Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-5, 2016
ISSN: 2454-1362, http://www.onlinejournal.in

Implementation And Comparative Analysis Of Image Segmentation Algorithms

Aishwarya Muktewar[1], Rajlakshmi Khode[2], Aishwarya Kasulkar[3], Runali Dongare[4], Neha Dhobale[5] & Pawan Khade[6]
[1,2,3,4,5]Computer Science & Engineering , Rajiv Gandhi College of Engineering & Research, Nagpur, Maharashtra, India
[6]Assistant Professor, Computer Science & Engineering , Rajiv Gandhi College of Engineering & Research, Nagpur, Maharashtra, India

Abstract*: In day to day life new technologies are emerging in the field of image processing. Since a vast amount of data is generated in almost every area such as business ,engineering ,health science etc ,people are more interested in the clustering algorithms. This is done in order to manage and maintain the such a vast amount of data. In this paper, we give a short review of segmentation algorithms based on clustering which are implemented on android platform. Image segmentation is the process of partitioning any digital. image into multiple segments. The difference is typically based on various parameters based on pixel color, intensity, texture etc. Clustering is one of the method of segmentation. The main objective of paper is to provide a comparative analysis of clustering algorithms.*

1. INTRODUCTION

This paper is based on the wide domain named as image processing. Image processing can be defined as the method in which image is converted into some digital form by performing some operations on it. Thus this paper gives a short review of some of the clustering algorithms which is the method of segmentation .

A. Image Segmentation:
Image processing is a field which is concerned with extracting information from images. Image segmentation is the process of partitioning an image into multiple segments. The task of image segmentation is to simplify the problem by

combining the pixels in the image. Image segmentation is an important image processing technique which is used to analyze what is inside the image. Image segmentation is used to separate an image into several “meaningful” parts and there are many possible ways of meaningfully grouping pixels.[3] Segmentation is used to provide additional information about any image which uses standard parameters as intensity, color, texture etc. There are various image segmentation techniques but very few provides better results. Some of the image segmentation techniques won’t work for all the images. There are many methods to solve image segmentation problems. In image segmentation each technique has its own advantages and also disadvantages, so it’s hard to tell which one is better in all the techniques. Clustering is one of the image segmentation algorithms.

B. Image Clustering
Clustering is the “process of organizing objects into groups whose members are similar in some way”. Thus we can define a cluster as “collection of objects which are similar between them and arranging the dissimilar to the objects belonging to other cluster”.
Clustering is a typical method of grouping data points in some or the other ways. The performance of most clustering algorithms is dependent on the exact value of the cluster numbers which is not possible to estimate in the real time application.

2. ARCHITECTURE

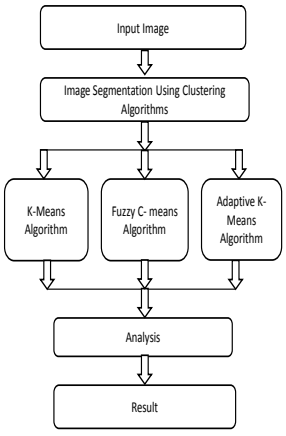


Fig. System Architecture.

This is the proposed architecture for the implementation of the various image segmentation

algorithms .Here user has to provide the input image which has to undergo the clustering algorithm as that of the K Means, Adaptive K Means and the fuzzy C Means algorithm. At the end analysis is provided of all above algorithms and the analysis result is generated.

1. CHALLENGES OF CLUSTERING

Since there are various algorithms available for the clustering, it is very tedious to choose the appropriate algorithm according to the required condition. Therefore, many problems in the existing algorithms; some algorithms have limitations when the clusters differ in sizes, densities, having different shapes. Some algorithms are sensitive to noise and outliers. Each algorithm has its own run time, complexity, error frequency etc. Another issue may be that the outcome of a clustering algorithm mainly depends on the type of dataset used. If the size of the dataset is increased then it becomes difficult to maintain the such a large datasets . The complexity of data set increases with the data like audios, videos, pictures and other multimedia data which form very heavy database. We always need to choose between various different clustering algorithms and it becomes really tedious sometimes.

Thus the main challenge is to select the correct type of clustering algorithm which is based on the type of dataset, accuracy, fault tolerance etc in order to get the required result.[2]

III .CHALLENGES IN CLUSTER ANALYSIS

In our survey we came across some of the problems in the analysis of cluster analysis.
1].The main challenge is faced during the identification of distance measure that are used for the numerical attribute that can be used for the standard equation like Euclidean distance.
2] Another challenge is faced in identifying the number of cluster which is not easy to determine beforehand.

3] At the same time it is tedious to choose the initial cluster. If the initial cluster is not properly chosen then after the few iterations it is found that cluster may even be left empty.

IV] RELATED WORK

A] K-MEANS CLUSTERING

K-Means is a commonly used clustering algorithm used for image processing. Clustering is a used for arranging the data points into required number of clusters where each cluster has maximal similarity as defined by an objective function. Each data point must belong to only a single cluster, and the union of all clusters contains all data points. Each data point is assigned to the cluster whose center (also called centroids) is nearest. The center is calculated by finding the average of all the data points in the cluster that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster.[4] The algorithm steps are:
1) select the number of clusters, k.
2) Randomly generate k clusters and determine the cluster centers, or directly generate k random points as cluster centers.
3) Assign each point to the nearest cluster center.
4) Re-compute the new cluster centers.
5) Repeat the two previous steps until some convergence criterion is met (usually that the assignment hasn't changed).
The main disadvantage of this algorithm is that it does not produce the same result with each run, since the resulting clusters depend on the initial random assignments. It minimizes intra-cluster variance, but does not ensure that the result has a global minimum of variance. If, however, the initial cluster assignments are heuristically chosen to be around the final point, one can expect convergence to the correct values.

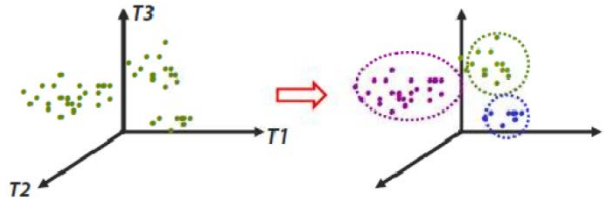


Fig b) The left-hand side is the original data points in the 3-dimensional space, and the right-hand side is the k-means clustering result(3 groups)

1] ADVANTAGES

1. For large number of variables, K Means algorithm may be faster than hierarchical clustering, when k is small.
2. K-Means may produce constricted clusters than hierarchical clustering, if the clusters are globular.

2] DISADVANTAGES
1. Difficulty in comparing quality of the clusters formed.
2. Fixed number of clusters can make it difficult to forecast what K should be.
3. Does not give good result with non-globular clusters. Different primary partitions can result in different final clusterrs
4. Different initial partitions can result in different final

B] FUZZY C-MEANS ALGORITHM

The Fuzzy C- Means (FCM) is another clustering algorithm that follows the same principle as that of the K-Means algorithm as stated above. The only point of difference between the K means algorithm and the Fuzzy C-Means algorithm is that the FCM algorithm takes the decision about which cluster the pixel should belong to, depends on the value that it assigns between 0 and 1which describes the belongingness of that particular pixel to that cluster.
As in the Fuzzy rule state that the summation of the all the membership value of the pixel of any particular cluster should be equal to 1.If the membership value is high the pixel belongs more to the cluster. Thus the formula given below is obtained by minimizing the objective function as shown in equation(1)

$$J = \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik}^{m} |p_i - v_k|^2 \quad (1)$$

Where :

J is the objective function
n is the number of pixel in the image E
c is the number of cluster
μ is the fuzzy membership value
m is the fuzzy membership value
pi is the pixel in image E
vk is the centroid in the cluster
|pi-vk| is the Euclidean distance between pi and vk

The Euclidean distance is calculated using the following equation (2)

$$|p_i - v_k| = \sqrt{\sum_{i=1}^{n} (p_i - v_k)^2} \quad (2)$$

Also the FCM algorithm can also be used for the colour images containing the RGB components colour space. Thus the Euclidean distance for the same can be calculated by using the following equation (3)

$$|p_i - v_k| = \sqrt{\sum_{i=1}^{n} \left(p_{iR} - v_{kR}\right)^2 + \left(p_{iG} - v_{kG}\right)^2 + \left(p_{iB} - v_{kB}\right)^2}$$

Thus as we know that it is the iterative process. Thus the algorithm steps are as follows.
1)Initially set the number of cluster, the fuzzy parameters, as well as the stopping condition of the algorithm.
2) Initialize the fuzzy partition matrix
3)Set the loop counter k=0
4) Calculate the cluster centroid and the fuzzy objective function J
5) For each pixel of every cluster calculate the membership values in the cluster
6) If the value of J between the consecutive iterations is less than the stopping condition, then stop otherwise set k=k+1 and go to step 4
7) At last do the segmentation.

ADVANTAGES

1) Gives best result for overlapped data set and comparatively better then k-means algorithm.
2) Unlike k-means where data point must exclusively belong to one cluster center here data point is assigned membership to each cluster center as a result of which data point may belong to more then one cluster center.

DISADVANTAGES
1) A priori specification of the number of clusters.
2) Euclidean distance measures can unequally weight underlying factors.

C] ADAPTIVE K MEANS ALGORITHM

Adaptive K Means Algorithm is another clustering algorithm as that of the above two stated algorithm. The adaptive k means clustering algorithm is similar to that of the traditional K Means algorithm some of the differences. It selects K elements from the input dataset and become the seed of the cluster and these are randomly selected. The properties of these seed elements become the properties of the cluster ultimately.
This algorithm is based on the ability to compute distance between the data point and the cluster center as well as the distance between the two data points. Thus in order to calculate the distance we use the Euclidean Distance formula as stated above.
Thus the algorithm steps are as follows:
1) Initially we need to compute the distance of each cluster from every other cluster which is stored in the matrix form. We also calculate the minimum distance between cluster C1 and C2 and denote it as dmin.
2) Now for each unclustered element E in the given dataset compute the distance for each cluster. Now in order to assign this element to a certain cluster we need follow the following three conditions:
A) If the distance of the element and the cluster center is 0 then assign that element to that cluster and then move on to the another cluster.
B) If the distance is less than dmin then assign it to the nearest cluster. In this case the cluster center is updated. This is done in such a way that the average properties of all the cluster data points is summed up and assign it to the cluster center.
C) If the distance dmin is less than the distance of the element from the nearest cluster then we select two nearest cluster C1 and C2 and then mergeC2 with

C1.Thus we destroy the cluster C2 and delete all its representation .Now the distance between all the clusters are recomputed.

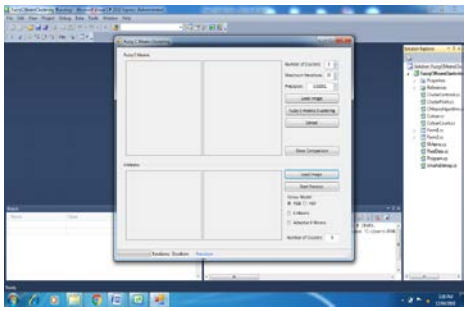
ADVANTAGES

1) In this method of clustering proper cluster are obtained as compared to the traditional K Means algorithms since the cluster center are updated each time.
2) Here we need not need to provide the number of iterations prior in the process.

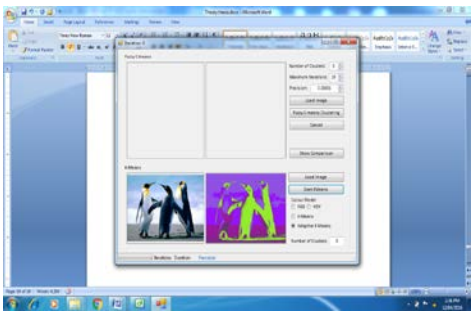
DISADVANTAGES

1) The time complexity of the algorithms is increased as the cluster centers are updated each time.

IMPLEMENTATION

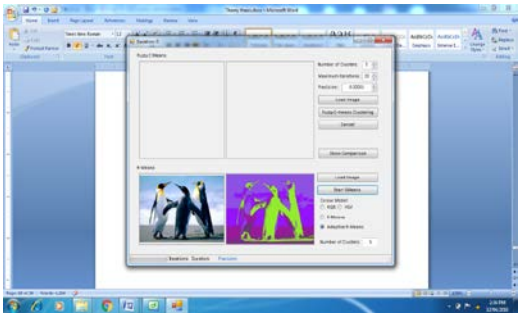


For implementing the K Means algorithm firstly click on the button of the load image for taking any image as input.



Click on the button named as that for the K Means algorithm and the clustered image using the K Means algorithm is been displayed.

Click on the fuzzy C Means algorithm and select the image of your choice.Also provide with the number of iteration that you have to give. The clusterd image using the Fuzzy C Means algorithm is been displayed in the block next to the original image after the completion of the number of iteration.



click on the Adaptive K Means algorithm and the clustered image using that algorithm is been displayed in the block.

4. CONCLUSION

Thus from the above study and implementation we concluded that the time required by the K Means algorithm is less as compared to the other two but it do not provide us with that appropriate clustering result. But at the same time the Fuzzy C Means algorithm requires more time for clustering as it is an iterative process but provides proper clustering result.
All this can be observed from the comparative analysis of these algorithms. Similarly adaptive K Means algorithm provides similar result as that of the K Means algorithm with a slight difference in the observation.

5. REFERENCES

[1] Soumi Ghosh Department of Computer Science and Engineering ,Amity University,Utttar Pradesh,Noida,India, "Comparative Analysis Of K-Means And Fuzzy C-Means Algorithms",(IJACSA)Interational Journal of Advance Computer Science and Application,Vol.4,No.4,2013.

[2] V. S. Rao and Dr. S. Vidyavathi, "Comparative Investigations and Performance Analysis of FCM and MFPCM Algorithms on Iris data", Indian Journal of Computer Science and Engineering, vol.1, no.2, 2010 pp. 145-151.

[3] S. Borah and M. K. Ghose, "Performance analysis of AIM-K-Means and K-Means in

quality cluster generation", Journal of Computing, vol. 1, Issue-1, 2009.

[4] Y. Yong, Z. Chongxun and L. Pan, "A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding", Measurement Science Review, vol. 4, no.1, 2004.

[5] https://sites.google.com/site/dataclusteringalgorithms/fuzzy-c-means-clustering-algorithm.

[6] https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm.

[7] BY S.VAMSI, " IMAGE SEGMENTATION" (Thesis).

[8] Yu-Ning Liu,Chung-Han Huang, Wei-Lun Chao "Image segmentation based on the normalized cut framework " DIP: Final project report(Thesis)

[9] Tapas Kanungo, Nathan S. Netanyahu"An efficient K Means Clustering algorithm: Implementation and analysis" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 7, JULY 2002

[10] Sanjiv K. Bhatia "Adaptive K-Means Clustering"

[11] Caroline Pantofaru, Martial Hebert" A Comparision of image segmentation algorithms"