units of analysis known as clusters while those that are dissimilar are separated in their distinct clusters. In image segmentation, clustering algorithm iteratively computes the characteristics of each cluster and segments the image by classifying each pixel in the closest cluster according to a distance metric. Through clustering technique, a much better results of segmentation can be obtained but over-segmentation is one of the problems that must be faced.

In digital image segmentation applications, clustering technique is commonly used to segment regions of interest and/or detect borders of objects in an image. There are many applications of clustering for image segmentation or edge detection.

II. LITERATURE REVIEW

Clustering is universal in science and engineering with numerous application domains ranging from bioinformatics and medicine to the social sciences and the web. Perhaps the most well-known clustering algorithm is the so-called “k-means” algorithm [3]. Image clustering and categorization is a means for high-level description of image content. The goal is to find a mapping of the archive images into classes (clusters) such that the set of classes provide essentially the same information about the image archive as the entire image-set collection. The generated classes provide a concise summarization and visualization of the image content that can be used for different tasks.

A. K-mean algorithm

K-mean is the digital image clustering technique, which is used to calculate mean in the sub-cluster and based on distance computation between data points and centre of individual sub-cluster which is found in the previous iteration, centers of each cluster gets updated. In the presence of noise, k-mean algorithm gives unsatisfactory result in mean calculation. K means tends to minimize eq. (1)

\[ J = \sum_{i=1}^{k} \sum_{j=1}^{n} \| x_{ij} - C_i \|^2 \]  

(1)

B. Moving K means

The K means segmentation earlier proposed by Hartigan and Wong was improved by Mashor. [4,5]. The cluster center can be computed at each iteration by eq. (2). The idea of
fitness was presented to guarantee number of data points in the cluster. The fitness for individual sub cluster can be computed using eq. (3)

\[ C_j = \frac{1}{n_j} \sum_{i \in c_j} v_i \]

\[ f(C_j) = \sum_{i \in c_j} \left( \| v_i - c_j \| \right)^2 \]

According to K.K and N.A. Mateach cluster center must follow the condition depicted in eq. (4)

\[ f(C_s) \geq \alpha F(C_i) \]

The MKM segmentation lessens dead centers. Moreover center redundancy problems are reduced in turn the problem of trapped center at local minima also lessens. However, MKM is sensitive to noise and cluster centers are not located precisely and efficiently. The concept of fitness tends to a misplaced largest fitness members to smallest fitness in turn creating noisy data and imprecise outcomes.

C. Fuzzy C Means Segmentation

Fuzzy C mean technique can be used to resolve the problems arises in the k-mean clustering. Previously hard clustering was used in which each sample is part of any single cluster out of n number of clusters. When illumination and noise kind of problems come into picture, hard segmentation seems to be difficult. Fuzzy set theory consists of partial membership theory using which fuzzy clustering also known as segmentation technique was being used. [6]

Consider c number of clusters is to be designed of k number of objective samples whose fuzzy level is l. Each sample corresponds to every cluster. Computation of center of the cluster is shown in eq. (5).

\[ C_j = \frac{\sum_{i=1}^{l} (u_{ij})^l x_i}{\sum_{l=1}^{l} (u_{ij})^l} \]

where, i = 1 to total number of clusters;

Distance is to be computed using center value of the cluster in each iteration by subtracting from each sample points k(i) using formula shown in eq. (6)

\[ d_{ij} = \left\| k_i^{(j)} - c_j \right\| \]

After calculating distance, Fuzzy membership matrix is need to be update using equation formulated shown in eq. (7)

\[ u_{ij} = \left[ \sum_{j=1}^{l} \frac{(u_{ij})^l}{d_{ij}} \right]^{-1} \]

Fuzzy C means segmentation does not needs prior information and supports membership concept. Although it is subtle to noise and undergoes computation time overhead.

D. Adaptive Fuzzy K Means (AFKM)

Adaptive fuzzy K means incorporates the features of fuzzy c means, K means clustering segmentation. This clustering segmentation introduces the idea of fuzziness and belongingness [7]. Conventional K means is used to form basic cluster using Euclidean distance and mean of data points while the concept of each data point can be part of more than one cluster based on their varying degree of membership also known as fuzziness. The AFKM segmentation proposes the idea of belongingness. The degree of belongingness specifies the strength of relationship between data points and center. The belongingness can be computed as eq.(8)

\[ B_j = \frac{c_j}{\mu_{ij}} \]

E. Adaptive Fuzzy Moving K Means (AFMKM)

Adaptive fuzzy moving K means was proposed as a consequence of adapting all the features of moving K means (MKM), fuzzy C means segmentation (FCM)Techniques. AFMKM use the membership concept of fuzzy c means to assign data point to nearest cluster. Whereas the fitness concept of moving K means is used to assign data point to in appropriate cluster according to their fitness value. AFMKM clustering algorithm allocates the centers of the clusters proficiently. It overcomes the dead centers trapped center and center redundancy problems. Furthermore, algorithm is also less subtle to noise and initialization process. Moreover AFMKM is fast robust and accurate.

III. RESULTS AND EVALUATION

The research on evaluation of image segmentation can provide crucial reference for those segmentation algorithms, and so this research deserves wide attentions. Understandably, the basic requirements are as follows: universal use for evaluation algorithm, its simplification and reliability, and whether referent images or manual intervention is needed.

Qualitative analysis of K means (KM), moving k means (MKM) fuzzy c means (FCM), adaptive fuzzy k means (AFKM) and adaptive fuzzy moving k means (AFMKM) yields similar results but the results of AFMKM are comparatively better.
Fig 1: Results of clustering segmentation (a) original (b) K means (c) Moving K means (d) FCM (e) AFKM (f) AFMKM

IV. CONCLUSIONS

This paper presents the analysis of the clustering techniques for segmentation. In this paper we have discussed five clustering techniques namely K means (KM), moving k means (MKM) fuzzy c means (FCM), adaptive fuzzy k means (AFKM) and adaptive fuzzy moving k means (AFMKM). The basic problems here we deal with are dead centers trapped center and center redundancy problems. Consequently, the results so obtained say the image obtained using adaptive fuzzy moving k means is having better quality. Also it overcomes the problem of other clustering techniques.

REFERENCES