Clustering Based Image Segmentation: A Case Study

Harendra Sharma[1], Akanksha Bisht[2], Ritu Payal[3]
[1]*Computer Science and Engineering Dev Bhoomi Institute of Technology Dehradun, India*
[2]*JRF, Research and Development University of Petroleum and Energy Studies Dehradun, India*
[3]*Computer Science and Engineering Dev Bhoomi Institute of Technology Dehradun, India*
(E-mail: harendra.pth@gmail.com[1], bisht.akanksha35@gmail.com[2], ritupayal1234@gmail.com[3])

***Abstract*—**Tumor is formed inside the brain or skull when there is an abnormal growth of cells. There are various imaging techniques like X-ray, CT scan and MRI to detect any abnormal changes in the human body. Out of these techniques, MRI is the safest and reliable technique.

Tumor can be segmented after processing the MRI image. This paper presents a study of image segmentation techniques that are fuzzy c mean algorithm, fuzzy k mean algorithm, and level set (polynomial method), and comparison of the result of these techniques.

Keywords*—MRI, Fuzzy C-Means, Fuzzy K-Means, Level Set, Tumor, Segmentation.*

I. INTRODUCTION

The most complicated part of human body is brain. The symptoms of brain tumor can be migraine, frequent headaches or vision loss over the time. Brain tumors are hard to detect and can be life threatening in many cases. So the techniques used for detection of brain tumor should be accurate. MRI scan provide more information even about the minor abnormalities, than other techniques. It uses radio waves and magnetism in order to produce accurate images. After retrieval of scanned images, the next step is to detect the tumor, its location and size. This is where Image processing techniques comes to the scenario, to detect the tumor.

The task of detection of tumor includes two processes- first is pre-processing and second is image segmentation. Image pre-processing includes conversion of MRI image to gray scale, noise removal and reduction, enhancing and reconstructing image, and in case of brain MRI it includes removal of skull from MRI. Segmentation is the process of splitting an image into multiple set of pixels that share similar characteristics. This makes it easier to analyse further and extract information [1]. For segmenting an image several methods are used that are described below:

A. Region based:

In this method, first a seed pixel is selected then other nearby pixels that follow similar property as the seed pixel, are put together to form a region [2]. This approach starts with the seed pixel and grows outward, until it meets the boundary of the region. The application of this method can be image or face identification, and pattern recognition [3].

B. Edge based:

In this method, a sober operator is used to detect the edge between the two regions, then that image is used for further processing. The physical extent of the objects is characterized by edges so the accurate detection of edge is very important in pattern recognition and image analysis [4].

C. Threshold based:

This method is used for segmenting the images that have light objects on grim background. For partitioning the image, slicing method and histogram thresholding is used. It transforms a multilevel image into binary image and it chooses a threshold T, to separate the object from its background and dividing pixels of the image in multiple regions [3].

D. Clustering based:

In this method multiple clusters are formed and pixels that belong to the same cluster have similar properties while pixels that belong to different clusters have dissimilar properties. It deals with unlabelled data in order to find the structure. This method is considered as the most important unsupervised learning problem [5]. To solve the clustering problems, the simplest algorithm used are fuzzy c-means algorithm and fuzzy k means algorithm.

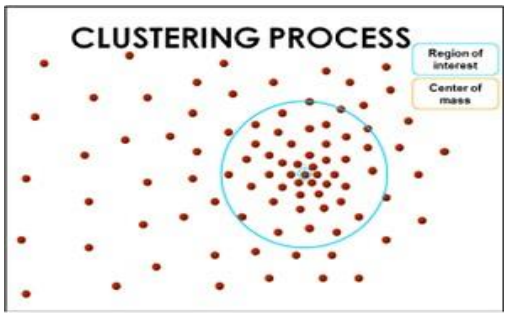

Figure 1: Clustering of Pixels.

II. LITERATURE REVIEW

Image segmentation can be done in easier and effective way by using the clustering based methods. It increases the efficiency of the process of accurate image retrieval and feature extraction. In this paper, three clustering technique that are fuzzy c means, fuzzy k means and level set (polynomial method) are reviewed and their performance is compared.

1. Fuzzy c means:

This algorithm is unsupervised fuzzy clustering algorithm. Traditional clustering algorithm used to be hard-partition, means the pixel or data set can only belong to only one cluster. While fuzzy c mean algorithm belongs to soft-partition category, the pixel or data set can belong to two or more cluster [6]. Generally, there is unsharp transition at the boundary of the cluster region, so it is hard to determine that a pixel belongs to that region or not. For characterization of an image by the member function, 3 basic features are involved that are- core, support and boundary. Core is the full member, support is non-membership value and boundary is partial member of the fuzzy set whose value lie between 0 and 1. For fuzzy partition, an iterative optimization of object function is used, while updating the cluster center and membership function. The distance between data point and cluster center decides the membership of data point towards the cluster center. Fuzzy c means [FCM] provide better results when a particular data point belongs to more than one cluster and overlapped region [7].

FCM algorithm steps:

1. Select the image in JPEG format.

2. If image in RGB format, convert it to grey scale.

3. For increasing the range of pixel value, convert it to double.

4. Define the number of clusters and iterations.

5. Get the size of image and convert image matrix into a vector.

6. Select the cluster centre randomly.

7. From the image vector, calculate the fuzzy Centre by the formula:

$$R_i = \sum_{i=1}^{N} xi.\, M_{ij}{}^{m} / \sum_{i=1}^{N} M_{ij}{}^{m}$$

Where,

R_i = d-dimension center of the cluster vector

M_{ij} = degree of the membership function of x in the cluster j

x_i = data measured in d-dimension

m = any real number greater than 1

8. Calculate the membership function of the fuzzy set using the distance formula,

$$M_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - x_j\|}{\|x_i - c_j\|} \right)^{\frac{2}{m-1}}}$$

9. Repeat steps 8 and 9 until achieving the minimum value of the equation,

$$Ym = \sum_{i=1}^{N} \sum_{j=1}^{C} M_{ij}^{m} \|x_i - c_j\|$$

10. If the condition is true in the equation then stop otherwise go to step 8.

$$Max_{ij} \left| M_{ij}^{(k+1)} - M_{ij}^{(k)} \right| < \delta$$

Where,

δ= termination value between 0 and 1

k= no of iteration steps.

2. Fuzzy K means:

In k means clustering, data points are grouped in k number of cluster based on attributes where k is a positive number. Euclidean distance between data and the cluster center is used to cluster data points. The cluster center which has shortest distance with the pixel, pixel is moved to that cluster [8]. A set of data points {x1, x2, …xN} are put in k clusters. It supports multidimensional vectors and has high computational efficiency. So by minimizing a cost function, fuzzy k means reduces the distortion measure as:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} b_{nk} \| x_n - c_k \|^2$$

$$b_{nk} = \begin{cases} 1 & \text{if k=arg min}_a \| x_n - c_a \|^2, \text{ a=1,....k} \\ 0 & \text{Otherwise} \end{cases}$$

$$c_k = \frac{1}{N_k} \sum_{x \in C_k} x$$

Where, $\| x_n - c_k \|$ is the distance between data point and cluster center.

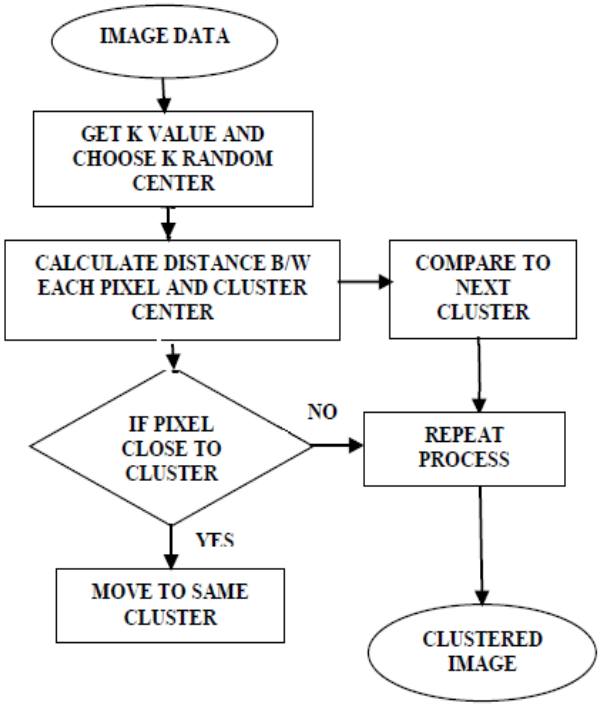

Figure 2: Flowchart of K-means clustering strategy.

K mean algorithm:

1. Randomly choose the number of clusters as, k.
2. Choose cluster center randomly.
3. Calculate the cluster center.
4. Calculate distance between pixel or data point and center of the cluster.
5. Move pixel to the cluster whose center is in minimum distance.
6. Re-evaluate the cluster center.
7. Until the cluster center stops changing, repeat the process.

3. Level Set

Fuzzy k means is simple and fast algorithm that gives better on larger data sets but it provides incomplete result in case of malignant tumor which can be easily detected by the system that use fuzzy c means. In level set method, the main idea is to reduce the iteration of finding cluster center in fuzzy c mean. As fuzzy k means, this method take k random number of clusters and choose random cluster center.

Algorithm:

1. Convert RGB image into grey scale.
2. Choose k number of clusters and cluster centre randomly, and start the iteration.
3. Convert pixel value into double for increasing the range.
4. Calculate the Euclidean distance between pixel or data point and cluster center.
5. Evaluate membership function and assign new center using the equation,

$$CCC1 = \frac{\sum(\sum U1 * U2 * double\ (image))}{\sum(\sum U1 * U2)}$$

6. Repeat the process until minimum value is achieved.

III. FINDING AND DISCUSSSION

We have discussed about clustering techniques like Fuzzy C- Means, K-Means and Level Set (Polynomial Method) in this paper. K means algorithm detects the brain tumour faster than the fuzzy c means algorithm, while fuzzy c means predicts brain tumour cell more precisely. Fuzzy c means can't segment brain tumour cells if the image is corrupted by the noise or outlier. On the other hand, level set (polynomial method) detects the brain tumour more precisely and in the lesser time. The accuracy, precision and recall value that we get by these methods are given in the table given below:

TABLE I. THE PERFORMANCES MATRICES

Clustering	Data Set	Accuracy	Precision	Recall
Fuzzy C- Means	DS 1	74.5	100	74.5
	DS 2	92.7	100	92.7
K- Means	DS 1	81.2	100	81.2
	DS 2	90.8	100	90.8
Level Set (Polynomial Method)	DS 1	95.7	100	95.7
	DS 2	94.8	100	94.8

IV. CONCLUSION

As from the above discussion on clustering method for MRI brain tumor segmentation we find that the Level Set (Polynomial Method) gives more accurate result than the other two algorithm i.e Fuzzy C-Means and K-Means. The results are found on two data set DS-1 and DS-2 in which we have found the Accuracy, Precision and Recall of the images in data sets.

REFERENCES

[1] Luxit Kapoor, Sanjeev Thakur. A Survey on Brain Tumour Detection using Image Processing Techniques. International Conference on Cloud Computing, Data Science & Engineering (2017).

[2] Angelina S, Suresh L P, Veni S H K. Image segmentation based on genetic algorithm for region growth and region merging. Computing, Electronics and Electrical Technologies (ICCEET)