Recent Trends in Power, Control and Instrumentation Engineering, PCIE

Extended Pillars K-Means Clustering for Automatic Brain Tumor Technique

[1]k. Sudharani, [2]Dr.T.C. Sarma and [3]Dr. K. Satya Prasad
[1]Associate Professor,VNR Vignana Jyothi IET, Hyderabad, Andhra Pradesh,
[2]Former Deputy Director, NRSA , Hyderabad, Andhra Pradesh,
[3]Professor, JNTUKakinada, Kakinada, AndhraPradesh ,
e-mail: sudhasarah@gmail.com,sudharani_k@vnrvjiet.in

***Abstract*—Tumor is an uncontrolled growth of tissue in any part of the body. This paper is to implement of simple Algorithm for detection of range and shape of tumor in brain MR Images. Performance of K-means algorithm which depends highly on initial starting points can be trapped in local minima and led to incorrect clustering results. The lack of K-means algorithm that generates the initial centroids randomly does not consider the placement of them spreading in the feature space. In this paper a new approach to optimize K-means with morphological segmentation techniques is presented. Statistical parameters like sensitivity, specificity, similarity index, accuracy of the proposed algorithm is presented along with the area calculations of the defective part of the brain. Comparative study is made between k-means and proposed k-means methodologies.**

***Index Terms*— Detection and Localization System; Area; Sensitivity; Specificity; Similarity Index; and Ground Truth Images.**

I. INTRODUCTION

The objective of clustering techniques is to identify similar patterns in data. A cluster usually contains a group of similar pixels that belongs to a specific region and different from other regions. Images can be grouped based on this content. In content based clustering, grouping is done depending on the inherited characteristics of the pixels like shape, texture etc. Cluster analysis is the task of grouping up of pixels in an image in such a way that pixels in the same group are more similar to each other than to those in other groups (or clusters) and this characteristic more important in highly accurate statistical results for the images. The most widely used and faster methods for clustering is K-means clustering. The simple understanding of K-means clustering made it possible for this algorithm to be employed in various fields. K-means clustering is a subdivided clustering method that separates data into *k* mutually similar groups. Through such the iterative partitioning, K-means clustering minimizes the sum of distance from each data to its clusters. K-means clustering is very popular because of its ability to cluster huge data and also outlines, quickly and efficiently [1].

K-means clustering is very sensitive to the designated initial starting points as cluster centers. K-means clustering generates initial clusters randomly. If a randomly designated initial starting point is close to a final cluster center, then K-means clustering can find the final cluster center. It, however is not always possible. If a designated initial point is far from the final cluster center, it will lead to incorrect clustering results [2]. Because of initial starting points generated randomly, K-means clustering does not guarantee the unique

Grenze ID: 02.PCIE.2015.3.505

clustering results [3]. K-means clustering is difficult to reach global optimum, but only to one of local minima [4]. An object density-based image segmentation [5], [6] is used, which incorporates intensity-based, edge based and texture-based segmentation techniques. The method consist of three main stages: preprocessing, object segmentation and final segmentation. Image enhancement, noise reduction and layer-of-interest extraction are some of the tasks of preprocessing. The FCM algorithm [7], [8], attempts to partition a finite collection of pixels into a collection of "C" fuzzy clusters with respect to certain defined criterion. Depending on the data and the application, different types of similarity measures may be used to identify classes. Several methods were proposed to solve the cluster initialization for K-means clustering [9]. Stochastic k-means algorithms were also proposed this method associates both vector and cluster according to the probability distribution and it depends on the vector and cluster gravity mean distance. Because of initial starting points generated randomly, K-means algorithm is difficult to reach global optimum, but only to one of local minima [10] which it will lead to incorrect clustering results [11]. In [12] performed poorly such that the error rate of K-means is more than 60% for well-separated datasets. To overcome this limitation we use our previous work regarding initial clusters optimization for K-means using Pillar algorithm [13], the approach in this paper designates positions of initial centroids by using farthest accumulated distance between them and the procedure presented in this paper is iterative. The Pillar algorithm is very robust, superior and extremely fast. The algorithm presented in this paper uses Morphological operations in addition to the pillar k-means algorithm which enhances the use of pillar technique by increasing its accuracy without affecting its previous qualities regarding dealing with the images.

II. Morphological Segmentation

Morphological operations are based on set theory of mathematics and rely only on the relative ordering of pixel values in the give image, sets in mathematical morphology represents objects in an image. Therefore, are especially suited for the processing of binary images. Morphology can also be applied to greyscale images such that their light transfer functions are unknown and their absolute pixel intensity values are of no or minor interest. Morphological techniques scrutinize an image with a small template called a structuring element. The structuring element is moved through the all possible locations in the main image and it compares with the corresponding neighborhood of pixels. Some operations test whether the element "fits" within the neighborhood, while others test whether it hits or intersects the neighborhood. The erosion of a grey image P by a structuring element Q produces a new image

$$g = P \ominus Q \qquad (1)$$

with ones in all the locations (x,y) of a structuring element's origin at which that structuring element Q fits the input image Q, i.e, g(x,y) = '1' is Q fits P and '0' otherwise, Repeating for all pixel coordinates (x,y), if the given pixels is or is not member of the structuring element set, the location is marked as don't care and is represented by any letter like 'q'. Erosion with small square structuring elements shrinks an image by stripping away a layer of pixels from both the inner and outer boundaries of the main image P.

Larger structuring elements have a more pronounced effect. The result of erosion with a large structuring element being similar to the results obtained by iterated erosion using a smaller structuring element of the same shape. If Q_1 and Q_2 are a pair of structuring elements identical in shape, with Q_2 twice the size of Q_1, then

$$a = (P \ominus Q_2) \approx (P \ominus Q_1) \ominus Q_1 \qquad (2)$$

Erosion removes small-scale details from a binary image but simultaneously reduces the size of the regions of interest, too. By subtracting the eroded image from the original image, boundaries of each region can be found by the relation

$$b = P - (P \ominus Q) \qquad (3)$$

where I is an image of the regions, Q is a 3×3 structuring element, and b is an image of the region boundaries. The Dilation of an image I by a structuring element Q produces a new binary image

$$h = P \oplus Q \qquad (4)$$

with ones in all locations (x,y) of a structuring element's origin at which that structuring element s hits the the input image P, i.e. $g(x,y)$ = '1' if Q hits P and "0" otherwise, Repeating for all pixel coordinates (x,y). Dilation has the opposite effect to erosion, it adds a layer of pixels to both the inner and outer boundaries of

regions. The holes enclosed by a single region and gaps between different regions become smaller, and small intrusions into boundaries of a region are filled in. Results of dilation or erosion are influenced both by the size and shape of a structuring element.
Dilation and erosion are dual operations with respect to set complementation and reflections. Let I^c denote the complement of an image I, i.e., the image produced by replacing 1 with 0 and vice versa. Formally, the duality is written as

$$P \oplus Q = P^c \ominus Q_{rot} \qquad (5)$$

The duality property is more useful when structuring element has symmetry about its own origin. Under such conditions the erosion of the image can easily be obtained by dilating its background with structuring element and then complementing the result.
Both Opening and closing are the other important operations of morphology. Opening smoothens the contour of an object, breaks narrow striations and eliminates thin eminences. Whereas closing also tends to smooth sections of contours but as opposed to opening, it generally fuses narrow breaks and long thin gulfs, eliminates small holes and fills gaps in the contour.

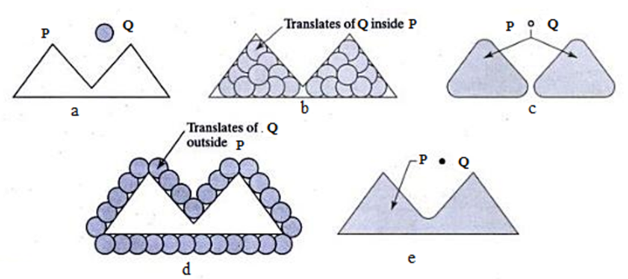


Fig 1. Opening and closing as unions of translated structuring elements. a) Set P and structuring element Q. b) translations of Q that fit entirely within set P. c) the complete opening (shaded). d) Translations of Q outside the border of P. e) the complete closing (shaded)

III. Image Segmentation using Pillar Algorithm

The Pillar k-means algorithm is used for clustering image data considering its ability to cluster huge data, as well as outlines, quickly and efficiently. However, because of initial starting points generated randomly, Pillar k-means algorithm is difficult to reach global optimum, but only to one of local minima which it will lead to incorrect clustering results. Performed that the error ratio of Pillar k-means is more than 60% for well-separated datasets. To overcome this drawback in this project uses initial cluster optimization through for Pillar k-means algorithm. The Pillar algorithm is very robust and superior for initial centroids optimization as Pillar 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. 1 illustrates the locating of two, three, and four pillars, in order to withstand the maximum pressure distributions of different roof structures composed of discrete points. It is necessary 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 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.

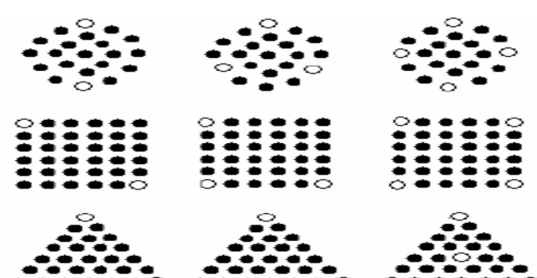


Figure.2 Illustrating of locating a set of pillars (white point) withstanding against different pressure distribution of roofs

The process flow steps for the k-means clustering and extended k-means clustering algorithms are presented in the following sections a and b respectively.

A. K-Means Clustering Algorithm

1. Input Image
2. Give the no of cluster value as k.
3. Randomly choose the k cluster centers
4. Calculate mean or centroid of the cluster
5. Calculate the distance between each pixel to each cluster centroid.
6. If the distance is near to the center then move to that cluster.
7. Otherwise move to next cluster.
8. Re-estimate the centroid iteratively.

This algorithm minimizes the total distance of data points with respect to the cluster centroid, of the cluster they are assigned to. Also it does not require the actual computation of distances. A drawback of k-means algorithm is that the number of desired clusters needs to be set before implementation. It implies that the data clusters are ball-shaped because it performs clustering based on the Euclidean distance only. Although hierarchical methods can be more accurate, partitioned methods are used in applications involving large data sets (related to images) due to the fact the construction of a dendrogram (a tree diagram) is computationally prohibitive **[14]**. In hierarchal method if once the merge/split is done, it can never be undone. This rigidity will be a limitation in some case and can be useful in some other as it leads to lesser computation costs as it will not worrying about a combinatorial number of different choices**.**

B. Extended K-Means Algorithm

1. Input image
2. Converting pixel intensities from single precision into double precision number.
3. Removing the salt and pepper noise using the median filter.
4. Applying Morphological Operations.
5. Give the no of cluster value as k.
6. Randomly choose the k cluster centroids.
7. Calculate mean or centroid of the cluster.
8. Calculate the distance between each pixel to each cluster centroid.
9. If the distance is near to the centroid then move to that cluster.
10. Otherwise move to next cluster.
11. Re-estimate the centroid iteratively.

IV. Comparative Study of Results Obtained By Both Pillar K Means and Extended Pillar K Means

Many techniques are being used on brain tumor analysis. One of them is clustering analysis. In clustering analysis fuzzy techniques are also used. In the present study an extension of Pillar k-means algorithm for tumor analysis is employed. A comparative study is also made here in between pillar k-means algorithm and extended pillar k-means algorithm. Brain tumor images are tested with both pillar k-means and pillar k-means technique with morphological operations. The result obtained with both the algorithms is presented in the following figure. From the figure it can be noted that the extended pillar method is fund to be superior to pillar method in dealing with the tumor regions. Noise content is more with Pillar k-means techniques after detecting the tumor inside the brain whereas extended pillar technique reduces the noise better than pillar technique.

Comparative study is enhanced by showing the result of individual brain tumor image. The results are presented in the following table-I. The table also represents the status of the tumor i.e, whether the tumor image is detected properly or not. No tumor images are also tested with both the algorithms.

V. Calculation Of Statistical Parametres

The detection status of all the brain tumor images analyzed through these two techniques are also presented in tabel-1. The statistical parameters like Sensitivity, Specificity, Accuracy and Similarity Index their definitions and standard relations for calculating them are presented in this section.

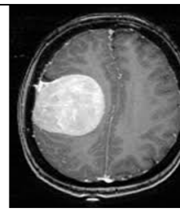
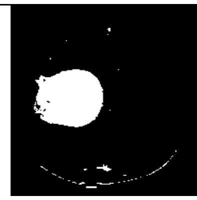
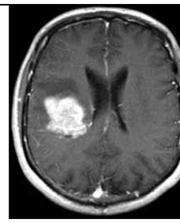
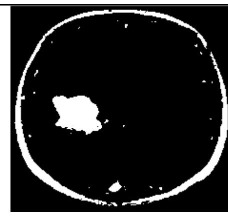
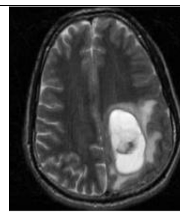
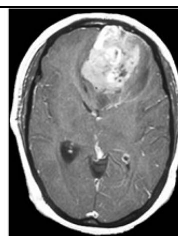
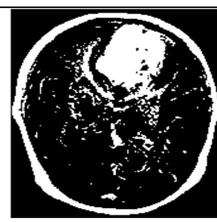
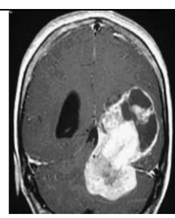
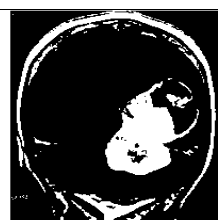
Patient details	Input image	pillar k means	Pillar k means and morphological operation (novel)
Patient 1			
Patient 2			
Patient 3			
Patient 4			
Patient 5			

Fig 3. Comparison between Pillar K Means and Extended Pillar K Means

- **True positives (TP)**: Brain tumor images are correctly recognized. Which means the people who have brain tumor are correctly identified.
- **False positives (FP)**: Non-Brain tumor images are incorrectly recognized. This indicates the people who do not have brain tumor are incorrectly identified as they have brain tumor. Simply Healthy people incorrectly identified as sick.
- **True negatives (TN)**: Non-Brain tumor images are correctly recognized as they do not have brain tumor. Easily Healthy people correctly identified as healthy.
- **False negatives (FN)**: Brain tumor images are incorrectly recognized. Which represents the people who have brain tumor are incorrectly identified as they do not have brain tumor. Sick people incorrectly identified as healthy.

TABLE.1 COMPARISSION BETWEEN PILLAR K-MEANS WITH MORPHLOGICAL AND PILLAR KMEANS

S.no	Patients	Brain Area(square mm)	Pillars K-Means Technique(tumor area sq.mm)	Status	Extended Pillar K-Means Technique(tumor area sq.mm)	Status
1	PATIENT1	21478	421.5	TP	413.0	TP
2	PATIENT2	25823	218.3	TP	213.6	TP
3	PATIENT3	18818.02	118.4	TP	118.4	TP
4	PATIENT4	24095	451.8	TP	448.6	TP
5	PATIENT5	27664	512.4	TP	480.0	TP
6	PATIENT6	29386	275.7	TP	270.4	TP
7	PATIENT7	33563	394.3	TP	390.0	TP
8	PATIENT8	35997	449.6	TP	449.6	TP
9	PATIENT9	34487	323.6	TP	325.0	TP
10	PATIENT10	25824	320.0	TP	300.0	TP
11	PATIENT11	14420	0	FN	0	FN
12	PATIENT12	24953	0	FN	69.0	TP
13	PATIENT13	34521	237.6	TP	510.0	TP
14	PATIENT14	27337	294.0	TP	291.0	TP
15	PATIENT15	29656	32.0	TP	32.0	TP
16	PATIENT16	24609	447.0	TP	413.0	TP
17	PATIENT17	27555	254.3	TP	250.0	TP
18	PATIENT18	19678	116.2	TP	103.0	TP
19	PATIENT19	24997.1	15.3	TP	14.0	TP
20	PATIENT20	29569	32.0	TP	38.0	TP

$$\text{Sensitivity} = \frac{\text{True Postives}}{\text{True Positives+False Negetives}} * 100\% \quad (6)$$

$$\text{Specificity} = \frac{\text{True Negetives}}{\text{True Negetives +False Positives}} * 100\% \quad (7)$$

$$\text{Accuracy} = \frac{\text{True Postives+True negetives}}{\text{TP+TN+FP+FN}} * 100\% \quad (8)$$

$$\text{Similarity index} = \frac{2(\text{True Postives})}{2(\text{TP})+\text{False Positives+FN}} * 100\% \quad (9)$$

To calculate above mentioned parameters we need tumor images and no tumor images.
The outcome of the statistical analysis:

TABLE-II : CONSOLIDATED DETECTION STATUS OF BOTH ALGORITHMS

Total Number Of Images Tested	Total Number Of Tumor Images Tested	Pillar K-Means Clustering				Extended Pillar K-Means Clustering			
		TP	TN	FP	FN	TP	TN	FP	FN
36	20	18	16	0	2	19	16	0	1

TABLE-III. THE RESULTS OF THE ANALYSIS OF THE STATISTICAL PARAMETERS OBTAINED THROUGH BOTH THE TECHNIQUES

Method	Sensitivity	Specificity	Accuracy	Similarity Index
K-Means Clustering	90	100	94.44	94.7
Extended Piller K-Means Clustering	95	100	97.22	97.44

VI. CONCLUSION

The combination of extended k-means clustering with morphological technique proposed by the author yields better results with low noise compared to simple pillar k-means technique. It is very important and a different task to analyze medical images with noise .Hence, it is essential to improve the quality of the images prior to analyzing them. In the present analysis, the images are enhanced prior to subjecting them for clustering and further analysis.

REFERENCES

[1] H. Ralambondrainy, "A conceptual version of the Kmeans algorithm", Pattern Recognition Lett, 16, 1147-1157, 1995

[2] Y.M. Cheung, "k*-Means: A new generalized kmeans clustering algorithm", Pattern Recognition Lett, 24, 2883-2893, 2003.

[3] S.S. Khan, A. Ahmad, "Cluster center initialization algorithm for K-means clustering", Pattern Recognition Lett, 2004.

[4] B. Kövesi, J.M. Boucher, S. Saoudi, "Stochastic Kmeans algorithm for vector quantization", Pattern Recognition Lett, 22, 603-610, 2001.

[5] M. Schmidt, I. Levner, R. Greiner, A. Murtha, and A. Bistritz. "segmenting Brain Tumours using Al ignment-Based Features".International Conference on Machine Learning and Applications,Los Angeles, December 2005.

[6] Juraj Horvath, 2006 "Image Segmentation using Fuzzy C-means" SAMI 2006.

[7] M. Rakesh, T. Ravi ― Image Segmentation and Detection of Tumour Objects in MR Brain Images Using FUZZY C-MEANS (FCM) Algorithm" International Journal of Engineering Research and ApplicationsVol. 2, Issue 3, May-Jun 2012, pp.2088-2094.

[8] Jie Yu, Peihuang Guo, Pinxiang Chen, Zhongshan Zhang and Wenbin Ruan, "Remote sensing image classification based on improved fuzzy c -means", Ceo-Spatial Information Science, vol.11, no.2, pp:90-94, 2008.

[9] P. Tamije Selvy, V. Palanisamy, T. Purusothaman" Performance Analysis of Clustering Algorithms in Brain Tumour Detection of MR Images" EuroJournals Publishing, Inc. 2011.

[10] B. Kövesi, J.M. Boucher, S. Saoudi, "Stochastic K-means algorithm for vector quantization", Pattern Recognition Letters, Vol. 22, pp. 603-610,2001

[11] Y.M. Cheung, "k*-Means: A new generalized k-means clustering algorithm", Pattern Recognition Letters, Vol. 24, pp. 2883-2893, 2003.

[12] A.R. Barakbah, A. Helen, "Optimized K-means: an algorithm of initial centroids optimization for K-means", Proc. Seminar on Soft Computing, Intelligent System, and Information Technology (SIIT), Surabaya, 2005.

[13] A.R. Barakbah, Y. Kiyoki, "A Pillar Algorithm for K-Means Optimization by Distance Maximization for Initial Centroid Designation", IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Nashville-Tennessee, 2009.

[14] Arnau Oliver, Xavier Mu˜noz, Joan Batlle, Llu´ıs Pacheco, and Jordi Freixenet, "Improving Clustering Algorithms for Image Segmentation using Contour andRegion Information", IEEE transactions, 2006.