International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 3, Issue 1, January 2014

The Survey on Various Clustering Technique for Image Segmentation

Dr. K. SATHIYASEKAR
Professor, Department of EEE
S.A Engineering College, Thiruverkadu,
Chennai, India

S.KARTHICK
Assistant Professor, Department of EEE
The Kavery Engineering College,
Salem, India

P.RAJA KANNIYA
Post Graduate Student, M.E (Embedded systems)
The Kavery Engineering College,
Salem, India

Abstract— **Image segmentation places an important role in image processing. This segmentation process can be done using various techniques like clustering, thresholding, edge detection and region extraction. This paper gives introduction to image processing operations and clustering process. Then the overview and algorithmic process of each clustering technique such as K-Means clustering, Kernel K-Means clustering, Fuzzy C-Mean and Graph based clustering is discussed.**

Index Terms— **Fuzzy C-Mean, Graph based clustering, Image segmentation, Kernel K-Means clustering, K-Means clustering.**

I. Introduction

Image Processing can do on different type of images like real time image, satellite image, and also in medical images. The Image Processing involves various steps namely; Image preprocessing, Restoration, Analysis and Compression. Preprocessing includes geometric correlation and radiometric correlation. The correlated image is then fed for restoration task. In restoration process, effective noise removal is done. Followed by this, image analysis is involved. Here it includes the feature extraction, segmentation and classification steps for the further analysis. Then the image is compressed to reduce the memory consumption which is observed as output image.

With the above mentioned steps, Image segmentation is one of the major key concepts in Image processing techniques. The goal of Image segmentation is to simplify or change the illustration of an image into something that is more meaningful and easier to examine. Image segmentation is normally used to locate objects and boundaries in images. In particular, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the similar label share certain visual characteristics. Set of contours extracted from the image or Set of segments that collectively cover the entire image is obtain as the result of image segmentation.

Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, texture or intensity. Adjacent regions are considerably different with respect to the same characteristic(s). This Image segmentation can done using various techniques. Thresholding, edge detection, region extraction and clustering are four main image segmentation techniques.

In this paper, clustering methods for image segmentation will be considered. The rest of this paper is organized as follows. In next section, concept of clustering is discussed. Section III will describe the K-Mean clustering algorithm. Section IV will describe the Kernel K-Mean clustering algorithm. Section V will describe Fuzzy C-Mean clustering algorithm. Section VI will describe Graph based clustering algorithm. Conclusion will be drawn in section VII.

II. Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of investigative data mining, and a common method for statistical data analysis used in various fields, including machine learning, pattern recognition, image analysis, information retrieval with bioinformatics. Cluster analysis itself is not one specific algorithm, but the common task to be solved. It can be achieved by different algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them.

Popular concept of clusters includes dense areas of the data space, groups along with small distances among the cluster members, intervals or particular distributions. So Clustering can be formulated as a multi-objective optimization problem. The apt clustering algorithm and parameter settings (including values such as a density threshold or the number of expected clusters, the distance function to use) depend on the individual data set and intended use of the results.

ISSN: 2278 – 909X

Cluster analysis is not an automatic task, but it an iterative process of knowledge discovery or interactive multi-objective optimization which involves trial and failure. It will frequently be needed to modify data preprocessing and model parameters until the result achieves the desired properties. The design of a "cluster" cannot be exactly defined which is one of the reasons why there are so many clustering algorithms. But they have common denominator i.e. a collection of data objects. However, different researchers make use of different cluster models, and for all of these cluster models again different algorithms can be used. Some of the clustering algorithms are: K-means clustering, Kernel based clustering, Fuzzy c-means clustering, Graph-based clustering.

III. K-MEANS CLUSTERING OVERVIEW

The K-Means (KM) is one of the premature center-based clustering methods. K-Means clustering generates a specific number of disjoint, flat (non-hierarchical) clusters. This is well suitable to generate globular clusters. The K-Means technique is numerical, unverified, non-deterministic and iterative. It adaptively assigns each pixel to each cluster in a closest-distance manner. The means of these clusters are then set as the new centers. The procedure is repeated until the sum of a dissimilarity measure, such as the Euclidean distance between centers and pixels, converges to a local minimum. The probability of each data point belonging to different clusters is usually referred to as the membership function. In the case of k-means, the membership function can only take values such as 0 or 1 and which shows data point can only belong to one cluster.

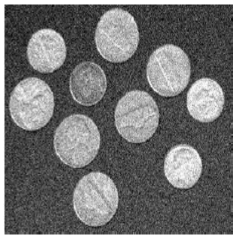
In this case, a hard membership function is implied and such detection schemes are often called “winner-takes-all” approaches. K-Means may be computationally faster than hierarchical clustering, with a large number of variables, (if K is small). K-Means may produce tighter clusters than hierarchical clustering, in particular if the clusters are globular. The KM algorithm has been widely applied to clustering problems mainly due to its simplicity and efficiency. But the k-means algorithm have some demerits such as difficulty in comparing quality of the clusters produced (e.g. for different initial partitions or values of K affect outcome), Fixed number of clusters can make it difficult to predict what K should be and it does not work well with non-globular clusters. Different final clusters can result based on different initial partitions

K-Means Algorithm Properties

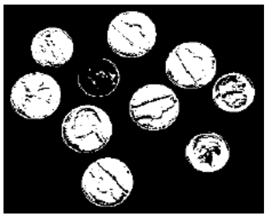
- KM always has K clusters.
- Each cluster will have at least one item.
- Due to non-hierarchical cluster they do not overlap.
- Every member of a cluster is closer to its cluster than any other cluster because closeness does not always involve the 'center' of clusters.

The K-Means Algorithm Process

- K clusters are formed by partitioning the dataset and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points.
- For all data point the distance from the data point to each cluster is calculated.
- Leave the data point where it is only if it closes to its own cluster. If the data point is not close to its own cluster, shift it into the closest cluster.
- Repeat the above step until a complete pass through all the data points’ results in no data point moving from one cluster to another cluster. On this point the clusters are stable and the clustering process ends.
- The choice of initial partition can greatly affect the final clusters that result, in terms of inter-cluster and intra-cluster distances and cohesion.

(a) Original Image (b) SPN Image



(c) KM Clustering Image

Fig.1. (a) original image, (b) “Salt & Pepper” noise (SPN) image, (c) K-Means (KM) Clustering image

IV. KERNEL K-MEANS CLUSTERING OVERVIEW

Kernel K-Means Clustering generates a specific number of hierarchical clusters. This algorithm applies the same trick as k-means but with one different that here in the calculation of distance; kernel method is used instead of the Euclidean distance. Kernel functions can be viewed as a non-linear transformation that increases separation of the input data and the integration of kernel function enables the K-Means algorithm to explore the inherent data pattern in the new space. And a major drawback to k-means is that it cannot separate clusters that are non-linearly separable in input space. This can be overcome by kernel k-means, where, before clustering, points are mapped to a higher-dimensional feature space using a nonlinear function, after that kernel k-means partitions the points by linear separators in the new space. This algorithm is able to identify the non-linear structures and is best suited for real life data set. But the drawback is number of cluster centers need to be predefined and it is complex in nature and time complexity is large.

ISSN: 2278 – 909X

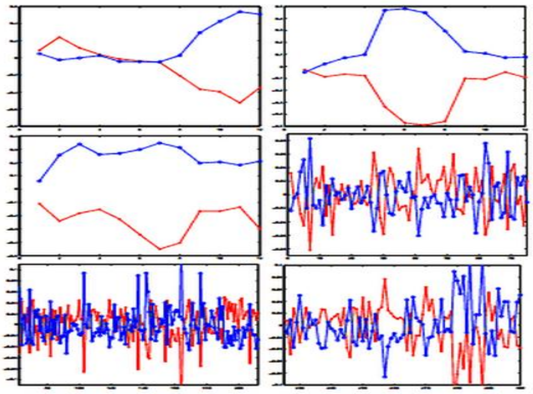


Fig.2. In each figure is plotted the mean expression profiles of two opposed clusters obtained on the human fibroblast dataset (first 3 plots) and the Rosetta dataset (last 3 plots). The clustering algorithm used in degree 2 polynomial kernel k-means.

Kernel k-means clustering algorithm Process

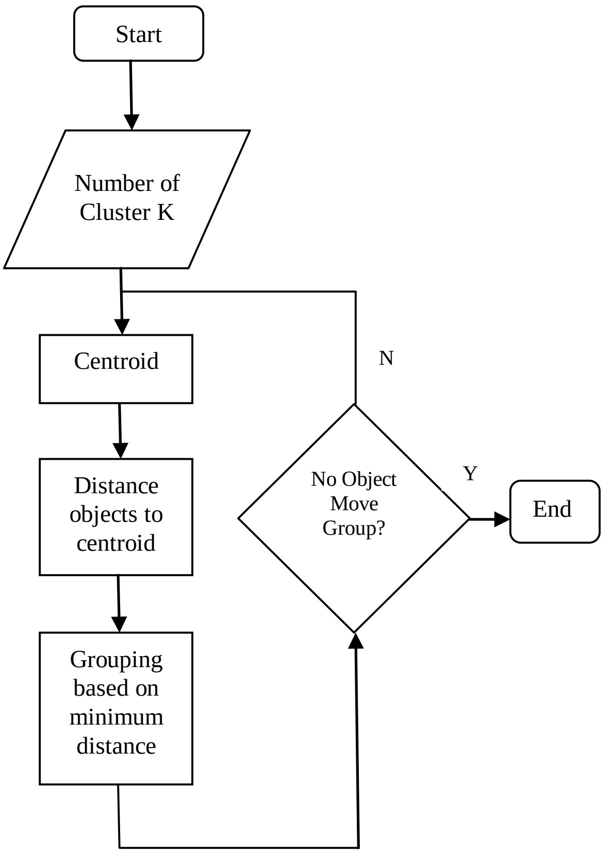


Fig.3. Flow chart of Kerel K-means clustering algorithm process

- Begin with a decision on the value of the number of clusters (k).
- Put any initial partition that classifies the data into k clusters. And then assign the training samples randomly or systematically as following: First take the initial k training sample as single-element clusters. Secondly assign each of the remaining (N-k) training samples to the cluster with the adjacent centroid. After each assignment, recomputed the centroid of the gaining cluster.
- Take each sample in sequence and compute its distance from the centroid of each of the clusters. If a sample is not currently in the cluster with the nearby centroid, switch this sample to that cluster and update the centroid of the cluster gaining the new sample and the cluster losing the sample.
- Repeat the previous step until convergence is achieved, that is until a pass through the training sample causes no new assignment.

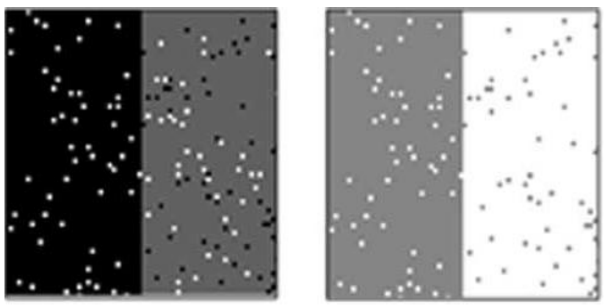
V. Fuzzy C-MEAMS CLUSTERING OVERVIEW

There are two main clustering strategies: the hard clustering scheme and the fuzzy clustering scheme. The usual hard clustering methods classify each point of the data set just to one cluster. As a consequence, the results are frequently very crisp, i.e., in image clustering each pixel of the image belongs just to one cluster.

The efficiency of hard (crisp) clustering technique is reduced in many real situations due to some issues like partial spatial resolution, intensity of overlapping, poor contrast, noise and intensity in homogeneities. The idea of partial membership value was introduced by Fuzzy set theory and which is described in a membership function. Fuzzy clustering scheme, as a soft segmentation method, has been generally studied and successfully applied in image clustering and segmentation. Due to robust characteristics for ambiguity and can retain much more information than hard segmentation methods fuzzy c-means (FCM) algorithm is most popularly used than other fuzzy clustering techniques.

Fuzzy C-Means (FCM) Algorithm Process

- Set values for c, m and €. Where 'c' is number of clusters, 'm' is fuzzy factor and '€' is stopping condition.
- Do initialization of fuzzy partition matrix.
- Set the loop counter b.
- Calculate the c cluster centers
- Calculate the membership matrix
- Set b= b+1 and go to step 4.



(a) (b)

Fig.4. Comparison of segmentation results on synthetic test image. (a) Original image with "salt and pepper" noise.
(b) FCM result.

VI. GRAPH BASED CLUSTERING OVERVIEW

Graph-based method for image segmentation can be applied to either grayscale or color images. The assumption

ISSN: 2278 – 909X

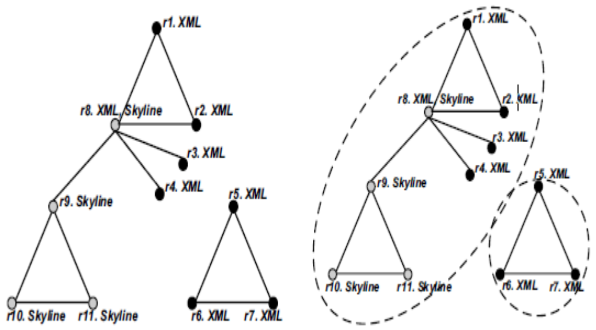
made here is that nearby pixels with similar colors or grayscale intensities may belong to the same region or segment of the image. A graph representation for an image is derived from the similarity among the pixels, and then partitioned by computationally efficient graph clustering method, which first identifies representative nodes for each cluster and then expands them to obtain complete clusters of the graph. Here image can be represented by a proximity graph in which nodes represent image pixels and edges reflect pair wise similarities between the pixels. Using similarity function of properties of corresponding pixels such as location, brightness and color the weight of edges are computed. With this representation, the segmentation task can be solved by graph clustering methods.

Graph Based Clustering Algorithm Process

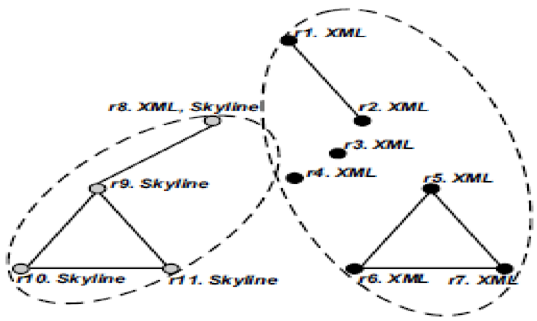
This algorithm partitions an image by building and clustering its proximity graph. The Process of algorithm is drawn below.

- A proximity graph G for the input image I is constructed.
- The sequence of density variation for G is computed.
- Recognize a set of core pixels depend on the sequence of density variation.
- Split the set of core pixels into groups.

Spread out the groups of core pixels to get the image segmentation



(a) Co-author Graph (b) Structure Based Cluster



(c) Attribute Based Cluster

Fig.5. A Co-author Network example for Graph Based Cluster. (a) Co-author Graph, (b) Structure Based Cluster, (c) Attribute Based Cluster

VII. CONCLUSION

There is no objectively "correct" clustering algorithm, but as it was noted, clustering is the eye of image segmentation. The most appropriate clustering algorithm for a particular problem often needs to be chosen experimentally, except there is a mathematical reason to prefer one cluster model than another. It should be noted that an algorithm that is designed for one kind of models has no chance on a data set that contains a radically different kind of models. For example, k-means cannot find non-convex clusters.

REFERENCES

[1] Qi Li, Nikolaos Mitianoudis and Tania Stathaki ,"Spatial Kernel K-Harmonic Means Clustering for Multi-spectral Image Segmentation"

[2] Soumi Ghosh, Sanjay Kumar Dube, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms."

[3] Pedro F. Felzenszwalb, Daniel P. Huttenlocher,"Efficient Graph-Based Image Segmentation."

[4] Radha Chitta, Rong Jin, Rong Jin, Anil K. Jain," Approximate Kernel *k*-means: Solution to Large Scale Kernel Clustering," Dept. of Computer Science and Engineering Michigan State University East Lansing, MI, 48824, USA.

[5] Grigorios Tzortzis and Aristidis Likas," The Global Kernel *k*-Means Clustering Algorithm."

[6] Yang Zhou,Hong Cheng, Jeffrey Xu Yu," Graph Clustering Based on Structural/Attribute Similarities," Department of Systems Engineering and Engineering Management ,The Chinese University of Hong Kong.

[7] Yong Yang,"Image Segmentation by Fuzzy C-Means Clustering Algorithm with a novel penalty term," School of Information Management Jiangxi University of Finance and Economics Nanchang 330013, P.R. China.

Dr.K.SATHIYASEKAR, obtained his Ph.D. Degree in High Voltage Engineering, from Anna University, Chennai. He has a

total teaching experience of 22 years in various Institutions at B.E and M.E levels. He has published / presented 35 research papers in International Journals / Conferences, and has received the **Best Paper Award** for his paper titled "Application of BPN Algorithm for Evaluating Insulation Behavior of High Voltage Rotating Machines" in the International Conference on Digital Factory - 2008, held at Coimbatore Institute of Technology. He was **awarded travel grant by the Department of Science and Technology, Government of India**, to present his research paper in the **International Conference INDUCTICA - 2010,atMesseBerlin,Germany,in2010.**

S.Karthick, obtained his Master Degree in Applied Electronics, from Anna University, Chennai. He has 7 years of teaching experience and 6 years in industries. He has presented 14 research papers in International / national Conference. His main area of research is image processing.

P. Raja Kanniya: Post Graduate Student, M.E (Embedded systems), The Kavery Engineering College, Salem, India.

ISSN: 2278 – 909X