IMPLEMENTATION OF HARD C-MEANS CLUSTERING ALGORITHM FOR MEDICAL IMAGE SEGMENTATION

[1]MANSUR ROZMIN, [2] PROF. CHHAYA SURATWALA, [3] PROF. VANDANA SHAH

[1] M.E.[Masters Of Engineering] Student, Department Of Electronics And Communication Engineering, Sarvajanik College Of Engineering & Technology, Gujarat Technological University, Surat, Gujarat, India
[2] Asst. Professor, Department Of Electronics And Communication Engineering, Sarvajanik College of Engineering Technology, Gujarat Technological University, Surat, Gujarat, India
[3] Asst. Professor, Department Of Electronics And Communication Engineering, Sarvajanik College Of Engineering & Technology, Gujarat Technological University, Surat, Gujarat, India

mansurrozmin@gmail.com , chhaya.suratwala@scet.ac.in , vandana.shah@scet.ac.in

ABSTRACT: *Now a day the technology took the paradigm shift in the area of medical field. It is now easy to diagnose the biological problems through different medical devices and software's like CT Scan, MRI, X-rays etc. and imageJ, MevisLab respectively. This paper highlights the application of brain tumour segmentation detection using Hard C-Means (K-Means) clustering algorithm. Hard C means clustering is an unsupervised classification. Clustering is partitioning of a data set in a set of similar subsets. Each object in dataset shares some common property- often proximity according to some defined distance measure. Among various types of clustering techniques, Hard C-means is one of the most popular algorithms. The objective of Hard C-means algorithm is to make the distances of objects in the same cluster as small as possible. In this paper hard C-means clustering algorithm is applied on different medical image of brain tumour to detect tumour. Also comparison of our results with the results which are obtained by using different software's like imageJ, MevisLab is shown in last section.*

Keywords: - Clustering, Hard C-means Algorithm, Medical Image, Image Segmentation

1. INTRODUCTION

Clustering requirements arise in many different applications, such as data mining, knowledge discovery, data compression, vector quantization, pattern reorganization and pattern classification in Biomedical, Bioinformatics [8]. There are various ways to cluster data depending on data set available. The more information on clustering and clustering algorithm can be available in [1].

Fig 1. Group of different colored balls

Fig 2. Group of clustered balls

Let say there are yellow, blue & Orange colored ten balls in room. For having database about balls like how many number of different color balls exist, grouping same color ball together, total number of balls, etc. These sort of data base can be prepared using clustering. Here ball of same color are grouped together to form one cluster. Thus, we see clustering means grouping of data or dividing a large data set into smaller data sets of some similarity in the figure 1 and 2. Clustering is a type of categorization imposed rules on a group of data points or objects. A broad definition of clustering could be "the process of categorizing a finite number of data points into groups where all members in the group are similar in some manner". As a result, a cluster is a aggregation of objects. All data points in the same cluster have common properties (e.g. distance) which are different to the data points lying in other clusters.

The most well known methods for clustering is Hard C-means that is Hard C- means developed by Mac Queen in 1967. The simplicity of Hard C-means made this algorithm used in various fields. Hard C-means is a partition clustering method that separates

data into k mutually excessive groups. By iterative such partitioning [6], Hard C-means minimizes the sum of distance from each data to its clusters. Hard C-means method is very popular because of its ability to cluster huge data, and also outliers, quickly and efficiently.

However, Hard C-means algorithm is very sensitive in initial starting points. Hard C-means generates initial cluster randomly. When random initial starting points close to the final solution, Hard C-means has high possibility to find out the cluster centre, Otherwise, it will lead to incorrect clustering result.

There are many approaches to image segmentation have been proposed [3] from these various methods, clustering is one of the simplest and widely used in segmentation of grey level images. There are different Techniques such as Hard C-means [4], ISODATA [4], and fuzzy c-means [5] have been developed but their application to color images has been limited. The Hard C-means and fuzzy c-means algorithms require the number of clusters to be known in advanced and the ISODATA algorithm has six parameters which must be supplied as input to the algorithm. In order to supply the information required by algorithms as input parameter, the user must have some knowledge about the image.

2. STANDARD HARD C-MEANS ALGORITHM

The Hard C-means algorithm [7] is effective in producing clusters for many practical applications. But the computational complexity of the standard Hard C-means algorithm is very high for large data sets. Moreover, this algorithm results in different types of clusters depending on the random choice of initial centroids. In K means objects are classified as belonging to one of k groups exclusively, k is chosen priori. Cluster membership is determined by calculating the centroids for each group during each iteration and updated and assigning each object to the group with the closest centroids. So that these iteration re-allocation minimizes the overall within cluster dispersion of cluster members [2].

The reason behind choosing the k means algorithm is its popularity for the following reasons:

1) Its time complexity is *O(nkl)*, where n is the number of patterns, k is the number of clusters, l is the number of iteration taken by the algorithm to converge.
2) Its space complexity is *O(k+n)*. it requires additional space to store the data matrix.
3) It is order independent; for a given initial seed set of cluster centres, it generates the same partition of the data irrespective of the order in which the patterns are presented to the algorithm.

In medical field, There are different types of algorithm were developed for brain tumour detection. But these algorithms may have some drawback in detection and extraction[9]. Tumour is due to the uncontrolled growth of the tissues in any part of the body. There are two types of tumour, it may be primary or secondary. If it is an origin, then it is known as primary level tumour and If the part of the tumour is spread to another place in the body and grown as its own then it is known as secondary. Brain tumour mainly affects CSF (Cerebral Spinal Fluid) and causes for strokes. The physician gives the treatment for the strokes rather than the treatment for tumour. But the detection of tumour is important for that treatment. The lifetime of the person who affected by the brain tumour will increase by couple of years if it is detected at current stage. Normally tumour cells are of two types, Mass and Malignant. The detection of the malignant tumour is somewhat difficult then mass tumour.

3. EXISTING METHOD

The existing method for brain tumour detection is based on the thresholding and region growing. The thresholding method was ignored the spatial characteristics but spatial characteristics are important for the malignant tumour detection. In the thresholding based segmentation the image is considered as having only two values either black or white. But the bit map image contains 0 to 255 gray scale values. So sometimes it ignores the tumour cells also. In case of the region growing based segmentation it needs more user interaction for the selection of the seed. Seed is nothing but the centre of the tumour cells; it may cause intensity in homogeneity problem. And also it will not provide the acceptable result for all the images. The typical output for the thresholding is given below.

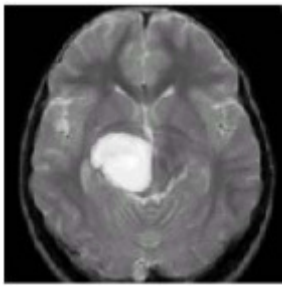

Fig.3 input image for thresholding

Fig.4 output Image for thresholding

4. HARD C-MEANS ALGORITHM AND FLOWCHART

Hard C-means is the one of the unsupervised learning algorithm for clusters. Clustering the image is the process of grouping the pixels according to the some characteristics such as pixel intensity, distance criterion etc. In the Hard C-means Algorithm algorithm [9] initially the number of clusters k is been already defined. Then k-cluster centre or centroids are chosen randomly. Then based on the distance between the each pixel to each cluster centres are calculated. The distance measurement may be of simple Euclidean distance function. Single pixel is compared to all cluster centres using the distance formula. The pixel is moved to particular cluster which has shortest distance among all. Then the centroid is re-estimated. Again each pixel is compared to all centroids.

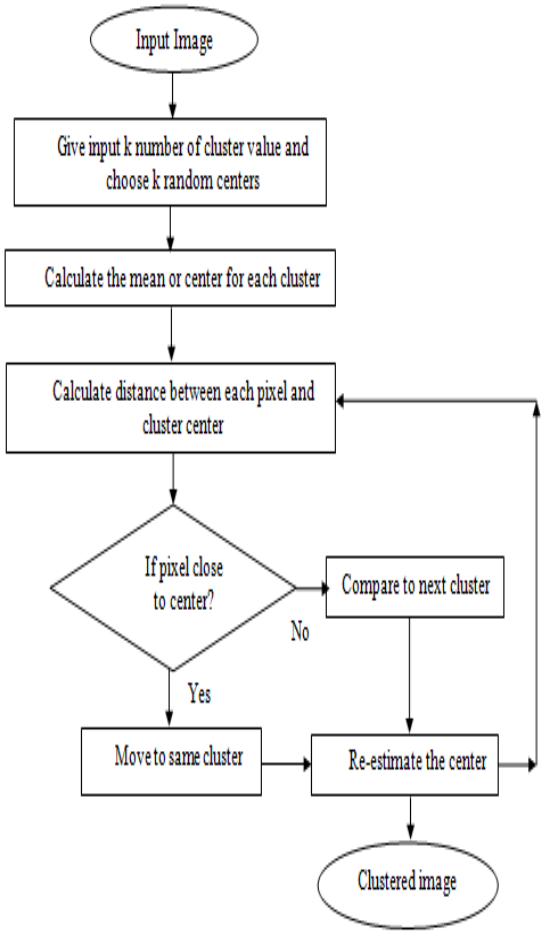

Fig 5. Flowchart of hard C-means algorithm

The process continuous until the centre converges.K-mean distributes all objects to K number of clusters at random. The Hard C-means method aims to minimize the sum of squared distances between all points and the cluster centre.

This procedure consists of the following algorithmic steps [9]:

1. Give the no of cluster value as k.
2. Randomly choose the k cluster centres
3. Calculate mean or centre of the cluster
4. Calculate the distance between each pixel to each cluster centre
5. If the distance is near to the centre then move to that cluster.
6. Otherwise move to next cluster.
7. Re-estimate the centre.
8. Repeat the process until the centre doesn't move

The Hard C-means algorithm is the most extensively studied clustering algorithm and is generally effective in producing good results. The major drawback of this algorithm is that it produces different clusters for different sets of values of the initial centroids. Quality of the final clusters heavily depends on the selection of the initial centroids. The Hard C-means algorithm is computationally expensive and requires time proportional to the product of the number of data items, number of clusters and the number of iterations [10].

The require flowchart for our implemented hard c-means algorithm is shown in figure. Medical image is given as input and after applying hard c-means algorithm and as output clustered image which shows the detected brain tumour location.

5. SIMULATION RESULTS

These are the simulation results obtained for hard C-means algorithm using MATLAB for five cluster operation and different other software comparison like *imagej* and *MeVisLab*.

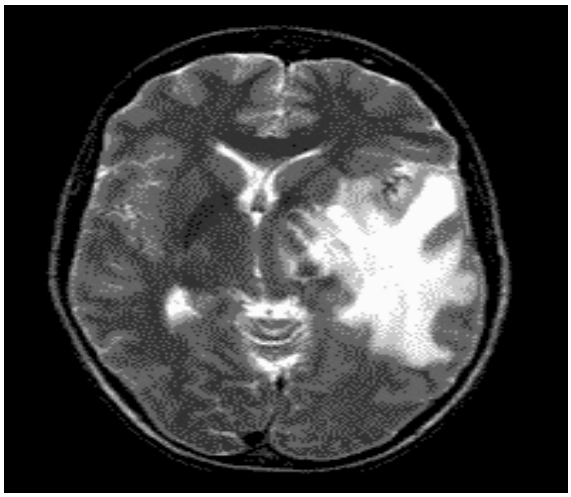

Fig 6. Original image

Comparison of results which are obtained by using different softwares like *imageJ*, *MevisLab* for hard C-means clustering algorithm for five cluster operation is shown in table.

Fig 7. MATLAB output

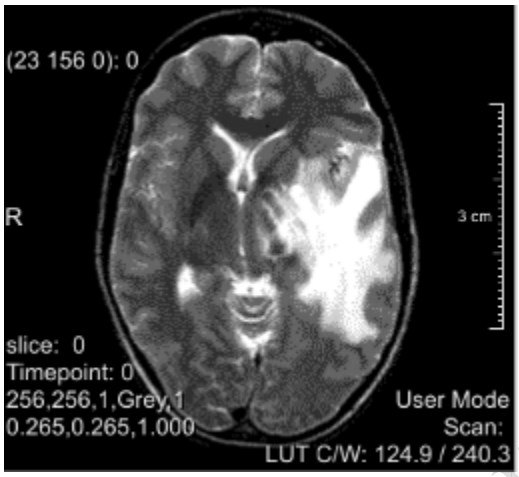

Fig 8. *Mevislab2*.1vc8 output

Fig 9. *Imagej* output

By observing this results we can comment that result obtained by our hard C-means algorithm implementation brain tumour detection gives better result compare to other result obtained by softwares. Here obtained value of results are approximate values.

Table 1.Observed Result

Observed parameter of brain tumour detection	Using MATLAB	Using *imagej*	Using *MeVisLab*
Resolution	>80%	70%	70%
Area detection	>80%	50%	90%
Area calculation	>80%	0%	80%

6. CONCLUSION

Hard C-means is an important clustering algorithm in the field of pattern recognition and data mining. There are different types of tumours are available. They may be as mass in brain or malignant over the brain. Suppose if it is a mass then K- means algorithm is enough to extract it from the brain cells. The noise free image is given as a input to the Hard C-means and tumour is extracted from the MRI image. For large cluster number, there are still other issues remaining for the design of Hard C-means hardware architectures.

For future developments, modules that can handle high-dimensional vectors might be integrated with the Hard C-means architecture to construct a complete clustering engine for multimedia content analysis. For hardware implementation FPGA device can be used for such complex algorithm which will reduce the amount requirement for power consumption. It has also advantage of high speed with high data computation with accurate area detection density.

7. REFRENCES

[1] Tapas kanungo, david m. mount, Nathan s. Netanyahu, christen d. piatko, ruth silverman, angela y. wu, "An Efficient Hard C-means Clustering Algo. Analysis And Implementation", IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 24, No. 7, 881-892, July 2002.

[2] Osama Abu Abbas, "Comparison Between Data Clustering Algorithms", The International Arab Journal Of Information Technology, Vol. 5, No. 3, July 2008.

[3] A. Perez and R.C. Gonzalez, "An iterative thresholding algorithm for image segmentation", IEEE Trans. Pattern Anal. Machine Intell., vol. 9, pp. 742-751, 1987.

[4] J.T. Tou and R.C. Gonzalez, "Pattern Recognition Principles", Massachusetts: Addison-Wesley , 1974.

[5] M.M. Trivedi and J.C. Bezdek, Low-level segmentation of aerial images with fuzzy clustering, IEEE Trans. Syst. Man. Cybern., vol. SMC-16, pp. 589-598,1986.

[6] Kohei Arai, Ali Ridho Barakbah, "Hierarchical Hard C-means: an algorithm for centroids initialization for k-means", International Journal of Intelligent Systems, Vol. 36, No.1, 2007.

[7] Stuart P. Lloyd, "Least squares quantization in pcm", IEEE Transactions on Information Theory, Vol.-28, No. 2, 129-136, March 1987.

[8] Daxin Jiang, Chum Tong and Aidong Zhang, "Cluster Analysis for Gene Expression Data", IEEE Transactions on Data and Knowledge Engineering, 16(11): 1370-1386, 2004.

[9] J.selvakumar, A.Lakshmi, T.Arivoli, "Brain Tumour Segmentation and Its Area Calculation in Brain MR Images using K-Mean Clustering and Fuzzy C-Mean Algorithm", IEEE International Conference On Advances In Engineering, Science And Management March 30, 31, 2012.

[10] Ritu yadav, Anuradha Sharma, "Advanced Methods to Improve Performance of k-means Algorithm: A Review", Global Journal of Computer Science and Technology, Vol 12, No 9, 2012.