COMPRESSION AND SEGMENTATION OF MEDICAL IMAGES USING K-MEAN CLUSTERING ALGORITHM

[1]P. Hariharan,[2] V.Divya
[1]Assistant Professor,[2] Research Scholar,
[1, 2] PG & Reaserch Department of Computer Science & Application,
[1, 2] Adhiparasakthi College of Arts and Science, G.B.Nagar, Kalavai, 632506, Vellore District, Tamilnadu, India.

Abstract : In computer vision, image compression mainly refers to the problem of reducing the memory size of a digital image. Mainly problems regarding data transferring over internet might require an image data to be of comparatively lesser memory size. Moreover high bandwidth is also required for transmission of high quality image data. Image compression provides solution to this problem. Now in order to retrieve the image of lesser size there is a considerable amount of loss of image data. For this problem optimized method is chosen to provide a final image which is comparatively of smaller memory size than the original image, yet it is quite visually similar to the original image. In this paper, we are trying to segment the original image using K-Means clustering method. To overcome the noise sensitiveness of conventional clustering algorithm, a novel FCM algorithm for image segmentation is presented in this paper. The algorithm is developed by modifying the objective function of the standard FCM algorithm. Experimental results on segmentation of synthetic and real images demonstrate that the proposed algorithm is effective and robust

***IndexTerms* - Fuzzy C means, Clustering, K-means, Compression, Segmentation**

I. INTRODUCTION

The objective of image compression is to reduce irrelevance and redundancy of the image data in order to be able to store or transmit data in an efficient form. Generally, compression technologies can be classed into lossless and lossy. Lossless compression allows the exact original images to be reconstructed from the compressed data. Lossless compression technologies are used in cases where it is important that the original and the decompressed data are identical. Avoiding distortion limits their compression efficiency. When used in image compression where slight distortion is acceptable, lossless compression technologies are often employed in the output coefficients of lossy compression. Lossy compression technologies usually transform an image into another domain, quantize and encode its coefficients. During the last three decades, transform-based image compression technologies have been extensively researched and some standards have appeared. Two most common options of transformation are the Discrete Cosine Transform (DCT) [1] and the Discrete Wavelet Transform (DWT) [2].

The DCT-based encoder can be thought of as compression of a stream of 8 × 8 small block of images. This transform has been adopted in JPEG [3]. The JPEG compression scheme has many advantages such as simplicity, universality and availability. However, it has a bad performance at low bit-rates mainly because of the underlying block-based DCT scheme. For this reason, as early as 1995, the JPEG-committee began to develop a new wavelet-based compression standard for still images, namely JPEG 2000 [4], [5]. The DWT-based algorithms include three steps: a DWT computation of the normalized image, quantization of the DWT coefficients and lossless coding of the quantized coefficients. The detail can be found in [6] and [7]. Compared with JPEG, JPEG 2000 provides many features that support scalable and interactive access to large-sized image. It also allows extraction of different resolutions, pixel fidelities, regions of interest, components and etc. There are several other DWT-based algorithms, such as Set Partitioning in Hierarchical Trees (SPIHT) Algorithm [8].

Image segmentation is an important and challenging problem and a necessary first step in image analysis as well as in high-level image interpretation and understanding such as robot vision, object recognition, and medical imaging. The goal of image segmentation is to partition an image into a set of disjoint regions with uniform and homogeneous attributes such as intensity, colour, tone or texture, etc. Many different segmentation techniques have been developed and detailed surveys can be found in references [9– 11]. According to reference [9], the image segmentation approaches can be divided into four categories: thresholding, clustering, edge detection and region extraction. In this paper, a clustering based method for image segmentation will be considered.

II. RELETAD WORKS

Clustering and classification are both fundamental tasks in Data Mining. Classification is used mostly as a supervised learning method, clustering for unsupervised learning (some clustering models are for both). The goal of clustering is descriptive, that of classification is predictive (Veyssieres and Plant, 1998). Since the goal of clustering is to discover a new set of categories, the new groups are of interest in themselves, and their assessment is intrinsic. Since clustering is the grouping of similar instances/objects, some sort of measure that can determine whether two objects are similar or dissimilar is required. There are two main type of measures used to estimate this relation: distance measures and similarity measures. The main reason for having many clustering methods is the fact that the notion of “ cluster” is not precisely defined (Estivill-Castro, 2000). Consequently many clustering methods have been developed, each of which uses a different induction principle. Farley and Raftery (1998) suggest dividing the clustering methods into two main groups: hierarchical and partitioning methods. Han and Kamber (2001) suggest categorizing the methods into additional three main categories: *density-based methods*, *model-based clustering* and *gridbased methods*.

Traditional clustering approaches generate partitions; in a partition, each instance belongs to one and only one cluster. Hence, the clusters in a hard clustering are disjointed. Fuzzy clustering (seefor instance (Hoppner, 2005)) extends this notion and suggests a *soft clustering* schema. In this case, each pattern is associated with every cluster using some sort of membership function, namely, each cluster is a fuzzy set of all the patterns. Larger membership values indicate higher confidence in the assignment of the pattern to the cluster. A hard clustering can be obtained from a fuzzy partition by using a threshold of the membership value.

III. PROPOSED METHODOLOGY

The first step is used compress the large size image into small size image by using the K-Means Clustering algorithm. Then the compressed medical image is segmented using the Fuzzy c means clustering. The Fuzzy Segmented image is compared with OTSU method The Results shows Fuzzy Segmentation is better than other Conventional Methods. The K means And Fuzzy clustering algorithm is explained in below.

K-MEANS CLUSTERING

K-Means clustering intends to partition *n* objects into *k* clusters in which each object belongs to the cluster with the nearest mean. This method produces exactly *k* different clusters of greatest possible distinction. The best number of clusters *k* leading to the greatest separation (distance) is not known as a priori and must be computed from the data. The objective of K-Means clustering is to minimize total intra-cluster variance, or, the squared error function:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\| x_i - v_j \| \right)^2$$

Where,

'$||x_i - v_j||$' is the Euclidean distance between x_i and v_j.

'c_i' is the number of data points in i^{th} cluster.

'c' is the number of cluster centers.

Algorithmic steps for k-means clustering

Let $X = \{x_1,x_2,x_3,\ldots,x_n\}$ be the set of data points and $V = \{v_1,v_2,\ldots,v_c\}$ be the set of centers.

1) Randomly select 'c' cluster centers.
2) Calculate the distance between each data point and cluster centers.
3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..
4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, 'c_i' represents the number of data points in i^{th} cluster.

5) Recalculate the distance between each data point and new obtained cluster centers.
6) If no data point was reassigned then stop, otherwise repeat from step 3).

Advantages

1) Fast, robust and easier to understand.
2) Relatively efficient: O(tknd), where n is # objects, k is # clusters, d is # dimension of each object, and t is # iterations. Normally, k, t, d << n.
3) Gives best result when data set are distinct or well separated from each other.

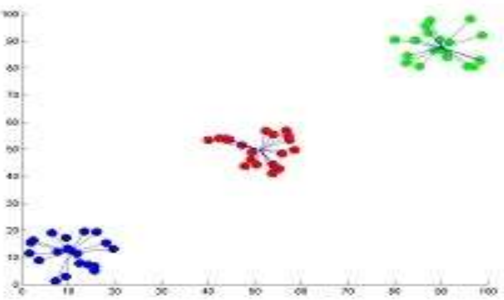

Fig I: Showing the result of k-means for *'N'* = 60 and *'c'* = 3

FUZZY C-MEANS CLUSTERING ALGORITHM

This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated according to the formula:

$$\mu_{ij} = 1 / \sum_{k=1}^{c} (d_{ij} / d_{ik})^{(2/m-1)}$$

$$v_j = (\sum_{i=1}^{n} (\mu_{ij})^m x_i) / (\sum_{i=1}^{n} (\mu_{ij})^m), \forall j = 1, 2, \ldots c$$

where,

n' is the number of data points.

'vj' represents the j^{th} cluster center.

'm' is the fuzziness index m € [1, ∞].

'c' represents the number of cluster center.

'µij' represents the membership of i^{th} data to j^{th} cluster center.

'dij' represents the Euclidean distance between i^{th} data and j^{th} cluster center.

Main objective of fuzzy c-means algorithm is to minimize:

$$J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{ij})^m \left\| x_i - v_j \right\|^2$$

where,

'$||x_i - v_j||$' is the Euclidean distance between i^{th} data and j^{th} cluster center.

Algorithmic steps for Fuzzy c-means clustering

Let $X = \{x_1, x_2, x_3 \ldots, x_n\}$ be the set of data points and $V = \{v_1, v_2, v_3 \ldots, v_c\}$ be the set of centers.

1) Randomly select ' *c*' cluster centers.

2) Calculate the fuzzy membership 'μ_{ij}' using:

$$\mu_{ij} = 1 / \sum_{k=1}^{c} (d_{ij} / d_{ik})^{(2/m-1)}$$

3) Compute the fuzzy centers 'v_j' using:

$$v_j = (\sum_{i=1}^{n} (\mu_{ij})^m x_i) / (\sum_{i=1}^{n} (\mu_{ij})^m), \forall j = 1, 2, \ldots c$$

4) Repeat step 2) and 3) until the minimum *'J'* value is achieved or $||U^{(k+1)} - U^{(k)}|| < \beta$.

where,

'k' is the iteration step.

β' is the termination criterion between [0, 1].

$U = (\mu_{ij})_{n*c}$' is the fuzzy membership matrix.

'J' is the objective function.

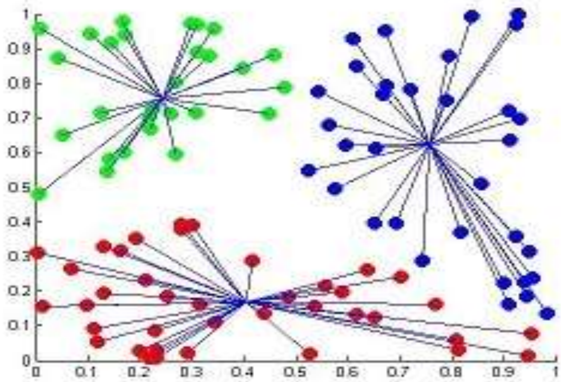

Fig III: Result of Fuzzy c-means clustering

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 than one cluster center.

IV. IMPLEMENTATION RESULTS

The Implementation can be done with the help of MATLAB with Fuzzy toolbox

The Figure 1 shows the Original Ultrasound B mode Image. The Figure 2 shows the compressed Image using K means Clustering .Figure 3 explains the image Segmentation OTSU Method. Figure 4 and 5 is image Segmentation using Fuzzy Methods. Figure 6 is Filtered image and Figure 7 is Artery Found in an Image.

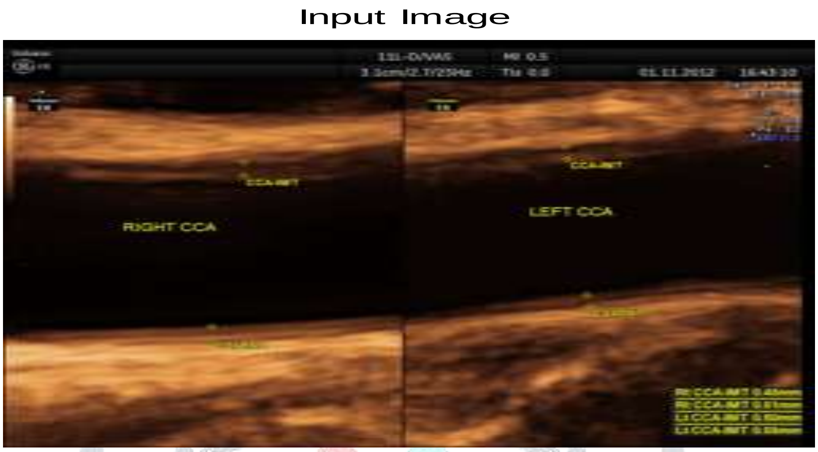

Fig 1: Original Ultrasound B mode Image

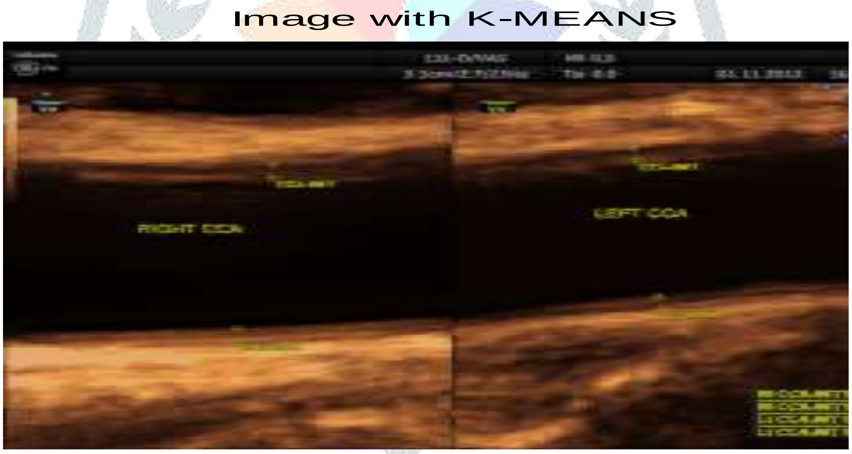

Fig 2: Compressed Image with Kmeans

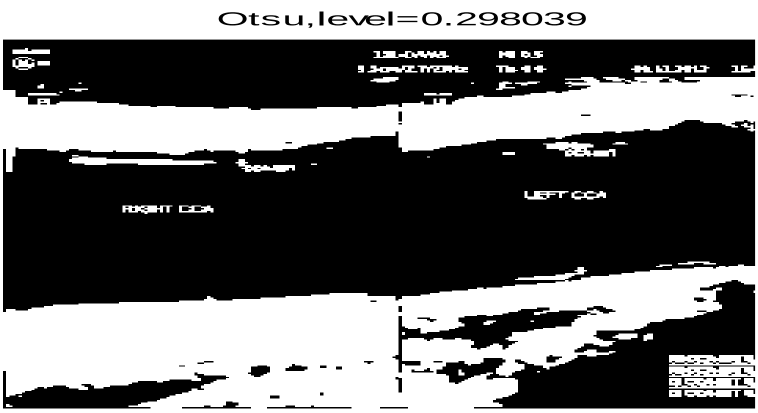

Fig 3: Otsu Image Segmentation

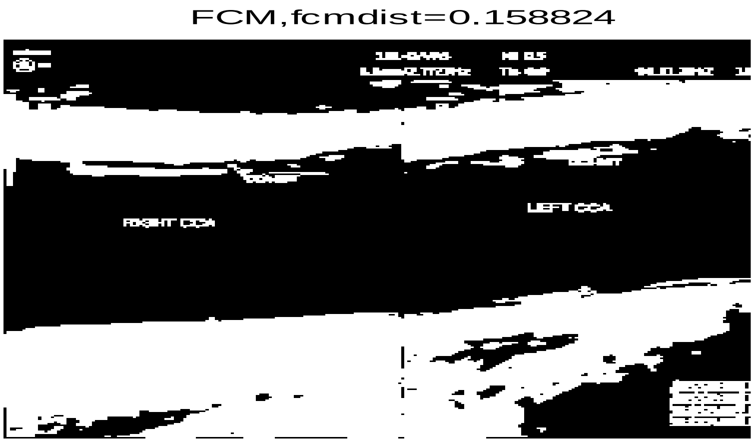

Fig 4: Fuzzy C means Clustering Based Image Segmentation-fcm distance

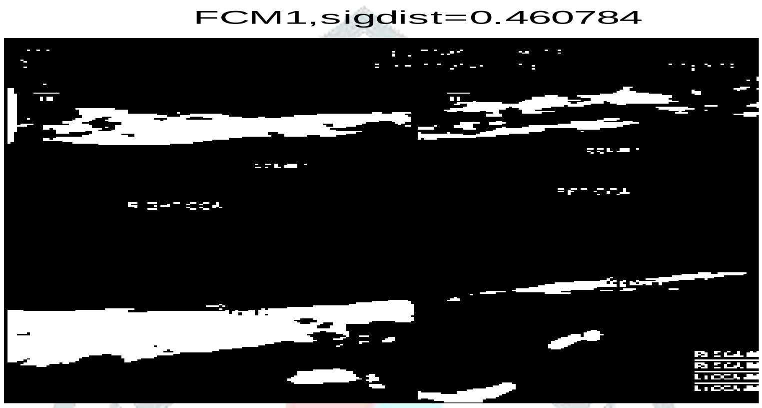

Fig 5: Fuzzy C means Clustering Based Image Segmentation-sigma distance

Fig 6: Filtered image

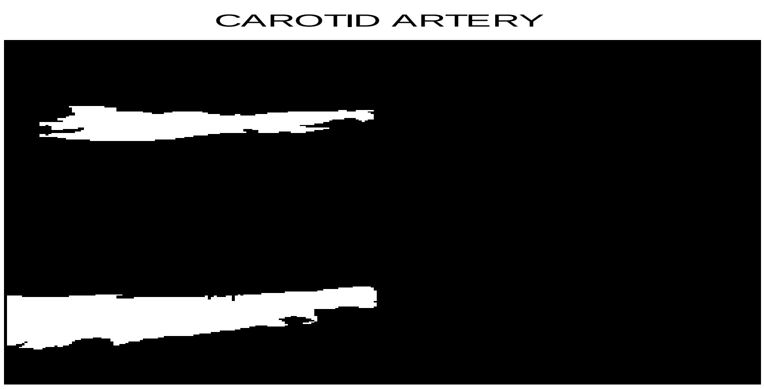

Fig 7: Carotid artery

CONCLUSION

K-means clustering is used for Image Compression. In this mat lab program, the feature vectors are simply the N X N non-overlapping blocks of pixels in the image. Like a scalar quantize, a vector quantize has a quantization levels called code vectors and the set of K such code vectors is called codebook of size K. K-means clustering is an iterative process in which the code vectors are refined every stage by computing the centroids of the input vectors which belong to the respective cluster. The image Segmentation done with Conventional FCM and the Results were analyzed. In Future Segmentation for noisy medical images with spatial probability, Novel Fuzzy C-Means Clustering (NFCM), Fuzzy Local Information C-Means (FLICM) Clustering Algorithm and Improved Spatial Fuzzy C-Means Clustering (ISFCM) algorithm can be used.

REFERENCES

[1] N. Ahmed, T. Natarajan, and K. R. Rao, " Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, no. 1, pp. 90– 93, Jan. 1974.

[2] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*. Upper Saddle River, NJ, USA: Prentice- Hall, 1998.

[3] W. Pennebaker and J. Mitchell, *JPEG—Still Image Compression Standard*. New York, NY, USA: Van Nostrand Reinhold, 1993.

[4] M. W. Marcellin, M. J. Gormish, A. Bilgin, and M. P. Boliek, " An overview of JPEG-2000," in *Proc. IEEE Data Compress. Conf.*, Mar. 2000, pp. 523– 541.

[5] A. Skodras, C. Christopoulos, and T. Ebrahimi, " The JPEG 2000 still image compression standard," *IEEE Signal Process. Mag.*, vol. 11, no. 5, pp. 36– 58, Sep. 2001.

[6] T. Hopper, C. Brislawn, and J. Bradley, " WSQ gray-scale fingerprint image compression specification," Federal Bureau of Investigation, Criminal Justice Information Services, Washington, DC, USA, Tech. Rep. IAFIS-IC-0110-V2, Feb. 1993.

[7] C. M. Brislawn, J. N. Bradley, R. J. Onyshczak, and T. Hopper, " FBI compression standard for digitized fingerprint images," *Proc. SPIE*, vol. 2847, pp. 344– 355, Aug. 1996.

[8] A. Said and W. A. Pearlman, " A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243– 250, Jun. 1996.

[9] Fu, S.K.—Mui, J.K.: A Survey on Image Segmentation. Pattern Recognition, Vol. 13, 1981, pp. 3– 16.

[10] Haralick, R.M.—Shapiro, L.G.: Image Segmentation Techniques. Comput. Vi- sion Graphics Image Process., Vol. 29, 1985, pp. 100– 132.

[11] Pal, N.—Pal, S.: A Review on Image Segmentation Techniques. Pattern Recognition, Vol. 26, 1993, pp. 1277– 1294.