International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 3, Issue 4, July-August 2014 ISSN 2278-6856

IMAGE SEGMENTATION IN VARIOUS DOMAINS USING TWO PHASE CLUSTERING APPROACH

Divya Sindhu*[1], Surender Singh[2]

[1]M. Tech. Scholar, Department of Computer Science & Engineering,
Om Institute of Technology and Management, Hisar

[2]Asstt. Professor, Department of Computer Science & Engineering,
Om Institute of Technology and Management, Hisar

Abstract

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. However, its performance in terms of global optimality depends heavily on both the selection of k and the selection of the initial cluster centers. Mean Shift clustering algorithm does not rely upon a priori knowledge of the number of clusters. Therefore, mean-shift can be utilized in initial phase for finding number of clusters and k-means in second phase for proper segmentation. In this paper, the importance of two-phase approach has been studied for images with non-uniform and noisy background like nano images, ultrasound images and Infrared images.

Keywords: Image Segmentation, Clustering, K-Means, Mean Shift.

1. INTRODUCTION

Segmentation is the process of partitioning a digital image into multiple segments known as super pixels. This is typically used to locate objects and boundaries in images. Images can be segmented into regions by widely used clustering algorithms. Clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. It is the computational task to partition a given input into subsets of equal characteristics. These subsets are usually called clusters. It is a main task of exploratory data mining, and a common technique for analysis of statistical data, used in many fields including image analysis, pattern recognition, machine learning, information retrieval and bioinformatics. Many variations of k-means clustering algorithm[1],[2],[3],[4],[5],[6] have been developed recently. The procedure of k means algorithm follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters). The first step is to find the k cluster centers and assign the objects to the nearest cluster center, such that the squared distances from the cluster are minimized. K centroids will be defined, one for each cluster. These centroids should be placed in such a way that cluster label of the images does not change anymore. Variations of k-means often include such optimizations as choosing the best of multiple runs, but also restricting the centroids to members of the data choosing medians (k-medians clustering), choosing the initial centers less randomly (K-means++) or allowing a fuzzy cluster assignment (Fuzzy c-means)[7]. The mean shift procedure was originally presented in 1975 by Fukunaga and Hostetler. Basic mean-shift clustering algorithms maintain a set of data points the same size as the input data set. Initially, this set is copied from the input set. Then this set is iteratively replaced by the mean of those points in the set that are within a given distance of that point. On the other hand, *k*-means restricts this updated set to *k* points usually much less than the number of points in the input data set, and replaces each point in this set by the mean of all points in the *input set* that are closer to that point.

2. TWO PHASE CLUSTERING APPROACH

The advantage of Mean Shift over *k*-means is that Mean Shift clustering does not rely upon a priori knowledge of the number of clusters. Therefore, we can utilize mean-shift in initial phase for finding number of clusters and k-means in second phase for proper segmentation. In this paper, the importance of two-phase approach has been studied for images with non-uniform and noisy background like nano images, ultrasound images and IR images. Mean shift algorithm clusters an n-dimensional data sets. For each point, mean shift computes its associated peak by first defining a spherical window at the data point of radius r and computing the mean of points that lie within the window. Algorithm then shifts the window to the mean and repeats until convergence[8]. At each iteration, the window will shift to a more densely populated portion of data set until peak is reached where data is equally distributed. Mean Shift is an iterative method consists of the following steps:

1. Initialise estimate *d*.
2. Initialise $K(d_i - d), K(d_i - d) = e^{c\|d_i - d\|}$ be a given Kernel function. The weighted mean of the density in the window is determined by *K*.

$$m(d) = \frac{\sum_{d_i \in N(d)} K(d_i - d) d_i}{\sum_{d_i \in N(d)} K(d_i - d) d_i} \quad (1)$$

is where *N* (*d*) is the neighbourhood of *x*, a set of points for which $K(d) \neq 0$.

3. Now, set $d_i \leftarrow m(d)$, and repeats the estimation until *m(d)* converges. An example illustrating mean shift procedure is shown in Fig. 1. The shaded and

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 3, Issue 4, July-August 2014 **ISSN 2278-6856**

black dots denote the data points of an image and successive window centres, respectively. Mean shift procedure starts at point Y1, by defining spherical window of radius r around it. Algorithm then calculates the mean of data points that lie within the window and shifts the window to the mean and iterates the same procedure until peak is reached. At each iteration, window is shifted to the more densely populated region.

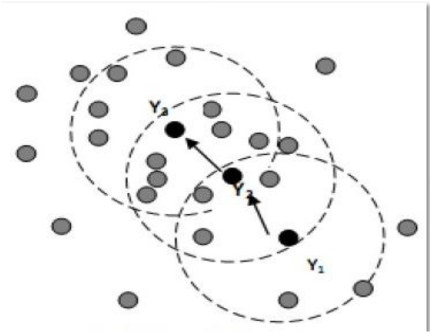

Fig.1. Mean Shift Procedure

The k means algorithm takes the input parameter K, and partitions a set of n objects into k clusters so that the resulting intra-cluster similarity is high whereas, the inter-cluster similarity is low. Cluster similarity is measured by the mean value of the objects in a cluster. k means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other[9]. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The points are clustered around centroids $\mu_i \forall i = 1 \ldots k$ which are obtained by minimizing the objective

$$V = \sum_{i=1}^{k} \Sigma_{x_j = s_i} \left(x_j - \mu_i \right)^2 \qquad (2)$$

where there are k clusters S_i, i = 1,2,…, k and μ_i is the entroid or mean point of all the points $x_j \in S_i$ [4].

In this study, an iterative version of the algorithm is presented for image segmentation. The algorithm takes a 2 dimensional image as input. Various steps used in the algorithm are as follows:

Input

The number of clusters k and a database containing n objects.

Output

A set of k clusters which minimises the square error criterion.

Method

1. Take K as the initial cluster centers as calculated by the mean shift algorithm.
2. Repeat the following steps 3 and 4 until the cluster labels of the image do not change anymore.
3. Cluster the points based on distance of their intensities from the cluster center using distance formula.

$$c_{(i)} := \arg \min_{j} \|x^{(i)} - \mu_j\| \qquad (3)$$

4. Compute the new centroid for each of the clusters.

$$_i := \frac{\sum_{i=1}^{m} 1\{c_{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c_{(i)} = j\}} \qquad (4)$$

where k is a parameter of the algorithm (the number of clusters to be found), i iterates over the all the intensities, j iterates over all the centroids and μ_i are the centroid intensities.

5. The resultant new values are used to find out again the mean value until convergence.

3 .RESULTS AND ANALYSIS

Nano Image

Nano-structured materials attract a growing attention due to their superior mechanical and physical properties. Such properties are inherently related to the unique structure which is controlled at the nano-scale. The ultra-high resolution images can be efficiently processed to obtain quantitative description of the nano-particles of cell segmentation.

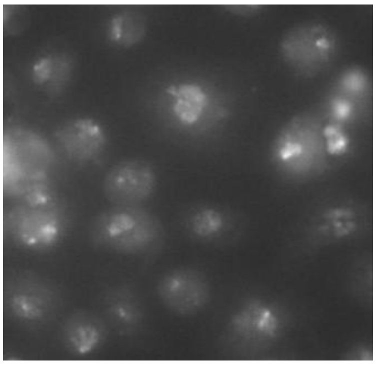

Fig. 2(a) Original Nano Image

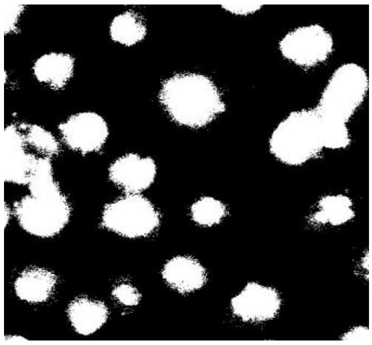

Fig. 2(b) with mean shift at k = 3

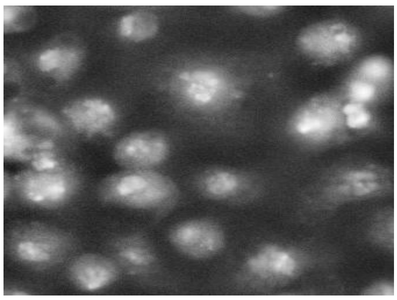

Fig. 2(c) Without mean shift

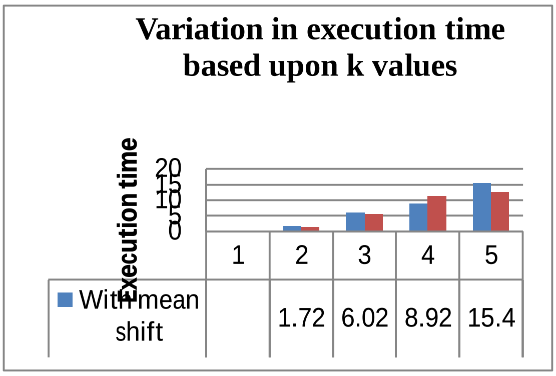

Fig. 2(d) Variation in execution time based upon k values.

Ultrasound Image

Fig. 3(a) Original Ultrasound Image

This image represents the cancer in the bile duct of human body. Bile duct cancer *begins near the liver and is an aggressive type of* cancer.

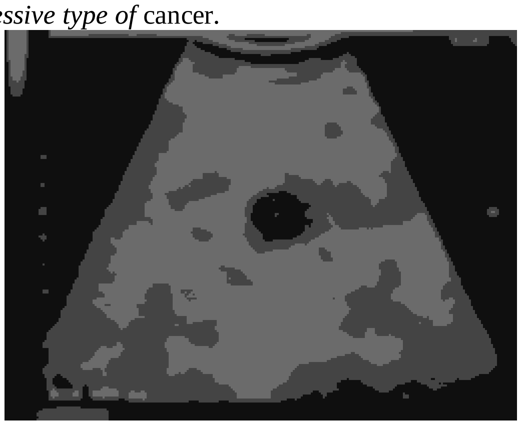

Fig. 3(b) With Mean Shift at k=3

Fig. 3(c) Without Mean Shift

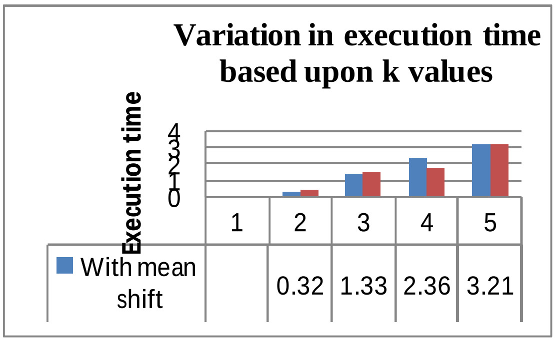

Fig. 3(d) Variation in execution time based upon k values.

Infrared Image

Many surveillance systems are used with human as an object. This image is an Infrared Image. So it is important and useful to segment the human body from an image.

Fig. 4(a) Original Infrared Image

Fig. 4(b) With Mean Shift at k=3

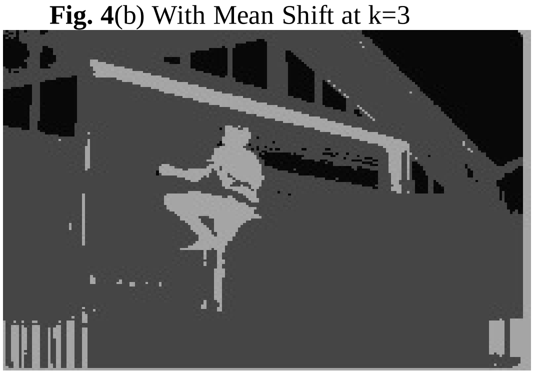

Fig. 4(c) Without Mean Shift

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 3, Issue 4, July-August 2014 **ISSN 2278-6856**

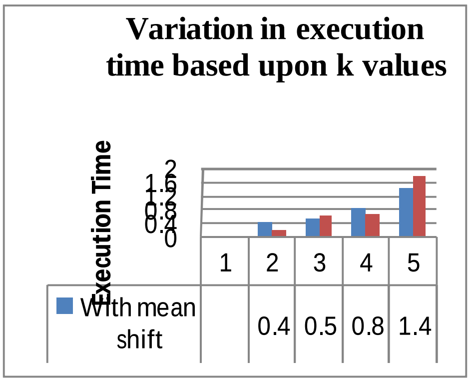

Fig. 4(d) Variation in execution time based upon k values.

4 .CONCLUSION

Graphs show that with increasing value of k, there is not much significant variation in execution time. Therefore, Two Phase approach outperforms the K-Means algorithm and this can be seen in the results shown above for the three domains. Thus, Mean-Shift can be used for initialisation purpose in k-means and results can be studied for these domains where execution time is important apart from segmentation.

REFERENCES

[1] F Gibou and R Fedkiw," A Fast Hybrid k-Means Level Set Algorithm for Segmentation", 4th Annual Hawaii International Conference on Statistics and Mathematics, pp. 1-11; 2002.

[2] K Sravya and S.V. Akram, "Medical Image Segmentation by using the Pillar K-means Algorithm", International Journal of Advanced Engineering Technologies Vol. 1-Issue 1, pp. 1080–1087; 2013.

[3] G Pradeepini and S. Jyothi, "An improved k-means clustering algorithm with refined initial centroids", Publications of Problems and Application in Engineering Research, Vol 04, Special Issue 01; 2013.

[4] G Frahling and C Sohler, "A fast k-means implementation using coresets", International Journal of Computer Geometry and Applications 18(6): pp. 605-625; 2008.

[5] Charles Elkan, "Using the Triangle Inequality to Accelerate k-means", Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, pp. 147-153; 2003.

[6] K Wagsta, C Cardie, S Rogers and S Schroedl, "Constrained K-means Clustering with Background Knowledge". Proceedings of the Eighteenth International Conference on Machine Learning, pp. 577-584; 2001

[7] M Alata, M Molhim and A Ramini, " Using GA for Optimization of the Fuzzy C-Means Clustering Algorithm", Research Journal of Applied Sciences, Engineering and Technology 5(3): pp.695-701; 2013.

[8] V Sumant, A. Joshi and N Shire, "A review of an enhanced algorithm for color Image Segmentation", International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 3, pp. 435-440; 2013.

[9] Suman Tatiraju, Avi Mehta ,"Image segmentation using K-means clustering, EM and Normalized Cuts", 2008.

AUTHOR

Divya Sindhu received B.Tech. and M.Tech. degree in Computer Science & Engineering from Om Institute of Technology & Management in 2012 and 2014 respectively. Her area of interest includes Data Mining, Software Engineering and Neural Networks.

Surender Singh received B.Sc. degree from KUK in 2004 and then received the MCA and M.Tech. degree from Guru Jambheshwar University of Science & Technology in 2007 & 2010 respectively. His area of interest includes Computer Algorithms, Data Structure Data Mining, Computer Networks and Mobile networks.