Medical Image Segmentation using PCA and K-mean Clustering Algorithm

Juilee Anil Katkar
Department of IT
Maharashtra Institute of Technology
Pune, India kat15792@gmail.com

Trupti Baraskar
Department of IT
Maharashtra Institute of Technology
Pune, India
trupti.baraskar@mitpune.edu.in

***Abstract*—Medical images are widely used by physician to find abnormalities in human bodies. However the image sometimes is corrupted with noise which normally exist or occur during storage or while transferring the image. Therefore need to improve the image quality, lowdown the computational complexity and signal to noise ratio to improve performance analysis of quantitative and qualitative capabilities of medical image so, there is need of segmentation. Segmentation is technique for partitioning image into meaningful sub regions or object with same attribute. Proposed method states Principle component analysis and K-means clustering method for segmentation of medical images and extracts meaningful part from medical image in efficient manner.**

Keywords: Image Segmentation; clustering; PCA; K-mean Clustering.

I. INTRODUCTION

Image Segmentation is process of extracting meaningful part from image. Image Segmentation filters the important part from rest of image. Image segmentation is process of partitioning the digital image into multiple segments these segment consist of set of pixels. The goal of segmentation is to simplify and change the representation of image into something meaningful and easier to analyze and process. Use of image segmentation is to locate object and boundaries in image and its process of assigning label to every pixel in image such that pixel with same label share certain characteristics [1].

Input of image segmentation is any digital form of image output is set of segments that collectively cover entire image set of contours extracted of entire image. Properties like gray scale, color, intensity, texture, depth help to recognize similar regions and similarity of such properties is used to construct groups of region. Image segmentation plays important role in human visual perception. Before high level reasoning can be applied to an image, it must broken into its major structural components. So apply image segmentation filter only those part which is important [1].

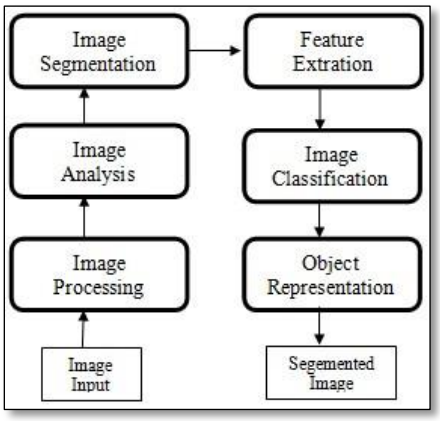

Figure 1: Image Segmentation

II. CLASSIFICATION OF IMAGE SEGMENTATION

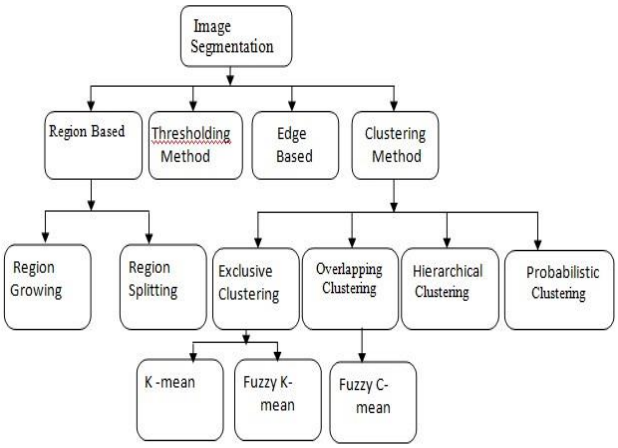

Figure 2: Classification of Image Segmentation

Following methods are use for medical image segmentation:

A. *Region based image segmentation*:

Region based methods, partition an image into regions that are similar according to a set of predefined criteria. Region base segmentation base on following methods [2].

1) Region growing:

Region growing is a procedure that group's pixels in whole image into sub regions based on predefined criteria.

(i) Selection of seed pixels from given image.

(ii) Selection of similarity criterion such as grey level intensity, shape, size, depth of pixel etc.

(iii) Grow regions by appending to each seed those neighboring pixels that have predefined properties similar to seed pixels.

(iv) Stop region growing when no more pixels met the criterion for inclusion in that region [2].

2) Region Splitting:

Instead of choosing seed points, user can divide an image into a set of arbitrary unconnected regions and then merge the regions. Region splitting and merging is usually implemented with theory based on quad tree data [2].

B. *Thresholding method***:**

Threshold T value object is separated from background.When there is single threshold T value, any pixel for which f(x,y)>T is consider as object pixel and f(x,y)<T is background pixel Thresholding techniques is mainly divided as local thresholding and global thresholding [2].

C. Edge based method:

Edge detection is the process of identifying and location discontinuities in an image.The discontinuities are abrupt changes in pixel intensity which characterize boundries of object [2].

D. *Clustering method:*

Clustering is the process of organizing objects into groups whose members are similar in some way. **Cluster** is the process of organizing objects into groups whose members are similar in some way.

Exclusive Clustering,In this type of clustering data that is grouped in an exclusive way, so that if a particular data belongs to a definite cluster then it could not be included in another cluster.**Overlapping Clustering**, It uses fuzzy sets to cluster data, so that each point may belong to two or more clusters with different degrees of membership. **Hierarchical Clustering**, Builds (agglomerative), or breaks up (divisive), a hierarchy of clusters. **Probabilistic Clustering**, Attempt to optimize the fit between the data and the model using a probabilistic approach. Each cluster can be represented by a parametric distribution, like a Gaussian (continuous) or a Poisson (discrete) and the entire data set is therefore modeled by a mixture of these distributions [2].

III. RELATED WORK:

Medical images are widely used by physician to find abnormalities in human bodies. However the image sometimes are corrupted with noise which normally exist or occur during storage or while transferring the image. Therefore need to improve the image quality, lowdown the computational complexity and signal to noise ratio to improve performance analysis of quantitative and qualitative capabilities of segmentation algorithms. Segmentation is technique for partitioning image into meaningful sub regions or object with same attribute.

So, while doing research there were our three main objective are given below:

A. Study of various medical image segmentation methods

B. Comparative analysis of various image segmentation methods [1][2].

C. Comparative study of clustering methods of medical image segmentation [4][5][6].

TABLE I: Comparative Study Of Segmentation Techniques

Segmentation technique	**Method description**	**Advantages**	**Disadvantages**	**Execution time**	**Application**
Clustering method	categorize the points into clusters	for implementation	results remains unclear	FCM 1.469	Measure tissue volumes,
Region based	Group pixels into homogeneous regions	Work best for region homogeneity	Expensive both in computational time and memory	Region Growing 5.78 Region Splitting 5.906	Neural network edge pattern, pixel aggregation
Edge Detection method	locate points with abrupt changes in gray level	works well for images having good contrast between regions	Doesn't work for ill-defined edges	Canny 1.786	Face detection

TABLE II: Comparison of clustering techniques

Clustering Techniques	Advantage	Disadvantage
K means	1) Fast, robust and easier to understand. 2) Relatively efficient 3) Gives best result when data set are distinct	1) Learning algorithm requires apriori specification of the number of cluster centers. 2) Two highly overlapping data cannot resolve into two clusters.
Fuzzy C means	1) Gives best result for overlapped data set 2) Data point may belong to more than one cluster center.	1) Apriori specification of the number of clusters. 2) With lower value of β we get the better result but at the expense of more number of iteration.

IV. PROPOSED METHODOLOGY

Central problem, called *segmentation*, is to distinguish objects from background. This is used in many fields such as medicine and biology. Image segmentation has come a long way. Using just a few simple grouping cues, one can now produce rather impressive segmentation on a large set of images. Firstly Principle component analysis(PCA) is applied on medical image and then clustering technique is applied to segment the region of interest from medical image. The number of clusters has to be defined in advanced so PCA is applied first. Using PCA, data is mapped into new feature space, K-mean algorithm is applied to data in feature space. The final objective is, to distinguish the different clusters in a better way. PCA is a stastical technique for unsupervised dimension reduction .PCA is used for visualization of complex data developed to capture as much of variation in data as possible and identify the patterns. After PCA, K-mean algorithm clusters data by trying to separate samples in n cluster of equal variance, minimizing criteria known as inertia.

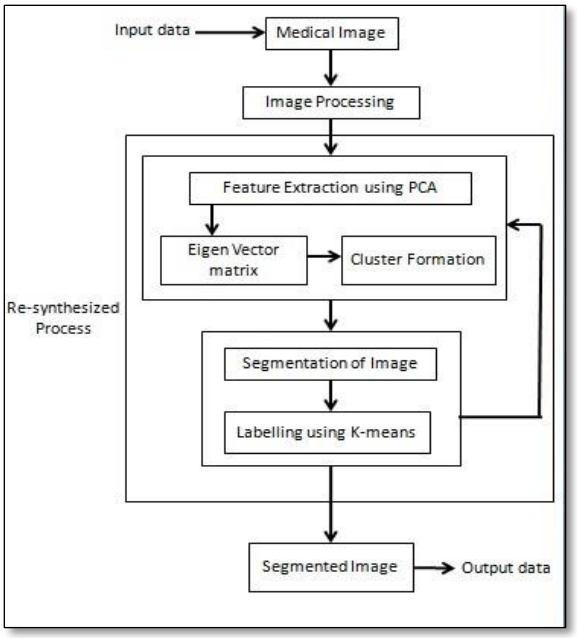

Figure 3: Proposed Architecture

A. *Proposed method steps*

Following are proposed method steps:

Step 1:

The image is split into blocks.

Step 2:

PCA is applied on all the blocks and the data set is reduced. This data set can be taken as the one which represents the whole data with the textures.

Step 3:

These values are taken and labeled using a K-means clustering algorithm

Step 4:

The image is reconstructed using the labels of these blocks. The blocks are colored depending on the label and the image contains these colors as blocks

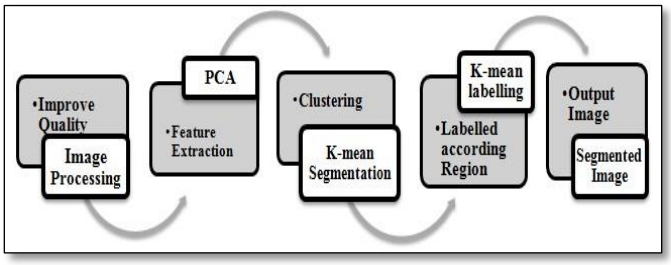

Figure 4: Steps of proposed Methodology

1) *Image Processing:*

Image Processing is first step, image processing is use for improvement of image data, image quality is improved. Enhances some image features important for further processing. Aims to correct some degradation in image such as brightness transformation is used to modify pixel brightness, grayscale transformation is used to change the

transformation without regard to position. Noise removal is done using low-pass filtering and high-pass filtering. Mathematical operation such as arithmetic operation and morphological operations are used to eliminate clustering background and smoothening purpose.

2) Principle Component analysis:

Principle component analysis is second step, it is use for feature extraction of image, visualization of complex data, capture the variation of data. Principle component analysis is to find number of cluster require for segmentation it find out number of precise clusters because large number of clusters loss the data and small number of clusters may loss region of interest(ROI) so need to find appropriate number of clusters**,** this value of appropriate number of cluster are given to k-mean for segmentation.

a) *Principle Component Analysis algorithm:*

i) Get data:

Take all pixels of image and represent it in matrix form. This is done by taking the first pixel from all the images arranging it in a column [3].

ii) Subtract mean:

Mean is calculated as follows [3]:

$$\textbf{Mean} = \frac{\textbf{Total pixel values}}{\textbf{No. of pixels}}$$

Take one column of the matrix and subtract from each pixel value, the mean value of all the pixels in that column. It shows the pixel variation from centre (mean) that is variance one dimensional but image data is two dimensional so find co-variance.

*iii) Calculate co-variance matrix***:**

following equation is use to calculate co-variance matrix [3]:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \quad (1)$$

Calculate co-variance matrix represent it in form of matrix. It shows variation of pixels from mean pixel to show spread of data from mean.

*iv) Calculate Eigen Vector***:**

Eigen vector tell us about pattern in data, Eigen vector are perpendicular to each other. Draw Eigen vector line through middle of points. Eigen vector shows us how dataset are related along the line. Process of taking Eigen vectors of covariance matrix we have been able to extract the lines that characterize the data [3].

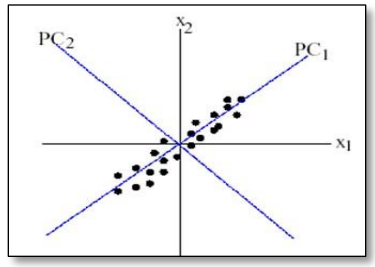

Figure 5: Principle Component

Eigen vector with highest Eigen value is 1st principle component (PC_1) [3].

v) Choosing the Components and forming feature vector:

Once Eigen vectors are found from co-variance matrix next step is to order them by Eigen value highest to lowest. This gives principle component with higher significant so we can ignore component with lower significance. That we have reduced the unwanted data and focus only on region of interest (ROI) [3].

Feature vector=(Eig_1,Eig_2,Eig_3.....Eig_n)

3) K-mean Segmentation:

K-mean algorithm take the number of appropriate cluster from PCA and segment the image according to number of clusters. K-mean clustering is unsupervised classification method of grouping set of data object into clusters and extracting region of interest. Highly depend on initialization of centroids as result computation is often done several times with different initialization of centroids. K-mean segmentation identify separate objects within the images, finds regions of connected pixels with similar properties, finds boundaries between regions and removes unwanted regions.

4) K-mean Clustering Algorithm:

Step 1 Compute the distribution of the intensity values.

Step 2 Using k random intensities initialize the centroids.

Step 3 Repeat the step 4 and step 5 until the labels of the cluster do not change any more.

Step 4 Cluster the image points based on the distance of their intensity values from the centroid intensity values .Following equation is use to calculate centroid [4]:

$$c^{(i)} := \arg\min_j \left\| x^{(i)} - \mu_j \right\|^2 \quad (2)$$

Step 5 Compute new centroid for each cluster [4].

$$\mu_i := \frac{\sum_{i=1}^{m} 1\{c_{(i)}=j\}x^{(i)}}{\sum_{i=1}^{m} 1\{c_{(i)}=j\}} \quad (3)$$

Where *k* is the number of clusters, *i* iterates over all the intensity values, *j* iterates over all the centroids (for each cluster) and μ_i are the centroid intensities

4)K-*mean labeling*:

Labeling the pixels of an image according to semantic context. Label according to region of clusters higher the region higher labeling [4].K-mean is efficient method than Fuzzy c-mean clustering method [4][5].

***5) Output Image*:**

Output image is segmented image. Segmented image shows region of interest (ROI). Thus meaningful part is extracted from medical image using Principle component analysis (PCA) and K-mean Clustering technique.

V. RESULT ANALYSIS:

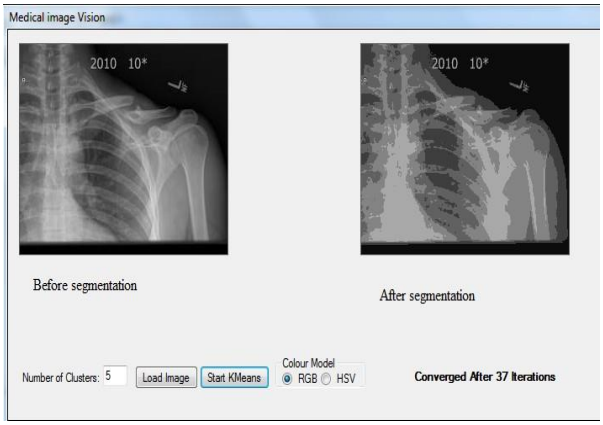

Figure 6: Result showing segmentation of medical image

We have applied proposed method to find out region of interest. Number of cluster and iteration are found efficiently and image is segmented. Above figure shows result of segmentation, it shows result in number of clusters and number of iteration.

Calculated Number of cluster for given medical image=5

Calculated Number of iteration for given medical image=37

Following table shows result of different number of clusters and its iteration according to number of cluster:

TABLE 3:Result analysis

Number of clusters	Number of iteration
3	13
5	37
7	44
9	50

As number of clusters increases, number of iteration increases. If number of clusters are maximum then region of interest disappears and if number of clusters are minimum then region of interest is not extracted. So, There is a need to find precise clusters to segment the image. Number of clusters has to be defined in advanced and then k-mean should be applied. So we use PCA before k-mean to find out precise number of clusters. PCA performs better representation of data without losing much of information.

VI. FUTURE WORK

We have proposed medical image segmentation using principle component analysis and K-means clustering method to extract meaningful part from image. Principle component analysis is used to find out precise number of clusters that is number of patterns, k-means clustering is use to group the pixels according to patterns and segment the image to extract meaningful region. The proposed method is observed to preserve patterns and details better than other competing techniques.

In future we would like to concentrate on compression ratio, execution time, signal to noise ratio.

VII. CONCLUSION

Until now, we have implemented K-mean clustering for segmentation. Later on we will implement PCA before k-mean to find out precise clusters. PCA performs image segmentation based on patterns of valueS in image. This pattern representS texture. The whole proposed method depends upon clustering algorithm'S use. K-means clustering algorithm is used for segmentation which is efficient method. Using texture re-synthesis and PCA together produces a good segmented image.

REFERENCES

1. Prof. Dinesh D. Patil , "*Medical Image Segmentation*", International Journal of Computer Science and Mobile Computing , January- 2013,Vol.2 Issue 1.
2. Jay Patel and Kaushal Doshi ,"Study of Segmentation Method for Detection of tumor in brain MRI",ISSN 2231-1297,2014,volume 4 pp 279-284.
3. Inamdar, R.S.; Ramdasi, D.S., "Active appearance models for segmentation of cardiac MRI data," Communications and Signal Processing (ICCSP), 2013 International Conference on , vol., no., pp.96,100, 3-5 April 2013
4. Hooda, H.; Verma, O.P.; Singhal, T., "Brain tumor segmentation: A performance analysis using K-Means, Fuzzy C-Means and Region growing algorithm," Advanced Communication Control and Computing Technologies (ICACCCT), 2014 International Conference on , vol., no., pp.1621,1626, 8-10 May 2014.
5. Shasidhar, M.; Raja, V.S.; Kumar, B.V., "MRI Brain Image Segmentation Using Modified Fuzzy C-Means Clustering Algorithm," Communication Systems and Network

Technologies (CSNT), 2011 International Conference on , vol., no., pp.473,478, 3-5 June 2011

6. Selvakumar, J.; Lakshmi, A.; Arivoli, T., "Brain tumor segmentation and its area calculation in brain MR images using K-mean clustering and Fuzzy C-mean algorithm," Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on , vol., no., pp.186,190, 30-31 March 2012
7. Acharya, J.C.; Gadhiya, S.A.; Raviya, K.S., "Objective assesment of different segmentation algorithm for underwater images," Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on , vol., no., pp.1,7, 4-6 July 2013