ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 1, January 2017

Modified Pillar K-Means Algorithm with Chessboard Distance & Accumulated Distance Metric for Centroid Designation to Detect Brain Tumors in MR Images

Hakeem Aejaz Aslam[1], Dr.T.Ramashri[2]

Research Scholar, Dept. of ECE in Rayalaseema University, Kurnool, A.P, India[1]

Professor, Dept. of ECE, Sri Venkateswara University College of Engineering, Tirupati, A.P, India[2]

ABSTRACT: This paper presents a new approach to image segmentation using Modified Pillar K-Means algorithm. This segmentation method includes a new mechanism for distance metric and grouping the elements of high resolution images in order to improve accuracy and reduce the computation time. The system uses modified Pillar K-means for optimized image segmentation algorithm. The Pillar algorithm considers the placement of pillars that should be located as far from each other to resist the pressure distribution of a roof same as the number of centroids between the data distribution. This algorithm is able to optimize the K-Means clustering for image segmentation in the aspects of accuracy and computation time. This algorithm distributes all initial centroids according to the Accumulated Distance Metric (ADM) & the distance is calculated by Chessboard distance measure.

This paper evaluates the proposed approach for image segmentation by comparing with K-Means, Fuzzy C-means, Pillar K-Means clustering algorithm with different sample MR Images for Precision, Sensitivity Factor, Efficiency Factor & Computation Time. Experimental results clarify the effectiveness of our approach to improve the segmentation quality and accuracy aspects of computing time.

KEYWORDS: Tumor, Segmentation, Detection, Centroids, Clustering and computing

I.INTRODUCTION

In image segmentation, one challenge is how to deal with the nonlinearity of real data distribution, which often makes segmentation methods need more human interactions and make unsatisfied segmentation results. Medical image segmentation plays an instrumental role in clinical diagnosis. An ideal medical image segmentation scheme should possess some preferred properties such as minimum user interaction, fast computation, and accurate and robust segmentation results.

Image segmentation is an image analysis process that aims at partitioning an image into several regions according to a homogeneity criterion. Image segmentation is a very complex task, which benefits from computer assistance, and yet no general algorithm exists. It has been a research field in computer science for more than 40 years now, and the early hope to find general algorithms that would achieve perfect segmentations independently from the type of input data has been replaced by the active development of a wide range of very specialized techniques. Most of the existing segmentation algorithms are highly specific to a certain type of data, and some research is pursued to develop generic frameworks integrating these techniques.

Segmentation can be a fully automatic process, but it achieves its best results with semi-automatic algorithms, i.e. algorithms that are guided by a human operator. This concept of semi-automatic process naturally involves an environment in which the human operator will interact with the algorithms and the data in order to produce optimal

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 1, January 2017

segmentations. The simplest example of the need of a human intervention during the task of segmentation results from the specificity of the existing algorithms. Depending on the type of input data, the operator will have to carefully pick the best adapted algorithm, which most of the time cannot be done in an automatic way. The subjective point of view of the human is required.

In this paper, we try to develop a segmentation algorithm for abnormal MR Images using Fuzzy C-means clustering technique and also to compare the results with the earlier techniques like Region growing method, K-Means clustering algorithms. Abnormal brain images from four different classes which are metastases, meningioma, glioma and astrocytoma are being used in this work.

II.LITERATURE SURVEY
SEGMENTATION ALGORITHMS

REGION GROWING

Region growing is a technique for extracting an image region that is connected based on some predefined criteria. These criteria can be based on intensity information and/or edges in the image. In its simplest form, region growing requires a seed point that is manually selected by an operator and extracts all pixels connected to the initial seed based on some predefined criteria. For example, one possible criterion might be to grow the region until an edge in the image is met. Like thresholding, region growing is seldom used alone but usually within a set of image-processing operations, particularly for the delineation of small, simple structures such as tumors and lesions.

This approach to segmentation examines neighboring pixels of initial “seed points” and determines whether the pixel neighbors should be added to the region. The process is iterated on, in the same manner as general data clustering algorithms. A general discussion of the region growing algorithm is described below:

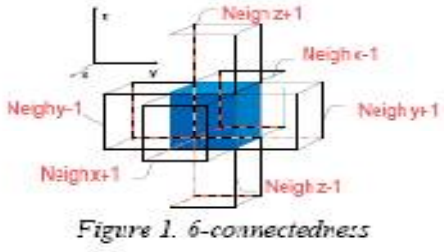


Figure 1. 6-connectedness

Figure 1: Connected region growing

Growth Algorithm

An auxiliary FIFO (First In First Out) structure is used when the seeds are first located, and where neighbors belonging to the area to visit are spooled. In this, we can see the growth of Voxelp seudo-code algorithm in detail. The algorithm takes successively the elements of the queue. Each of these elements is avoxel volume already accepted. For each of them should go to their neighbors and decide if it belongs to the neighboring region in accordance withthe selection criteria. To compare the6-neighbor connection is used.

One of the strengths of this technique is that it always grows by neighbors, thus maintaining connectivity between elements that are included in the segmented region.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 1, January 2017

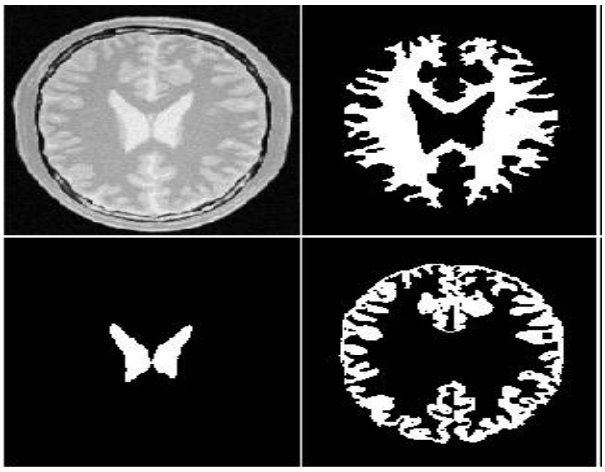


Figure 2: Region growing method with different initial seeds. Top-left: Original Image, Top-Right: the white matter, bottom–Left: a ventricle bottom right: The grey matter.

HISTOGRAM BASED METHODS:

Histogram-based methods are very efficient in comparison with other methods of image segmentation, and which generally require a single pass through the pixels. In this technique, a histogram is calculated from the set of pixels of the image, and peaks and valleys of the histogram are used to identify clusters in the image. A refinement of this technique is applied recursively searching the histogram method in the image groups divided into small groups. This is repeated with the poles of smaller groups and smaller until it is formed.

CLUSTERING METHOD

Clustering can be viewed as the problem of unsupervised learning is important because it provides information about the "right" answer to any of the objects. Here prior information about the classes is not compulsory, it is not the number of groups and the distribution rules in groups are known. They are to be found only in the data set without any reference to a training set. Cluster analysis allows many options on the nature of the algorithm to combine the groups. There are two basic approaches to clustering, which we call supervised and unsupervised. In the case of clustering the group, do not have labels. If we know the labels of our input data, the problem is considered supervised or unsupervised otherwise known.

Definition of Clustering: Clustering is a set of data with similar characteristics. In dividing the data into groups of similar objects, here the distance functions are used to determine the similarity of the two objects in the data set.

CLUSTER ANALYSIS

Cluster analysis or clustering is the assignment of a set of observations into subsets(called clusters)so that observations in the same cluster are similar in some sense. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics.

AGGLOMERATIVE HIERARCHIAL CLUSTERING

For example, suppose this data is to be clustered and the Euclidean distance is the distance metric for raw data then the hierarchical clustering dendogram would be as follows:

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 1, January 2017

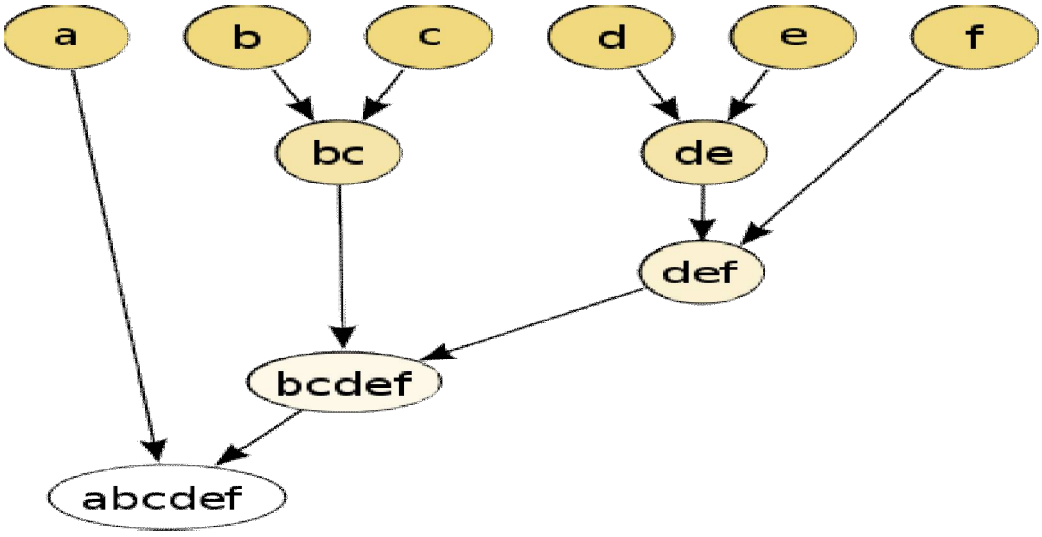


Figure 3: Traditional representation of raw data

This method builds the hierarchy of elements of particular groups of progressive melting. In our example, we have six elements {a} {b} {c} {d}{e}and{f}. The first step is to determine which elements to merge in a cluster. Usually we take the two closest elements, according to the selected distance.

Optionally, we can also create a distance matrix at this stage, where the number of i-th row j-th column is the distance between i-th and j-th elements. Then, as the group progresses, groups of rows and columns merged and the distances updated. This is a common way to implement this type of grouping, and has the advantage of catching distances between groups.
Suppose we have merged the two closest elements b and c, we now have the following clusters {a}, {b, c}, {d}, {e} and {f}, and want to merge them further. To do that, we need to take the distance between {a} and {b c}, and therefore define the distance between two clusters. Usually the distance between two clusters ***A*** and ***B***is one of the following:
The maximum distance between elements of each cluster (also called complete linkage clustering):

$$\max\{\, d(x,y) : x \in \mathcal{A},\, y \in \mathcal{B} \,\}.$$

The minimum distance between elements of each cluster (also called single-linkage clustering):

$$\min\{\, d(x,y) : x \in \mathcal{A},\, y \in \mathcal{B} \,\}.$$

The mean distance between elements of each cluster (also called average linkage clustering):

$$\frac{1}{|\mathcal{A}| \cdot |\mathcal{B}|} \sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} d(x,y).$$

Each agglomeration occurs at a greater distance between clusters than the previous agglomeration, and one can decide to stop clustering either when the clusters are too far apart to be merged (distance criterion) or when there is a sufficiently small number of clusters (number criterion).

K-MEANS CLUSTERING
One of the most popular and widely studied clustering algorithms to separate the input data in the Euclidian space is the K-Means clustering. It is a nonhierarchical technique that follows a simple and easy way to classify a given dataset

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 1, January 2017

through a certain number of clusters (we need to make an assumption for parameter k) that are known a priori. The K-Means algorithm is built using an iterative framework where the elements of the data are exchanged between clusters in order to satisfy the criteria of minimizing the variation within each cluster and maximizing the variation between clusters. When no elements are exchanged between clusters, the process is halted. The four steps of this algorithm are briefly described below:

Each cluster is defined by its centroid in n-dimensional space. Pixels are grouped by their proximity to cluster's centroids. Cluster centroids are determined using a heuristics: initially centroids are randomly initialized and then their location is interactively optimized.

Distance measure (usually Euclidian metric) between a data point $x_i^{(j)}$ and the cluster center c j (this is an indicator of the distance of the n data points from the cluster centers). There are situations when the K-Means algorithm doesn't find the optimal solution corresponding to the global objective function J and in addition is sensitive to the initialization process that selects the initial cluster centers that are usually randomly picked from input data. The main advantages of this algorithm are its simplicity and low computational cost, which allows it to run efficiently on large datasets. The main drawback is the fact that it does not systematically yield the same result each time the algorithm is executed and the resulting clusters depend on the initial assignments. The K-Means algorithm maximizes inter-cluster (or minimizes intra-cluster) variance, but does not ensure that the algorithm will not converge to local minima due to an improper starting condition (initialization of the cluster centers).

K- Means is a widely used clustering algorithm to partition data into k clusters.Clustering process is done for grouping data points with similar feature vectors into a single cluster and grouping data points with dissimilar feature vectors into different clusters. Let the feature vectors derived from l clustered data be X-{Xi/i=1,2-----l}. The generalized algorithm initiates k cluster centroids C={Cj/j=1,2,--------k} by randomly selecting k feature vectors from X. $Ci=\frac{1}{ni}\sum_{j=1}^{n(si)} mij \ \in si$

The centroids are said to be converged if their positions do not change in the iteration. It also may stop in the *t* iteration with a threshold ε if those positions have been updated by the distance as:

$$\left|\frac{c^t - c^{t-1}}{c^t}\right| < \varepsilon$$

Later the feature vectors are grouped into k clusters using a selected distance measure such Euclidean distance so that D=mod(Xi-Cj).

PILLAR K-MEANS CLUSTERING

The system uses the real size of the image in order to perform high quality of the image segmentation. It causes high-resolution image data points to be clustered. Therefore we use the K-Means algorithm for clustering image data considering that its ability to cluster huge data, and also outliers, quickly and efficiently. However, Because of initial starting points generated randomly, K-Means algorithm is difficult to reach global optimum, but only to one of local minima which it will lead to incorrect clustering results. Barakbah and Helen performed that the error ratio of K-Means is more than 60% for well-separated datasets. To avoid this phenomenon in this project uses our previous work regarding initial clusters optimization for K-Means using Pillar algorithm. The Pillar algorithm is very robust and superior for initial centroids optimization for K-Means by positioning all centroids far separately among them in the data distribution. This algorithm is inspired by the thought process of determining a set of pillars locations in order to make a stable house or building. Fig. 5 illustrates the locating of two, three, and four pillars, in order to withstand the pressure distributions of several different roof structures composed of discrete points. It is inspiring that by distributing the pillars as far as possible from each other within the pressure distribution of a roof, the pillars can withstand the roof's pressure and stabilize a house or building. It considers the pillars which should be located as far as possible from each other to withstand against the pressure distribution of a roof, as number of centroids among the gravity weight of data distribution in the vector space. Therefore, this algorithm designates positions of initial centroids in the farthest accumulated distance between them in the data distribution.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 1, January 2017

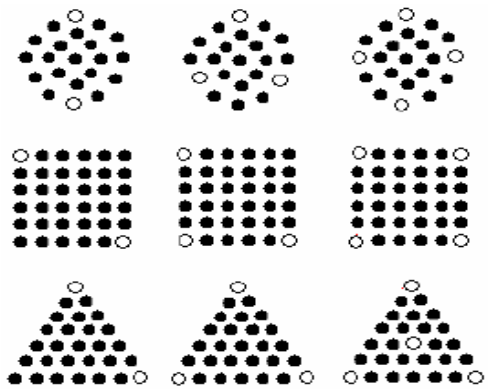


Figure 4: Illustrating of locating a set of pillars (white point) withstanding against different pressure distribution of roofs

MODIFIED PILLAR K-MEANS CLUSTERING

The M0dified Pillar algorithm is described as follows. Let $X=\{xi \mid i=1,\ldots,n\}$ be data, k be number of clusters, $C=\{ci \mid i=1,\ldots,k\}$ be initial centroids, SX be identification for X which are already selected in the sequence of process, $DM=\{xi \mid i=1,\ldots,n\}$ be accumulated distance metric, $D=\{xi \mid i=1,\ldots,n\}$ be distance metric for each iteration, and m be the grand mean of X. The following execution steps of the proposed algorithm are described as:

Algorithm:
1. Set $C=Ø$, $SX=Ø$, and $DM=[\ ]$
2. Calculate $D \leftarrow$ dis(X,m) ; **Hear I have used Chessboard distance instead of Euclidian distance method i.e., D←max[abs(x+s), abs(x-s)]**
3. Set number of neighbors $nmin = \alpha . n / k$
4. Assign $dmax \leftarrow argmax(D)$
5. Set neighborhood boundary $nbdis = \beta . dmax$
6. Set i=1 as counter to determine the i-th initial centroid
7. *DM = DM + D; Hear it is Accumulated Distance Metric (ADM)*
8. Select ж ←*xargmax*(*DM*) as the candidate for i-th initial centroids
9. SX=SX Uж
10. Set D as the distance metric between X to ж.
11. Set $no \leftarrow$number of data points fulfilling $D \leq nbdis$
12. Assign DM(ж)=0
13. If $no < nmin$, go to step 8
14. Assign $D(SX)$=0
15. $C = C$ Uж
16. $i = i + 1$
17. If $i \leq k$, go back to step 7
18. Finish in which C is the solution as optimized initial centroids

III.EXPERIMENTAL RESULTS

The practical results for **K-Means, C-Means, Pillar K-Means& Modified Pillar K-Means** for a particular image is shown below:

ENTER 1 FOR K-Means
ENETR 2 FOR C-Means
ENTER 3 for Pillar K-Means
ENTER 4 for Modified Pillar K-Means

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

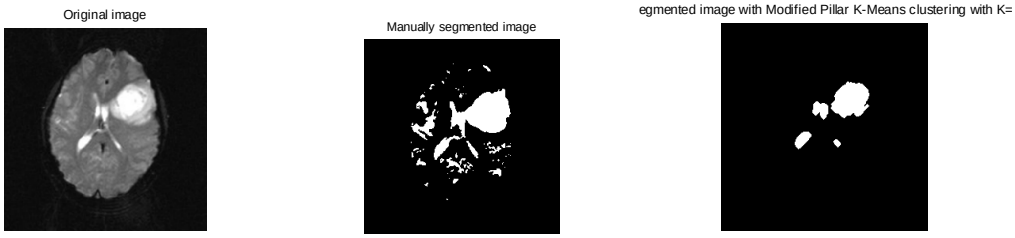
Vol. 6, Issue 1, January 2017

True positive value, Tp = [a(i,j)==1 || m(i,j)==1]
True negative value, Tn = [a(i,j)==0 || m(i,j)==0]
False positive value, Fp = [a(i,j)==1 || m(i,j)==0]
False negative value, Fn = [a(i,j)==0 || m(i,j)==1]

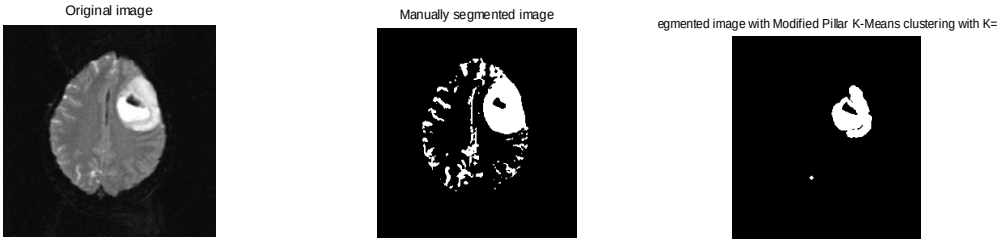
The formulae used to calculate Precision, Sensitivity Factor, Efficiency Factor & Computation Time are:
Similarity Index or Precision SI = [2*tp/(2*tp+fp+fn)]
Sensitivity Factor SF = [tp/(tp+fn)]
Efficiency Factor EF = [fp/(tp+fn)]
Computation Time Tm = cputime-t1 where t1 = time taken
Different sample MR Images are taken and simulated in MATLAB 7.5 version & results of modified Pillar K-Means clustering is shown below:

For sample image 1

For sample image 2



For sample image 3



ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 1, January 2017

For sample image 4

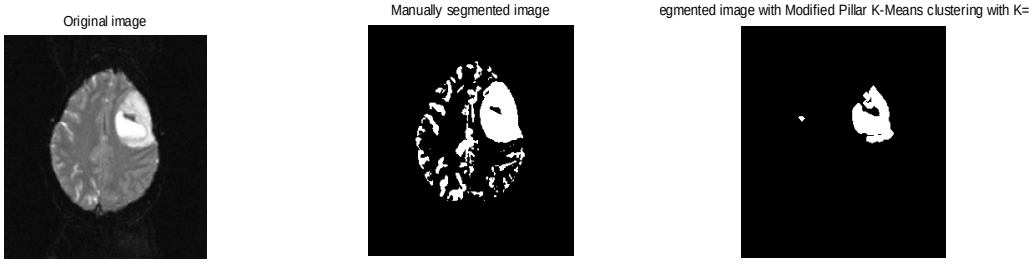


Figure 5: Shows the original MR Images, Manually segmented Images& Segmented Images with Modified Pillar K-Means clustering images with K=3 for different sample images 1, 2, 3 & 4.

The following tables give the information of different parameters such as Similarity Index or Precision, Sensitivity Factor, Efficiency Factor &Computation Time for all the four methods for different MR Images (The mean value is taken for all the parameters):

For sample image 1

Parameter	K-means	Fuzzy C-means	Pillar K-means	Modified PK-means
Precision	0.6336	0.6039	0.6203	0.6203
Sensitivity Factor	0.4703	0.9892	0.4613	0.4613
Efficiency Factor	0.0144	1.2869	0.0261	0.0261
Computation Time	**0.5781**	**5.9219**	**0.2031**	**0.1875**

For sample image 2

Parameter	K-means	Fuzzy C-means	Pillar K-means	Modified PK-means
Precision	0.7162	0.7097	0.6056	0.6056
Sensitivity Factor	0.5590	0.5507	0.4371	0.4371
Efficiency Factor	0.0019	0.0010	0.0064	0.0064
Computation Time	**0.2813**	**1.8594**	**0.3125**	**0.1719**

For sample image 3

Parameter	K-means	Fuzzy C-means	Pillar K-means	Modified PK-means
Precision	0.6768	0.6743	0.5925	0.5925
Sensitivity Factor	0.5123	0.5092	0.4232	0.4232
Efficiency Factor	0.0016	0.0012	0.0053	0.0053
Computation Time	**0.2969**	**1.6094**	**0.1875**	**0.1732**

For sample image 4

Parameter	K-means	Fuzzy C-means	Pillar K-means	Modified PK-means
Precision	0.6205	0.6169	0.5178	0.5178
Sensitivity Factor	0.4500	0.4462	0.3501	0.3501
Efficiency Factor	0.0040	0.0040	0.0020	0.0020
Computation Time	**0.3281**	**1.7188**	**0.2031**	**0.1406**

IV.CONCLUSION

In this Paper, we proposed a new approach to image segmentation using Modified Pillar K-Means algorithm for brain MR Images. The system applies the modified Pillar K-Means algorithm optimized for optimized segmentation. Pillar algorithm considers the placement of pillars should be located as far from each other to resist the pressure distribution of a roof, as same as the number of centroids between the data distribution. This algorithm is able to optimize the K-

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 1, January 2017

Means clustering for image segmentation in the aspects of accuracy and computation time. A series of experiments with five different brain MR Images was conducted and executed. The experimental results show that our proposed approach for image segmentation using Modified PillarK-Means algorithm is able to improve the accuracy and enhance the quality of image segmentation. We also made the computation time faster than Pillar K-Means and maintaining the quality of results.

REFERENCES

1. Hakeem AejazAslam, Prof.Tirumala.Ramashri and Mohammed Imtiaz Ali Ahsan "A New Approach to Image Segmentation for Brain Tumor detection using Pillar K-mean Algorithm", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 3, March 2013.
2. Ali RidhoBarakbah and Yasushi Kiyoki "A Pillar Algorithm for K-Means Optimization by Distance Maximization for Initial Centroid Designation", IEEE, 2009.
3. Ali RidhoBarakbah and Yasushi Kiyoki "A New Approach for Image Segmentation using Pillar-Kmeans Algorithm", World Academy of Science, Engineering and Technology 59, 2009.
4. RaselAhmmed and Md.FoisalHossain "Tumor Detection in Brain MRI Image Using Template based K-means and Fuzzy C-means Clustering Algorithm", International Conference on Computer Communication and Informatics (ICCCI-2016)", IEEE, 2016.
5. Swapnil R. Telrandhe, AmitPimpalkar and AnkitaKendhe " Detection of Brain Tumor from MRI images by using Segmentation & SVM", world conference on Futuristic Trends in Research and Innovation for Social Welfare (WCFTR-2016)", IEEE, 2016.
6. Anupurba Nandi "Detection of human brain tumour using MRI image segmentation and morphological operators", International Conference on Computer Graphics and Information Security (CGVIS)", IEEE, 2015.
7. Ms. NehaMathur, Mr. PankajDadheech and Mr. Mukesh Kumar Gupta "The K-Means Clustering based Fuzzy Edge Detection Technique on MRI Images", fifth International Conference on Advances in Computing and Communications", IEEE, 2015.
8. KailashSinha and G.R.Sinha "Efficient Segmentation Methods for Tumor Detection in MRI Images", Students Conference on Electrical, Electronics and Computer Science, IEEE, 2014.
9. Pavel Dvorak, Walter Kropatsch and KarelBartusek "Automatic Detection of Brain Tumor in MR Images", TSP, IEEE, 2013.
10. Satish Chandra, Rajesh Bhat and harinder Singh "A PSO based method for Detection of Brain Tumor from MRI", World Congress on Nature & Biologically Inspired Computing, IEEE, 2009.

BIOGRAPHY

Hakeem Aejaz Aslam did his bachelors from Visvesvaraya Technological University, Belagavi, Karnataka and Masters from Sri. Krishnadevaraya University, Anantapur, Andhra Pradesh. Pursuing Ph.D from Rayalaseema University, Kurnool, A.P. He is working as an Assistant Professor from last 10 years in MuffakhamJah College of Engineering and Technology, Hyderabad, Telangana, India.

Prof. T. Ramashri is a professor in Department of ECE, Sri Venkateswara University College of Engineering, Tirupati, A.P. has more than 25 journal papers with 20years of teaching experience & supervising 8 Ph.D scholars.