Application of clustering Technique for Tissue image segmentation and comparison

Mohit Agarwal, Gaurav Dubey, Ajay Rana

Abstract— through this paper we proposed the methodology that incorporates the K-means and fuzzy c means algorithm for the color image segmentation. The image segmentation may be defined as the process of dividing the given image into different parts. Here we are taking color image as the input and we are supposed to segment the given image on the basis of its color. The clustering algorithm also proved to be very efficient in the process of image segmentation. The clustering can be defined as the method of grouping the similar kind of objects from the input. The grouping can be on the basis of the attributes like color, shape, texture, size and other. The motive behind the segmentation is to extract the some meaning information from the input like image so that it can be utilized in effective manner. The K-means and fuzzy c means are the most popular clustering algorithm. The k-means use the iterative approach to partition the given image into k clusters while in the fuzzy c means clustering method of image segmentation a group of images are allowed to form a cluster on the basis of similarity checks.

Index Terms— Segmentation, K-means, Fuzzy c means, clustering, unsupervised clustering.

1 INTRODUCTION

IN today's world lots of method for the information conveying is used by the users and images are also considered as one of the important method of information passing. So it is important to understand the images and extract the information efficiently so that it can be used for other useful purposes. The problem associated with broad view of vision is that interference. Segmentation of image proving very helpful in extraction of the information from given image [1]. The image is first supposed to be divided into different parts so that one can easily concentrate on the desired portion or the object in them. For this purpose segmentation using clustering is proving as the better technique. Image segmentation is the process of splitting the given image into the similar regions on the basis of the available attributes like color, texture, shape, size and etc. The similar or homogeneous region of the image is the result of segmentation which used to put the pixels of similar kind attribute in a particular region. Segmentation plays a very important role in extracting the information from the image to form a similar region by classifying pixels on some basis to group them into a region of similarity. The result obtained from segmentation is a set of regions that collectively constitute the entire image. One natural view of segmentation is that we are attempting to determine which components of a data set naturally "belong together". This methodology is termed as "clustering". The image segmentation has wide area of application like object recognition, face recognition, medical imaging and video retrieval [2] [3].

- *Mohit Agarwal is currently pursuing masters degree program in computer science & engineering in Amity University, India, PH-09258707066. E-mail: mohit.hcst@gmail.com*
- *Gaurav Dubey is currentlyAssistant Professor in Amity University, India, PH-09958759459. E-mail: g.dubey@amity.edu*
- *Dr. Ajay Rans is currently Director/Professor in AmityUniversity, India, PH- 098 91374031 E-mail: ajay_rana@amity.edu*

2 IMAGE SEGMENTATION

Image Segmentation is defined as the process of splitting the given image in group of pixels into homogeneous regions on the basis of common features [4]. Features can be represented by matrix of color, texture and many more each of which help in finding the similarity of the pixels in that region [5]. Segmentation of an image is the process of dividing the given image into multiple regions. The aim of the segmentation is to convert the image into simple format and change the representation into something more constructive to analyze. Segmentation is basically used for determining the objects and their boundaries (curves, lines, etc) in the given image [6]. The image segmentation results in set of regions that collectively represent the entire image. Each region represents the pixels which are similar with respect to some features or characteristics like color, shape, size etc. So image segmentation may also be defined as the process of understanding and analyzing the image [7]. The Segmentation may prove helpful in recognition of the object [8], image compression, editing many more. The Segmentation process basically depends upon the quality of image which is to be processed [9]. This proves effective in the case of simple images due to very less variations in the pixels and required some preprocessing in case of complex image. The image segmentation process can be categorized into five categories such as Region Based Segmentation, Pixel based segmentation, Edge based segmentation, Edge and region Hybrid segmentation, and clustering based segmentation.

3 CLUSTERING

Clustering means the process of grouping the objects which is based on their attributes, so that the objects with attributes should lie in the same cluster . Clusters may be defined as the group of objects which are similar in nature on the basis of their attributes and dissimilar to the objects of other groups. Clustering is data mining technique which is used for data

analysis, image processing, pattern recognition and more. An image can be think as a group which is grouped on the basis of keyword (metadata) and its content i.e. description. Clustering used to divide the given images into various regions or classes without any prior information.

In metadata based clustering a keyword is a form of font which is used to refer the different features of an image. The similar features of an image are grouped into same clusters by assigning them a value. On the other hand in content based clustering, the content refers to the shape, color, texture or any other information that can be extracted from the image itself.

Clustering method includes partioning clustering which involves the creation of clusters that partitions the data into similar groups. In hierarchical clustering a hierarchy of clusters from each and every element is maintained.

A number of clustering techniques are available to make the image segmentation more efficient and effective. The clustering methods which we are using in this paper are relevance feedback, log based clustering, hierarchical clustering, graph based clustering, Retrieval-dictionary based and filter based clustering and many more.

3.1 K-Means Clustering

K-means clustering algorithm is one of the simplest clustering algorithms which lie in the category of unsupervised clustering algorithm. The main concept of this algorithm is that it used to classify the given dataset into various clusters. The number of clusters in this technique is fixed. The purpose of using the k-means algorithm is to get the primary segmentation of the input image. The main advantage of using the k-means algorithm is that it is simple and has relatively low computational complexity [9]. The second important advantage of using this particular algorithm is, it is suitable for biomedical image as the number of cluster (k) is known in advance for the image. For the tissue image which we used we are taking K=3.

The K-means clustering algorithm can be divided into two basic versions which are named as non- adaptive version which was introduced by Lloyd and an adaptive version which was introduced by McQueen.

Here in this paper we are going to concentrate on adaptive k-means which is based on Euclidean distance.

The first step involves the selection of the cluster centers. The data set is then partitioned into K clusters and the pixels are assigned to some clusters. The Euclidean distance is calculated from the each pixel to the mean of each cluster. The clusters should have no overlapping with each other. The dimension for the centroid is the same as the dimension of the data vectors. Every pixel which is assigned to some clusters is on the basis of the closeness which is obtained by using the Euclidian distance measure . After the clustering of all pixels, the mean of each cluster is calculated and this process will be continued until no significant changes will encountered.

The K-means clustering algorithm is as follows:

1. Place K points which are supposed to be clustered into the space represented by the objects.
2. Each object is assigned to the group that has the closest centroid.
3. When each object is assigned, recalculate the position of each k-centroid.
4. Repeat the steps 3.1.2 and 3.1.3 until the centroids no longer change. This will produce a division of the objects into the groups from which the metric which is to be minimized can be computed

3.2 Fuzzy C Means Clustering

Fuzzy c-means clustering algorithm is one of the most prominent clustering algorithms for color image segmentation. Ruspini in 1969 has come up with the idea of using fuzzy set theory for clustering. But the first fuzzy c-means clustering algorithm was introduced by Dunn [10] and later it was generalized and framed by Bezdek[11]. Unlike k-means clustering technique which use the concept of hard partitioning the fuzzy c-mean employs the concept of the fuzzy partitioning according to which a pixel can belong to all the groups with different membership grades 0 and 1.

The fuzzy clustering algorithm receives the data as a matrix of order n*m where n is the number of the data and m is the number of the parameters. The fuzzy c-means is a iterative algorithm which is used to separate the given set S in K compact clusters. Every cluster is denoted by its Centre ci, fuzzy c means also associates with every feature vector xj which belongs to fuzzy membership grade vij belongs to [0,1], 1<i<K in each of the K classes. The membership function vij represents the degree to which xi belongs to class i. The set V = {vij} {1≤i≤K, 1≤j≤N} is called the fuzzy partition of S. Let A = {ai} 1≤i≤K is the set of clusters centers. The fuzzy c-means clustering algorithm employs the iterative optimization procedure to calculate the minima of constrained objective function:

$$F(A, V, S) = \sum_{i=1}^{k}\sum_{i=1}^{N} v_{ij}^{m}\, d(xj, ai) \qquad (1)$$

In the above equation m can be defined as the parameter that controls the degree of fuzzification if m=1 it shows no fuzziness. The proposed algorithm creates the relative memberships, which shows the degree of sharing the pixels between the clusters.

The fuzzy c-means algorithm is very much similar to the k-means algorithm, the fuzzy algorithm here which we used for color image segmentation uses the extended feature of image i.e. the L*U*V transformation. The technique for the segmentation of color image segmentation is different from the segmentation of the grey image. The L*U*V color space is also known as CIELAB or CIE L*a*b which is used to quantify the visual differences.

The fuzzy c-means algorithm is as follows:

1. First we are supposed to choose the number of clusters.
2. Randomly assign the coefficients for each point to be in the cluster.
3. Repeat the following step until the algorithm has converged:
   3.1 Computation of the centroid for each cluster.
   3.2 The coefficient for each pixel is computed to ensure them to be in cluster

4 RESULT AND SIMULATION

The simulation is done is done on MATLAB on color tissue image with K=3. Figure 1 shows the simulation of the k-means clustering algorithm which is applied on the color tissue image. Figure 1(b) presents the initial cluster index image by k-means and figure 1(c), (d), (e) presents the objects in cluster1, cluster2 and cluster3 respectively

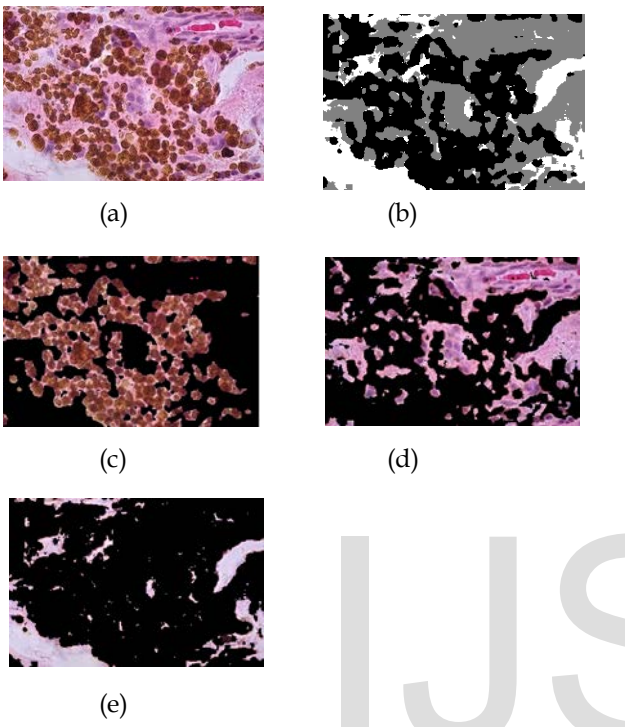

Figure 1 Result of K means clustering on color tissue (a)

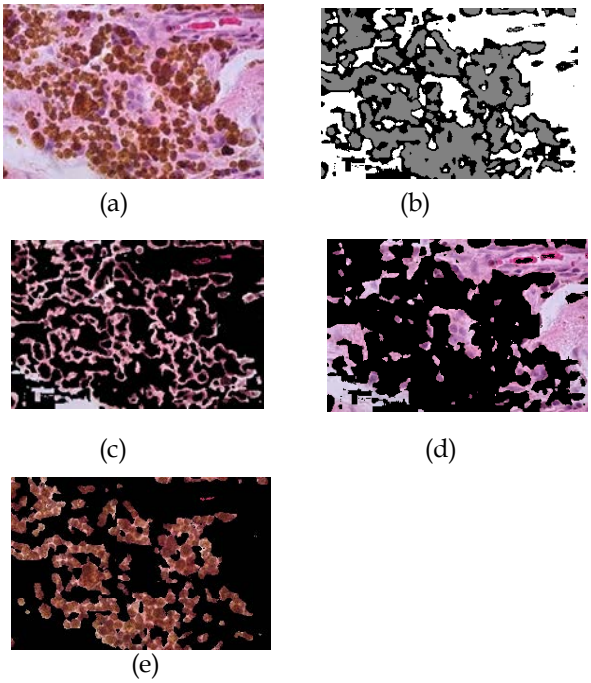

Figure 2 Result of FCM on color tissue (a)

while the figure 2 (c), (d), (e) presents the objects in the cluster1, cluster2 and cluster3 by using Fuzzy c-means clustering algorithm.

5 CONCLUSION

In this paper we segment the image and extract the features and clustered the given image on the basis of these features by using the k-means and fuzzy c-means clustering technique. The cluster analysis is done on the basis of the number of clusters, speed and other factors.We can easily make out that the clusters obtained from the FCM are better in comparison to the K-means. The FCM also proves faster in comparison to its counterpart if the number of clusters is same. The color image segmentation of tissue can be helpful in the diagnosis and medical purpose.

REFERENCES

[1] N.R. Pal and S.K. Pal, A review on image segmentation techniques, Pattern Recognition, vol. 26, pp. 1277-1294, 1993.

[2] Parag Puranik, Dr. P. R. Bajaj, Prof. P. M. Palsodkar "Fuzzy based Color Image Segmentation using Comprehensive Learning Practical Swarm Optimization (CLPSO)- A Design Approach ", Proc. International Multi-Conference of Engineers and Computer Scientists Vol I, IMECS 2009, March 18-20, Hong Kong, 2009.

[3] M.Ameer Ali, Laurence S Dooley and Gour C Karmakar "Object Based Image Segmentation Using Fuzzy Clustering"IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'06), 14-19 May 2006, Toulouse.

[4] Puzicha, J., Hofmann, T. and Buhmann, J. M.,"Histogram Clustering for Unsupervised Image Segmentation"., Computer Vision and Pattern Recognition , Vol.2.IEEEpress,602-608, 2000.

[5] Li Wenchao Zhou Yong Xia Shixiong China Univ. of Min. & Technol., Xuzhou, "A Novel Clustering Algorithm Based on Hierarchical and K-means Clustering" Print ISBN: 978-7-81124-055-9, On page(s): 605, 2009

[6] G. Salton and M.J. McGill, Introduction to modern information retrieval, New York, McGraw-Hill, 1988.

[7] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", 2nd ed., Beijing: Publishing House of Electronics Industry,2007

[8] Shi. J, and Malik. J, "Normalized cuts and image segmentation", IEEE Conf. Computer Vision and Pattern Recognition, IEEE Computer Society, vol. 3(2), pp. 731-737, 1997.

[9] Han, M.Kamber, "Data Mining concepts and Techniques",Morgan Kaufmann Publishers, 2002

[10] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," 1973.

[11] J. C. Bezdek, Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers, 1981.