



Application of clustering technique for Image Segmentation

Mohit Agarwal*

Computer Science & Engineering,
Anand Engineering College, Agra, India

Gaurav Dubey

Master of Computer Application
Amity University, Noida, India.

Abstract— *In Computer Science, the term segmentation means the process of splitting the digital image into different parts that is the set of pixels. The aim of this paper is to survey the different clustering methods to perform the segmentation in an efficient manner. The clustering method is recommended to carry out the segmentation of an image in a more efficient manner. Clustering can be defined as the process of grouping of similar kinds of objects in the given sample space. This is basically performed on the basis of different attribute like shape, size, color, texture and other. The main purpose of clustering is to get meaningful outcome from the huge database like image and also for the effective utilization of the image.*

Keywords— *Segmentation, Clustering, Content based image retrieval, supervised clustering, Unsupervised Clustering, Filters*

I. INTRODUCTION

Images are now considered as one of the important method of information conveying. The biggest challenge is to exactly understand the image and extract the information which it contains so it can be used [1]. The first step to understand the image is to partitioned it and find out the different objects in them. Image segmentation means the process of splitting the given image into similar regions on the basis of certain attribute like size, color, shape and texture etc. 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.

Clustering of an image is one of the good techniques which is used for the segmentation of images. After extraction of the feature, these features are treated as vector and are put together into well separated clusters on the basis of each class of image. Clustering can be categorized broadly into two categories: - Supervised and unsupervised clustering. Supervised Clustering requires a human interaction to decide the clustering rules or criteria on the other hand unsupervised clustering requires no human interaction and used to decide the criteria by itself.

II. IMAGE SEGMENTATION

Image Segmentation is the process of splitting the given image in group pixels into homogeneous regions on the basis of common features [2]. Features can be represented by vector of color, texture and many more each of which help in finding the similarity of the pixels in that region [3]. Segmentation of an image is the method of method 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 analyse. Segmentation is basically used for determining the objects and their boundaries (curves, lines, etc) in the given image [4]. 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 analysing the image [5]. The Segmentation may prove helpful in recognition of the object [6], image compression, editing many more. The Segmentation process basically depends upon the quality of image which is to be processed [7]. This proves effective in the case of simple images due to very less variations in the pixels and required some pre-processing in case of complex image.

III. CLUSTERING

Clustering may be defined as the process of grouping the objects which is based on some attributes, so that the objects with similar attributes should lie in the same cluster [8]. Clusters may be defined as the collection of objects which are very much similar in nature on the basis of the attributes and distinct to the objects of other groups or collection. Clustering is one of the important data mining techniques which is used for the purpose of data analysis, image processing, pattern recognition and more. An image can be think as a collection which is grouped on the basis of keyword. In the clustering which is based on the concept of metadata, a keyword is a form of font or attribute which is used to represent the different features or attributes of an image. The common features of an image are supposed to grouped into same clusters by assigning them some value. While in content based clustering, the content actually refers to the color, texture, shape or any other feature or information that can be extracted from the given image..Clustering method also includes partitioning of the clustering which involves the formation of clusters that partitions the given data into similar groups. In hierarchical based clustering, a hierarchy of clusters for each and every element is maintained. A good number of clustering based techniques are available for the image segmentation process, to carry out the same in

more efficient and effective manner. The clustering methods which we are using in this paper to carry out our survey are relevance feedback, graph based clustering log based clustering, hierarchical clustering, Retrieval-dictionary based and filters based clustering and many more.

IV. CLUSTERING TECHNIQUES

An image used to contain more than one objects so to partition it in accordance with the available features to extract the meaningful results from it now become a big challenge. One of the good methods to achieve the segmentation is clustering.

Clustering algorithms can be classified into two broad categories namely supervised which also include semi-supervised and unsupervised techniques. The latter techniques include density based algorithm e.g. fuzzy c – means. The earlier one require human interaction for good results.

Through this paper we are going to review clustering techniques from segmentation point of view.

A. **Relevance Feedback**

A relevance feedback is one of the oldest methods which was introduced in information retrieval field [4] but it find the more importance in CBIR field [5]. This algorithm allows a user to interact with the image retrieval mechanism by providing the basic feature information which the user finds relevant to the query. This algorithm proved very effective over keyword based image retrieval which involves retrieval of the image on the basis of the keyword as provided by the user and image in the database.

The main challenge for keyword based retrieval is that some image may not have appropriate keyword to describe them and this make the search quite complex. Because of these reasons the relevance feedback algorithm is used for segmentation which use users feedback to reduce the possible errors and duplicacy [6][7]. The underlying technology uses by this method is Bayesian Classifier which used to deals with both positive and negative feedback.

The main drawback with the relevance feedback algorithm for image segmentation is that it is static in nature i.e. it cannot handle the changes in user requirement like addition of new topic. For the improvement in image segmentation process the log –based clustering algorithm brought into the practice.

B. **Log –Based Clustering**

Image segmentation using clustering can also be achieved by the logs which are maintained by information retrieval process (like web servers access logs). But log –based clustering cab result in incomplete information. This is because log –based clusters are based on those documents which are accessed by some users. To overcome this problem a log –based vector is maintained for each session vector which is based on log –based documents [8]. One vector used to represent a given document in a hybrid matrix form. And for those documents that are in a log based document is created in a hybrid matrix form. And also the un accessed document creates their own vector. A hybrid matrix is used to represent the documents which are a cluster of content based clustering algorithm.

The log –based clustering algorithm is quite accurate in comparison to previous algorithm but this is not very efficient to use.

C. **Hierarchical Clustering**

The hierarchical clustering is one of the important clustering techniques for the retrieval of information [9]. This process involves integration of different images and the creation of clusters in the form of tree from those images. There are two main method of hierarchical clustering.

First one is known as agglomerative approach, which is also known as bottom up approach in which we start from the bottom where all objects are presents and then we used to move up by merging the objects. This process is continued until all the objects are merged into a single cluster.

Second one divisive method which is also known as top down approach where we think all objects as the single group and then we split them recursively into two until all groups contain a single object.

The steps required for the agglomerative hierarchical clustering are as follows [10]:-

1. First convert the objects features into the distance matrix.
2. Treat each object as the cluster itself.
3. Iterate until number of is 1.
 - a. Now merge two closest clusters.
 - b. Update the distance matrix.

D. **Retrieval Dictionary based algorithm**

The supervised clustering algorithm involves a formation of rough classification retrieval system. This is incorporated by calculating the distance between two learned patterns, then these learned are classified into different clusters which is followed by image retrieval stage. The challenge faced by this technique is the determination of the distance which is used in classification.

The solution to the above problem can be retrieval system which is developed by retrieval dictionary based clustering [11]. This technique involves a retrieval dictionary generation unit which used to classify the learned pattern into plural clusters which help in creating the retrieval dictionary using these clusters[R]. The image retrieval in this technique is based on the distance between two spherical clusters that have distinct radius. Each radius refers the similarity measure between the centrally located cluster and an image retrieved. So that the image similar to the queried image or pattern will be obtained using the retrieval dictionary algorithm.

E. **K means Clustering**

In this unsupervised clustering technique the data vector which is obtained are grouped into predefined clusters [12][13]. In the beginning the centroids of the predefined clusters are initialized at randomly. There is always at least one

object in each clusters. 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 [14]. 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.

F. Fuzzy Clustering

In this method of image segmentation a group of images are allowed to form a cluster on the basis of similarity checks by using the fuzzy clustering, in which each image is represented by a node of a graph [15]. The graph obtained is finally divided into the sub-graphs which used to represent the real clusters from the various available images. The division or partitioning is based on the generation of the true cluster at each and every stage and include removal of the nodes that are belongs to some true clusters and allows the ungrouped nodes to be clustered again. In the beginning of each stage, the nodes to be clustered are considered as all of them are belonged to the single clustered. And then nodes are used to be removed from this single cluster in accordance with the similarity criteria to form a true cluster. The existing member nodes of this cluster are then removed by following some process before the next step of clustering will take place.

The drawback with this system of segmentation is that the image content is usually inadequately defined which used to affect the performance of the existing system like recall and precision. The new system is developed with the intention of maximizing these performance factors [16]. This new system used to identify the similarity between the structures of the images which is present in the one or more images. The characterization parameter set is maintained which allocated in accordance. The similarity index is also maintained which is based on the characterization parameter set. To enhance the features and for improving the efficiency of the image the various kinds of filters like convolution filter, Look-up-table (LUT), frequency domain filter and many more are used.

V. CONCLUSION

Through this paper we have presented a very rigorous survey for identifying the clustering algorithms which can be used for the image segmentation. This is also proved that with the clustering we can perform the image segmentation in a very effective manner. Clustering algorithm on which we focused in this survey are from both the categories i.e. supervised and unsupervised. The Hierarchical clustering and relevance feedback are the example of supervised one while the retrieval dictionary based clustering, k means clustering, fuzzy clustering are the example of unsupervised clustering.

In spite of the improvement in the performance by using the clustering the image segmentation process still needs some improved technology for the feature extraction because large database used cause variation in feature representation. The supervised clustering techniques are suffering from the drawbacks like they cannot effectively handle the situation if the available feature vector does not cover the query raised by the user then the result obtained will be less meaningful. The unsupervised algorithms are suffering from inefficiency problem due to this they require a serious computational efforts for image segmentation. In future we will try to explore the probable solutions to the current challenges

REFERENCES

- [1] Datta R, Joshi D, Li J, Wang JZ. Image Retrieval: ideas, influences, and trends of the new age. *ACM Comput Surv* 2008; 40: 1-60.
- [2] 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.
- [3] 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.
- [4] G. Salton and M.J. McGill, *Introduction to modern information retrieval*, New York, McGraw-Hill, 1988.
- [5] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", 2nd ed., Beijing: Publishing House of Electronics Industry,2007
- [6] 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.
- [7] J. Winn and J.Shotton, "The Layout Consistent Random Field for Recognizing and Segmenting Partially Occluded Objects", *proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, vol. 1,pp. 37-44, 2006.
- [8] Giacinto G., Roli F.: Bayesian relevance feedback for content-based image retrieval. *Pattern Recognition* 37(7) (2004) 1499-1508
- [9] Han,J. ; kamber, M. *Data Minining: Concepts and Techniques*, Morgan Kaufmann.2001
- [10] Zhang, H.J., Su, Z., Zhu, X.: US20067113944 (2006).
- [11] Huang Min,Sun bo,Xi Jianqing"An Optimized image retrieval method based on Hierarchical clustering and genetic algorithm"Intl forum on Information technology and applications,978-0-7695-3600-2/09-IEEE,2009
- [12] Parvinder S. Sandhu and Hardeep Singh, "A Neuro-Fuzzy Based Software Reusability Evaluation System with Optimized Rule Selection", *IEEE 2nd International Conference on Emerging Technologies (IEEE ICET 2006)*, Peshawar, Pakistan, Nov. 13-14, 2006, pp. 664-669
- [13] Smeulders A.W.M., Worring M., Santini S., Gupta A., Jain R.: Content-based image retrieval at the end of the early years. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 22(12) (2000) 1349-1380
- [14] Irani, A.A.Z. Belaton, "A K-means Based Generic Segmentation System B.Dept. of Comput. Sci., Univ. Sains Malaysia, Nibong Tebal, Malaysia Print ISBN: 978-0-7695-3789-4 On page(s): 300 – 307, 2009.
- [15] Jafarkhani, H., Tarokh, V.: US20077260263 (2007).
- [16] Galperin, M.: US20097483919 (2009).