

# ADAPTIVE K-MEANS CLUSTERING FOR MEDICAL IMAGE SEGMENTATION

Jyothsna C<sup>1</sup>, Dr.G.R.Udupi<sup>2</sup>

<sup>1</sup>Assistant Professor, Electronics and Communication Department  
KLS's VDRIT, Haliyal, India

<sup>2</sup>Professor, Electronics and Communication Department  
KLS's Gogte Institute of Technology, Belgaum, India

<sup>1</sup>jyothsnasc18@gmail.com

<sup>2</sup>grudupi@gmail.com

**Abstract**— One of the major challenges in image analysis is image segmentation. In medical applications, skilled operators usually extract the desired regions that may be anatomically separate but statistically indistinguishable. It is subjected to manual errors and biases, which is time consuming, and has poor reproducibility. The problem faced in clustering is the identification of clusters in given data. A widely used method for clustering is based on K-means in which the data is partitioned into K number of clusters. In this method, clusters are predefined which is highly dependent on the initial identification of elements representing the clusters well. Several researchers in clustering has focused on improving the clustering process such that the clusters are not dependent on the initial identification of cluster representation. The proposed method advances an adaptive technique that grows the clusters without the initial selection of elements representing the cluster. It is found to be capable of segmenting the regions of smoothly varying intensity distributions. The technique has been used to achieve an notable accelerated search process.

**Keywords**— Segmentation, Clustering, Adaptive K means.

## I. INTRODUCTION

### A. Segmentation

Digital imaging technologies have become indispensable components for clinical procedures. Major advances in the field of medical imaging and computer technology have created opportunity for quantitative analyses of medical images and provided powerful techniques to probe the structure, pathology and function of the human body [1]. The availability of many different imaging modalities increased the requirement for significant innovations to obtain accurate and fast results in all aspect of image processing.

Some of the current researches in medical imaging are focused on pathology diagnosis of abdominal organs such as spleen, liver, kidney and gallbladder. Segmentation is an important step for radiological operations such as diagnosis, study of anatomical structures, quantification of volumes of tissues, treatment planning, localization of pathologies and computer aided surgery [1]. Manual segmentation process is not only tedious and time consuming due to the high number of slices but also depends on the skills and experience of the operator.

For example, Figure 1.1 shows liver edges identified manually by different radiologists and also by the same radiologist at a different time on a sample Magnetic Resonance Image (MRI). Therefore, automatic segmentation is a fundamental step.

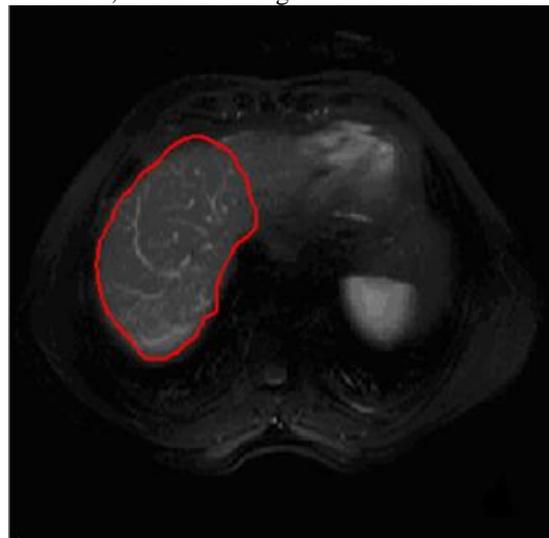


Figure 1 (a). Liver edges defined by radiologist R1 at time T1

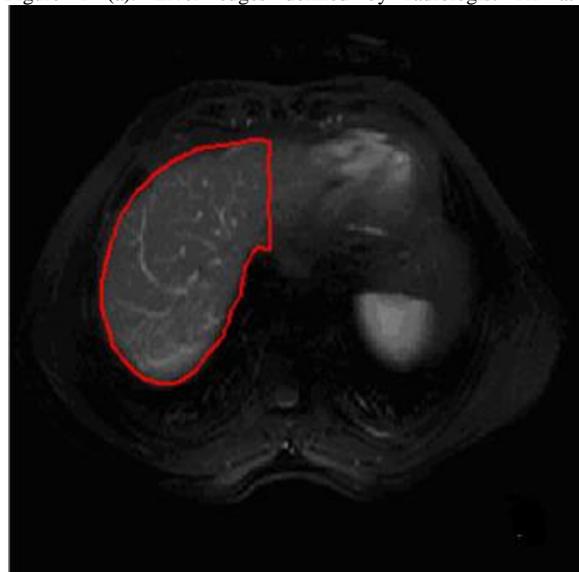


Figure 1(b). Liver edges defined by radiologist R2 at time T2

Segmentation is performed to partition an image into multiple parts. The aim of this process is to simplify or/and change the representation of the digital image into something that is easier to analyze and more meaningful. Some examples for biomedical image segmentation applications can be given as examination of computer guided surgery, anatomical structure, planning of treatment, diagnosis, measurement of tissue volumes to locate tumors and other pathologies [2]. Manual segmentation is error prone, very time consuming because of the high number of slices and tedious. It is performed by an expert using a device that is generally a mouse and a software program for labeling edge pixels. Therefore, it depends on the skills and experience of the expert that means manual segmentation method is a subjective

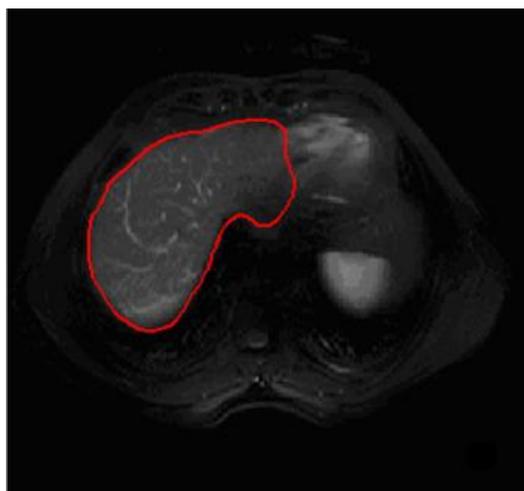


Figure 1(c). Liver edges defined by radiologist R2 at time T3

process and not reproducible. Therefore, computer-aided medical image analysis is indispensable. There is not any general solution for image segmentation problems. Several researchers used many different fully automatic and semi-automatic image segmentation techniques for different images. In practice, segmentation algorithms usually require knowledge and experience of radiologists to adjust segmentation parameters to get optimal performance. Mostly computerized systems run in an interactive or semi-automatic manner because of the complexity of medical image segmentation [1] which means that the radiologists initiate the segmentation and interrupt it when necessary, and finally stop the algorithm. Generally, initial information such as an initial seed point or an initial boundary is inserted to the process for initialization by the radiologists in these methods. Semi-automatic techniques are usually known by region growing and active contour methods. In active contour method, a cost function obtained by the initial boundary is minimized. In region growing method, connections of pixels are checked by starting from initial information according to a criterion. The methods in fully automatic segmentation approaches are applied without any user intervention by using several

parameters. However, the values of these parameters can not be identified in a strict bound because of the variations in anatomy and also characteristics of images.

An image segmentation technique can be in the group of semi-automatic or fully automatic segmentation methods according to its implementation way. Segmentation techniques can be classified as thresholding based, clustering based, atlas based, model based, Artificial Neural Network (ANN) or PDE based approaches [3].

Thresholding is a simple method that determines an intensity value, which is the threshold value, and then separates pixels, which have upper, between or lower threshold values. But thresholding is not useful for organ segmentation because of several reasons such as adjacent organs with similar intensity, intensity in homogeneity, noise or.

Clustering method uses an iterative algorithm. Parameters of a cost function of the current partition are computed and then used to minimize this function. K-means [2] is the commonly known clustering method. In the K-means method, mean intensities are calculated iteratively for each class and then the image is segmented by classifying pixels into the classes according to the closest mean values [3]. Initial centers and the number of classes are given as prior information to the algorithm.

In atlas based segmentation approaches, an atlas is used as the reference and one-to-one transformation is applied to do a mapping with a pre-segmented atlas image and the image that we want to find the required segment [4].

Several model based segmentation methods have been proposed. The assumption of the model based approaches is that the structures of organs have a repetitive form of geometry. Therefore, a probabilistic model is searched to explain the variation of the shape of the organ and this model is used as prior for segmentation of the image. This process involves three steps. One of them is registration of the training images to a common type [4]. Another step in this method is probabilistic representation for variations of the registered example images. The last step is statistical inference by using the model and the image.

ANNs have also been used for image segmentation. ANNs simulate biological learning by applying elementary computations. Learning process is performed by the adaptation of weights that are assigned to the connections between these elements. ANNs can be used both for classification and clustering [5]. In the case of classification, training data is used to determine the weights of ANN at first. Next, new data is segmented with adjusted weights. Spatial information can be incorporated into its classification procedures by using several interconnections in a neural network. In the case of clustering, ANNs are used in an unsupervised approach as a clustering technique and also for deformable models [6].

PDE based segmentation methods are based on active contours. The central idea for image segmentation using PDE is evolution of an initial contour towards the lowest potential of an identified cost function so that the definition of the cost function reflects the process to be addressed [6]. Curve

propagation can be performed by parametric or geometric active contour (i.e. level set) methods.

The most important difficulty in MR image segmentation is the intensity in-homogeneity [7]. Therefore, the segmentation methods that are based on the intensity values by assuming that intensities in a tissue class are constant over an image are not successful. The difficulties for organ segmentation from CT images are generally due to the artifacts [8]. The artifacts are caused by low attenuation adjacent to bones or patient motion such as beam-hardening artifacts or streak artifacts [9]. Ultrasound imaging is usually used for motion estimation instead of image segmentation due to the real-time acquisition. High levels of speckles cause limited segmentation from ultrasound images [10].

### B. Clustering

Clustering is a technique that is used to partition elements in a data set such that similar elements are assigned to same cluster while elements with different properties are assigned to different clusters. The purpose of clustering is to perform efficient search of elements in a data set. Clustering is mainly effective in multi-dimensional data that may be otherwise difficult to organize in an effective manner. One of the earliest clustering techniques in the literature is the K-means clustering method [11]. In this, clustering is based on the identification of K elements in the data set that can be used to create an initial representation of clusters. These K elements forming the cluster seeds. Left over elements of the data set is then assigned to one of these clusters. Even though the method seems to be simple, it bears from the fact that it may not be easy to clearly identify the initial K elements, or the seeds for the clusters [12]. This shortcoming led the researchers to look into alternative methods that provide an improvement over K means. Some of these techniques include genetic algorithm based clustering and fuzzy clustering [13].

In this paper, we have presented a technique that allows the partitioning of a given data set without having to depend on the initial identification of elements to represent clusters. Proposed technique is based on rearranging the clusters to better represent the partitions when new elements are added.

## II. ALGORITHM FOR ADAPTIVE K MEAN CLUSTERING

The adaptive K means clustering algorithm starts with the selection of K elements from the input data set. These K elements form the seeds of clusters and are randomly selected. The properties of each element also form the properties of the cluster that is constituted by the element. The algorithm is based on the ability to compute distance between a given element and cluster. This function is also applied to compute distance between two elements. An important consideration for this function is that it should be able to account for the distance based on properties that have been normalized so that the distance is not dominated by one property or some property is not ignored while distance computation [14]. In most cases, the Euclidean distance may be sufficient. For example, in the case of spectral data given by 'n' dimensions, the distance between two data elements  $E_1$  and  $E_2$ , is equal to

$$\sqrt{(E_{11} - E_{12})^2 + (E_{12} - E_{22})^2 + (E_{1n} - E_{2n})^2},$$

Where  $E_1 = \{E_{11}, E_{12}, \dots, E_{1n}\}$  and  $E_2 = \{E_{21}, E_{22}, \dots, E_{2n}\}$

It should be noted that due to better performance, the square root function may be neglected. In other cases, there is need to modify the distance function. Such cases can be shown by data in which one dimension is scaled different compared to other dimensions or the properties may be required to have different weights during comparison. With the distance function, the algorithm is as follows:

Compute the distance of each cluster from every other cluster and stored as a triangular matrix in a 2D array. Minimum distance  $D_{\min}$  between any two clusters  $C_{m1}$  and  $C_{m2}$  and also identification of these two closest clusters is noted down.

For each non clustered element  $E_i$ , distance of  $E_i$  from every cluster is to be computed. For assignment of this element to a cluster, there may be following three cases:

1. If the distance element and the cluster is 0, assign the element to that cluster, and begin working with the next element.
2. If the distance between the element and cluster is less than the distance  $d_{\min}$ , assign this element to its nearest cluster. As a result of this assignment, the cluster representation, or centroid, may change. The centroid is recomputed as an average of properties of all elements in the cluster. In addition, the distance of the affected cluster from every other cluster, as well as the minimum distance between any two clusters and the two clusters that are closest to each other is recomputed.
3. When the distance  $D_{\min}$  is less than the distance of the element from the closest cluster, select the two closest clusters  $C_{m1}$  and  $C_{m2}$ , and merge  $C_{m2}$  into  $C_{m1}$ . Demolish the cluster  $C_{m2}$  by removing all the elements from the cluster and by deleting its representation and add the new element into this empty cluster, successfully creating a new cluster. The distances between all clusters are recomputed and the two closest clusters identified over again.

Table 1. Performance Analysis for Quality Measures Using K means and Adaptive K-Means clustering in segmentation of Medical images

Segmentation Method	Input image	MSE	PSNR	Elapsed Time In Sec
K Means	im_15.bmp	1.5324	46.2771	5.48
Adaptive K Means		2.1993	67.0620	0.295
K Means	im_18.bmp	7.3430	39.4720	8.25
Adaptive K Means		6.7147	59.7908	0.243
K Means	im_17.bmp	9.8092	38.2144	5.44
Adaptive K Means		14.1326	54.9429	0.295

### III. RESULT



Fig.2(b).K mean Segmented Image



Fig.2(a). Input Image

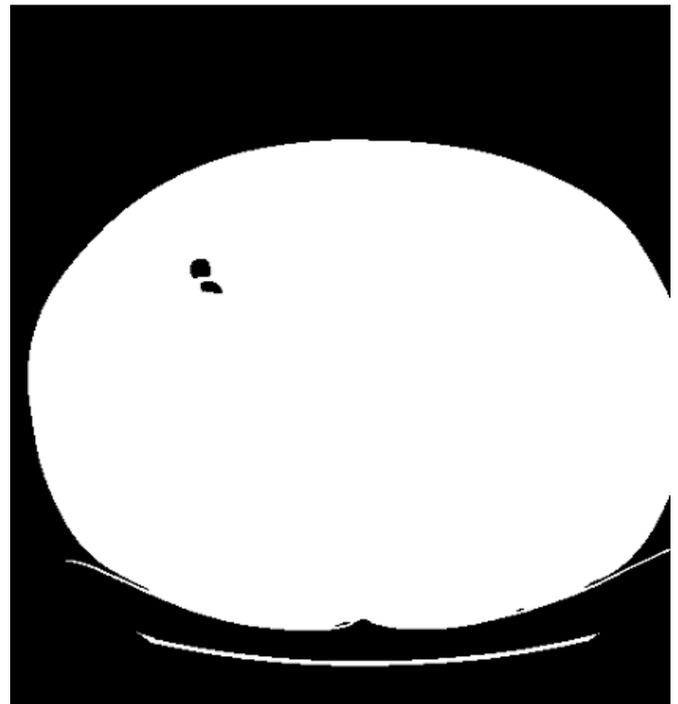


Fig.2(c). Adaptive cluster segmented Image

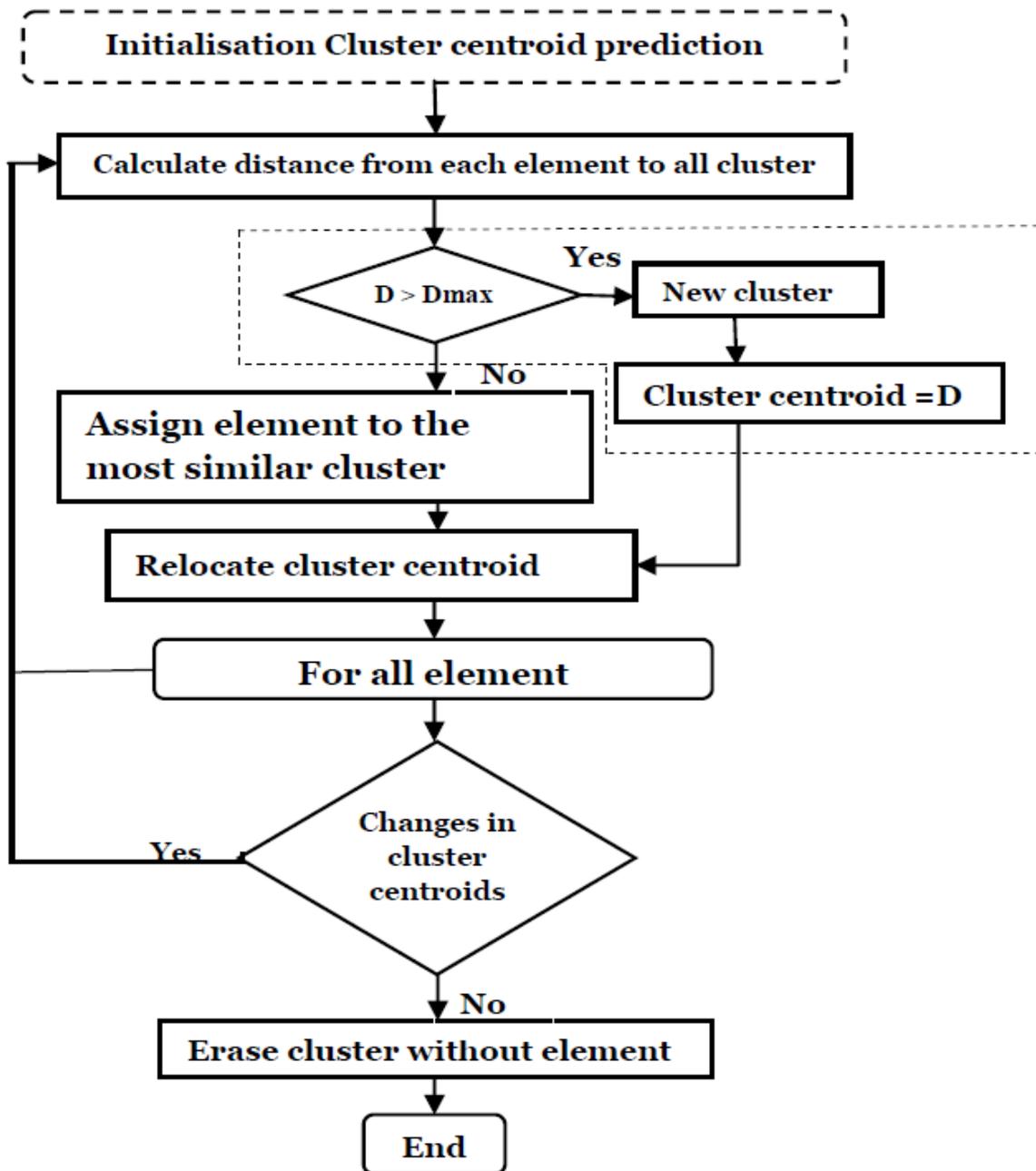


Fig.3.Flow Chart Adaptive Clustering Algorithm

#### IV. CONCLUSION:

An adaptive algorithm to perform K-means clustering, and is not dependent on the selection of K seeds to initialize the clusters is presented in this paper. The algorithm is tested for several images and found to perform fine, ensuing in good data partitioning and remarkable accelerated search in the resulting data structures. Based on the PSNR values It can be concluded that the adaptive k means performs better and speed more than K means.

#### REFERENCES

- [1] Levinski, K., Sourin, A., Zagorodnov, V., 2009. Interactive surface-guided segmentation of brain MRI data, *Computers in Biology and Medicine*, 39, 1153–1160.
- [2] Coleman, G. B., Andrews, H. C., 1979. Image segmentation by clustering. *Proceedings of IEEE*, 5, 773–785.
- [3] Jian, B., Vemuri, B. C., 2011. Robust point set registration using Gaussian mixture models. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 33 (8), 1633 – 1645.
- [4] Maintz, J. B. A., Viergever, M. A., 1998. A survey of medical image registration. *Medical Image Analysis*, 2, 1–36.

- [5] Hall, L. O., Bensaïd, A. M., Clarke, L. P., Velthuizen, R. P., Silbiger, M. S., Bezdek, J. C., 1992. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *IEEE Transactions on Neural Networks*, 3, 672–682.
- [6] Vilarino, D. L., Brea, V. M., Cabello, D., Pardo, J. M., 1998. Discrete-time CNN for image segmentation by active contours. *Pattern Recognition Letters*, 19, 721–734.
- [7] Condon, B. R., Patterson, J., Wyper, D., Jenkins, A., Hadley, D. M., 1989. Image non-uniformity in magnetic resonance imaging: its magnitude and methods for its correction. *The British Journal of Radiology*, 60, 83–87.
- [8] Simmons, A., Tofts, P. S., Barker, G. J., Arridge, S. R., 1994. Sources of intensity nonuniformity in spin echo images at 1.5T. *Magnetic Resonance in Medicine*, 32, 121–128.
- [9] Slone, R. M., Fisher, A. J., Pickhardt, P. J., Gutierrez, F. R., Balfé, D. M., 2000. *Body CT: A practical approach*. New York: McGraw-Hill.
- [10] Mojsilovic, A., Popovic, M. V., Neskovic, A. N., Popovic, A. D., 1997. Wavelet image extension for analysis and classification of infarcted myocardial tissue. *IEEE Transactions on Biomedical Engineering*, 44, 856–866.
- [11] Anderberg, M. R., 1973. *Cluster Analysis for Applications*. Academic Press, New York, NY.
- [12] Bandyopadhyay, S., and Maulik, U. 2002. Generic clustering for automatic evolution of clusters and applications to image classification. *Pattern Recognition* 35:1197–1208.
- [13] Belacel, N.; Hansen, P.; and Mladenovic, N. 2002. Fuzzy J-means: A new heuristic for fuzzy clustering. *Pattern Recognition* 35:2193–2200.
- [14] Bandyopadhyay S and Maulik U. Generic clustering for automatic evolution of clusters and applications to image classification, *Pattern Recognition* 35:1197–1208, 2002.
- [15] Yu C.T, Wang Y.T, and Chen C. H., Adaptive document clustering. In *Proceedings of the Eighth ACM SIGIR*, 197–203, 1985.