An Analysis of K Means Clustering Algorithm for Image Segmentation with IQI

Ramaraj.M[1] Dr.S.Niramathi[2]
[1]Research Scholar [2]Assistant Professor
[1,2]Department of Computer Science and Engineering
[1,2]NGM College Pollachi India,

Abstract— Image mining is used to discover the knowledge from the image dataset. Similar to data mining tasks, image mining task can also be categorized into classification, clustering, association rule mining, and characterization based summarization. Many researches have been done in the area of image segmentation using clustering. There are different methods and one of the most popular methods is k-means clustering algorithm. K-means clustering algorithm is an unsupervised algorithm and it is used to segment the interest area from the background. In this paper we discuss above the analysis of k means clustering algorithm for image segmentation process using different clustering methods have been applying for images to the image dataset.

Key words: Data Mining, Clustering Techniques, Image Segmentation, Image Dataset

I. INTRODUCTION

Image segmentation is an important technology for image processing. There are many applications whether on synthesis of the objects or computer graphic images require precise segmentation. With the consideration of the characteristics of each object composing images in MPEG4, object-based segmentation cannot be ignored. Nowadays, sports programs are among the most popular programs, and there is no doubt that viewers' interest is concentrated on the athletes. Therefore, demand for image segmentation of sport scenes is very high in terms of both visual compression and image handling using extracted athletes.

One of most used clustering algorithm is *k*-means clustering. It is simple and computationally faster than the hierarchical clustering. And it can also work for large number of variable. But it produces different cluster result for different number of number of cluster. So it is required to initialize the proper number of number of cluster, *k*2. Again, it is required to initialize the *k* number of centroid. Different value of initial centroid would result different cluster. So selection of proper initial centroid is also an important task [4]. Nowadays image segmentation becomes one of important tool in medical area where it is used to extract or region of interest from the background. So medical images are segmented using different technique and process outputs are used for the further analysis in medical. But medical images in their raw form are represented by the arrays of numbers in the computer3, Image mining is still at the experimental stage and growing filed for research. Lack of understanding the research issues of image mining is the obstacle to rapid progress. Image mining is not just the expansion of data mining to image domain. It can be considered to be an efficient hybridization of image processing and data mining concepts to extract the useful knowledge [6]. Various application domains of image mining includes natural scene recognition, remote sensing, Egeria detection, weather forecasting, criminal investigation, image segmentation, etc.

II. RELATED WORK

There are many authors describes the image segmentation process, T. Y. Gajjar et al 2012 [11] It can be used to group the images on web, efficient retrieval of images, or to extract hidden meaningful information from image datasets which is not explicitly available from image sources. Pankaj Agrawal et al this process can help to find out the best suitable value of parameters for the segmentation of different types of imagery. In this paper, one best algorithm has considered for each method of image segmentation. The interactive based method provides the facility to select the desired area as an object and produces better result. The proposed process also displays the duration of segmentation of each algorithm. A.Kannan et al 2010 [10] Mining Image data is the one of the essential features in this present scenario since image data plays vital role in every aspect of the system such as business for marketing, hospital for surgery, engineering for construction, Web for publication and so on. Image mining normally deals with the extraction of implicit knowledge, image data relationship, or other patters not explicitly stored from the low-level computer vision and image processing techniques. i.e.) the focus of image mining is the in the extraction of patterns from a large collection of images. Hamed Shamsi et al 2012 [8]a new modify spatial FCM that incorporates the spatial information into the membership function to improve the segmentation results. In the new spatial function we used two contribution factors. The first one was according to distances between central pixels with neighbor pixels. The second factor was calculated according to value difference of central pixel with neighbor pixels. Using of these contribution factors caused that spatial function is made of according to distance and value pixels. PRABHJOT KAUR et al Fuzzy clustering can be applied to a wide variety of applications like image segmentation, pattern recognition, object recognition, and customer segmentation etc. The clustering output depends upon various parameters like distribution of points inside and outside the cluster, shape of the cluster and linear or non-linear separability. A.Hema et al 2013 [12] This increase in number of images and image databases has necessiated the need for image mining techniques. Image mining is an extended branch of data mining that is concerned with the process of knowledge discovery concerning digital images. The main aim of this paper is to present a survey of the various techniques used for image mining applications like image retrieval, Matching, Pattern recognition given by different researchers. A.Rajendran et al 2011 [1] a method that combine region based fuzzy clustering and deformable model for segmenting tumor region on MRI images. Region based

fuzzy clustering is used for initial segmentation of tumor then result of this is used to provide initial contour for deformable model, which then determines the final contour for exact tumor boundary for final segmentation using gradient vector field as a external force field. S. Patel et al 2012 [2] The analysis of clustering algorithms for medical MR (magnetic resonance) images using IQI (image quality index) is done. The clustering algorithms used are classical C means, fuzzy C means and rough fuzzy C means. With the assistance of the lower and upper approximation of rough sets, the rough fuzzy C means clustering algorithm improves the objective function and further the distribution of membership function for the traditional fuzzy C means clustering. Zhensong Chen 2015 Image segmentation problem is a fundamental task and process in computer vision and image processing applications. Eman Abdel-Maksoud 2015 [6] Image segmentation refers to the process of partitioning an image into mutually exclusive regions. an efficient image segmentation approach using K-means clustering technique integrated with Fuzzy C-means algorithm. It is followed by thresholding and level set segmentation stages to provide an accurate brain tumor detection [7].

III. METHODOLOGY (CLUSTERING METHODS)

It's the on the techniques for the clustering concept in the data mining process and is very famous algorithm for the K-means clustering, because it is similar or simpler and easier incomputat ion of an efficient K-means clustering algorithm it is the simplest unsupervised learning algorithms that solve the well-known clustering problems.

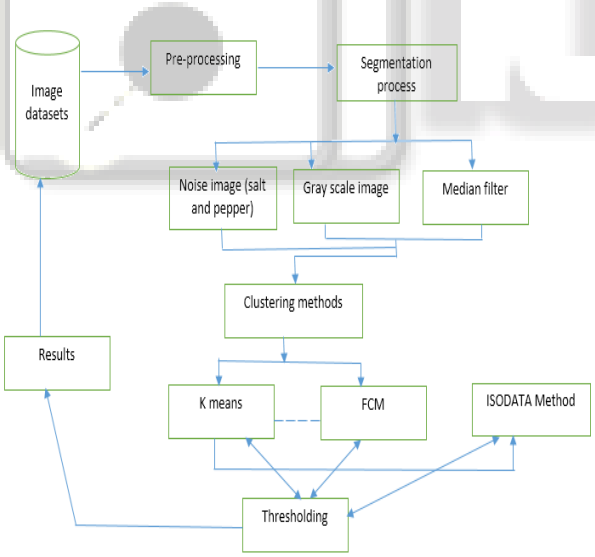

Fig. 1: block diagram for image segmentation

Figure 1 is represent the process of image and store the result of the image value in the data base and fellow the process of block diagram as pre-processing, segmentation process, methods and result of the image to be stored on data base.

IV. *K*-MEANS CLUSTERING ALGORITHM

Clustering is a method to divide a set of data into a specific number of groups. It's one of the popular method is *k*-means clustering. In *k*-means clustering, it partitions a collection of data into a *k* number group of data. It classifies a given set of data into *k* number of disjoint cluster. *K*-means algorithm consists of two separate phases. In the first phase it calculates the *k* centroid and in the second phase it takes each point to the cluster which has nearest centroid from the respective data point.

```
kmeans_clustering( )
  while δ/N > threshold
    δ ← 0
    for i ← 0 to N-1
      for j ← 0 to K-1
        distance ← | objects[i] - clusters[j] |
        if distance < dmin
          dmin ← distance
          n ← j
      if membership[i] ≠ n
        δ ← δ + 1
        membership[i] ← n
      new_clusters[n] ← new_clusters[n] + objects[i]
      new_cluster_size[n] ← new_cluster_size[n] + 1
    for j ← 0 to K-1
      clusters[j][*] ← new_clusters[j][*] / new_cluster_size[j]
      new_clusters[j][*] ← 0
      new_cluster_size[j] ← 0
```

1) Initialize number of cluster *k* and centre.
2) For each pixel of an image, calculate the Euclidean distance *d*, between the center and each pixel of an image using the relation given below.
3) Assign all the pixels to the nearest centre based on distance *d*.
4) After all pixels have been assigned, recalculate new position of the centre using the relation given below.
5) Repeat the process until it satisfies the tolerance or error value.
6) Reshape the cluster pixels into image.

Although *k*-means has the great advantage of being easy to implement, it has some drawbacks. The quality of the final clustering results is depends on the arbitrary selection of initial centroid. So if the initial centroid is randomly chosen, it will get different result for different initial centers. So the initial center will be carefully chosen so that we get our desire segmentation. And also computational complexity is another term which we need to consider while designing the *K*-means clustering. It relies on the number of data elements, number of clusters and number of iteration [3].

A. Median Filter:

Median filtering is used as a noise removal in order to obtain a noise free image. After segmentation is done, the segmented image may still present some unwanted regions or noise. So to make the image a good and better quality, the median filter is applied to the segmented image. We can use different neighborhood of $n \times n$. But generally neighborhood of $n = 7$ is used because large neighborhoods produce more severe smoothing.

There are some medical image segmentation systems which use K-means algorithm, K-means algorithm is fast and simple to run on large datasets, other systems use Fuzzy C-means algorithm because it retains the more

information of the original image to detect malignant tumor cells accurately compared to the K-means. These systems are sensitive to noise and outliers, and they take long execution time [5].

B. IQI Measures:

Image quality index (IQI) is used on the image to improve the quality of image to the image segmentation and main purpose IQI, which is easy to calculate and applicable to various image processing applications. Instead of using traditional error summation methods, the proposed index is designed by modeling any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. Although the new index is mathematically defined and no human visual system model is explicitly employed, experiments on various image distortion types show that it exhibits surprising consistency with subjective quality measurement. It performs significantly better than the widely used distortion metric mean squared error.

V. FUZZY C-MEANS ALGORITHM

Traditional clustering approaches generate partitions where each pattern belongs to one, and only one, cluster. Hence, the clusters in a hard partition are disjoints. Fuzzy clustering extends this notion to associate each pattern to every cluster using a membership function. The Fuzzy c-Means clustering algorithm (FCM) was initially developed by Dunn [9], and generalized later by Bezdek in. This algorithm is based on optimization of the objective function, given by the following equation:

$$j_{fcm}=\sum_{i-1}^{c}\sum_{k=1}^{n}(\mu_{ik})^{m}\|z_k-v_i\|^2$$

where the membership matrix, U = [μik] ∈ Mfmc , is a fuzzy partition of Z, V = [v1, v2, . . . , vc] is the vector of prototypes of the clusters, which are calculated according to $D_{ikAi}=\|z_k-v_i\|^2$, a square inner-product distance norm, and m ∈ [1,∞]is a weighting exponent that determines the fuzziness of resulting clusters. The optimal partition U∗ of Z for a FCM algorithm is reached through the couple (U∗, V∗), which minimizes (locally) the objective function Jfmc according to the Alternating Optimization (AO).

VI. ISODATA METHOD

The ISODATA algorithm is similar to the k-means algorithm with the distinct difference that the ISODATA algorithm allows for different number of clusters while the k-means assumes that the number of clusters is known a priori. K-means (just as the ISODATA algorithm) is very sensitive to initial starting values. For two classifications with different initial values and resulting different classification one could choose the classification with the smallest MSE (since this is the objective function to be minimized). However, as we show later, for two different initial values the differences in respects to the MSE are often very small while the classifications are very different. Visually it is often not clear that the classification with the smaller MSE is truly the better classification.

VII. RESULT

A. Data set:

In order to check the performance of our image segmentation approach, we used three benchmark data sets. The data sets are collect from the web page and store the data bases on image and the different types of size of images and different images are collect them. Knee image, brain tumor, nature images and etc….. all the images are different format as .gif, .jpg, .png, .trf, and etc….

B. Implementation Process:

In this section, we show the result of the image segmentation process to apply for matlab tool has to be implemented. The version of tool is 8.1 and corei3 processor, graphics card on nvidia and support for other system facilities as to use.

The results of image segmentation using partitional clustering algorithms are shown to be a good alternative to improving the detection of MCs in digitized mammograms. On the other hand, k-Means and FCM were very helpful in carrying out a previous stage, applying image enhancement using tophat morphological transform and using only one feature, gray level intensity, and we also obtained good results. As the final conclusion to our work, we can say that the results for k-Means and FCM depend on several factors, such as the stage of image enhancement and the initial number of partitions in the algorithms. For K Means, the obtained results depend only on the threshold value being the most appropriate, since the parameters of this algorithm remain constant.

DATA SETS	IQI (brain)	IQI (knee)	IQI(breast image)
K MEANS	2.285714	2.714286	3.571429
FCM	3.047619	3.619048	4.761905
ISODATA	6.333333	8.333333	9.333333

Table 1: IQI value for the image segmentation of three data sets (brain, knee, breast image)

Table1 is represent by use of three algorithms in the given data set for brain knee and breast image data set and find the simple and easy to calculate the image quality index value of image and to display the different value will produce the three algorithms values in the table 1.

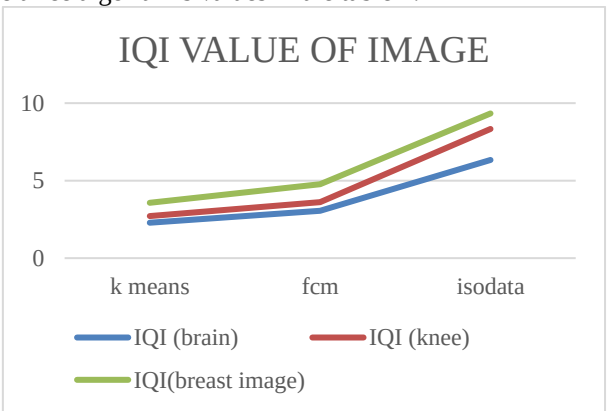

Fig. 2: overall Image Quality Index Value of image

Figure 1 is represented by the overall IQI value of the segmented images such as three different types of datasets and various algorithms result will be produced.

VIII. CONCLUSION

In this paper described to analysis of k means clustering algorithm for image segmentation. We have segmented an image by using k-clustering algorithm, it is using subtractive cluster to generate the initial centroid. At the same time is used to improve the quality of original image and median filter, IQI value is used to improve segmented image. And the final segmented result is compare with k-means clustering algorithm and FCM clustering algorithm and we can conclude that the proposed clustering algorithm has better segmentation. Implement with ISODATA clustering method to analyze the different areas and different measurement to apply for image segmentation. In the future, we can improve the quality of the output image more by using the morphological operation and get better performance measurement.

REFERENCES

[1] A.Rajendran R. Dhanasekaran" Fuzzy Clustering and Deformable Model for Tumor Segmentation on MRI Brain Image: A Combined Approach "ICCTSD- 2011 Procedia Engineering 30 (2012) 327 – 333.
[2] S. Patel, K.S.Patnaik "Analysis Of Clustering Algorithms for MR Image Segmentation Using IQI" ICCCS-2012-Procedia Technology 6 (2012) 387 – 396.
[3] V.Kalist, Ganesan P, B.S.Sathish , J.Merlin Mary Jenitha, Khamar Basha.shaik "Possiblistic-Fuzzy C-Means Clustering Approach for the Segmentation of Satellite Images in HSL Color Space" ICRTC-2015 Procedia Computer Science 57 (2015) 49 – 56.
[4] Nameirakpam Dhanachandra, Khumanthem Manglem and Yambem Jina Chanu "Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm " IMCIP-2015 Procedia Computer Science 54 (2015) 764 – 771.
[5] A.R. Kavitha , C. Chellamuthu "Implementation of Gray-level Clustering Algorithm for Image
[6] Segmentation" Procedia Computer Science 2 (2010) 314–320 ICEBT 2010.
[7] Eman Abdel-Maksoud, Mohammed Elmogy , Rashid Al-Awadi "Brain tumor segmentation based on a hybrid clustering technique" Egyptian Informatics Journal (2015) 16, 71–81
[8] Biju V.G, Mythili. P" Fuzzy clustering algorithms for cDNA microarray image spots segmentation " ICICT 2014 Procedia Computer Science 46 (2015) 417 – 424.
[9] Hamed Shamsi and Hadi Seyedarabi "A Modified Fuzzy C-Means Clustering with Spatial Information for Image Segmentation" International Journal of Computer Theory and Engineering, Vol. 4, No. 5, October 2012.
[10] Prabhjot Kaur Dr. A. K. Soni Dr. Anjana Gosain "Novel Intuitionistic Fuzzy C-Means Clustering for Linearly and Nonlinearly Separable Data" WSEAS TRANSACTIONS on COMPUTERS Issue 3, Volume 11, March 2012 E-ISSN: 2224-2872.
[11] A.Kannan, Dr.V.Mohan, Dr.N.Anbazhagan "Image Clustering and Retrieval using Image Mining Techniques" IEEE International Conference on Computational Intelligence and Computing Research 2010.
[12] T. Y. Gajjar, N. C. Chauhan" A Review on Image Mining Frameworks and Techniques" T. Y. Gajjar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,4064-4066.
[13] A.Hema, E.Annasaro "A SURVEY IN NEED OF IMAGE MINING TECHNIQUES" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 2, February 2013 ISSN (Print) : 2319-5940 ISSN (Online) : 2278-1021.

