ISSN 2319-7080
International Journal of Computer Science and Communication Engineering
IJCSCE Special issue on "Emerging Trends in Engineering & Management" ICETE 2013

A Modified Image Segmentation using K-means Clustering algorithm for Content Based Medical Images

Er. Surbhi Ghai[1],Dr.K.S.Mann[2]
[1]Research Scholar, Department of Computer Science Engineering, G.N.D.E.C Ludhiana,Punjab, India
[2]Associate Professor, Department of Computer Science Engineering & IT, G.N.D.E.C Ludhiana, Punjab,India
[1]ghai.surbhi@gmail.com,[2]mannkulvinder@yahoo.com

Abstract: **This paper gives a new k-means clustering algorithm for medical images using image segmentation. A new algorithm has been developed on the system of patient information. The model will be an attempt to reduce the semantic gap between the information retrieved and information stored in technical terms. In order to achieve the above said functionality a classification scheme is required based on which the identification of similar medical images can be searched and retrieved by the system. The classification scheme helps in reducing the semantic gap which in turns ensures faster image retrieved from the data source. The presented work is about the content retrieval in MRI medical images using K-means clustering algorithm.**

I .INTRODUCTION

A. Content Based Medical Image Retrieval (CBMIR)

With the development of the Internet, and the availability of image capturing devices such as digital cameras, image scanners, the size of digital image collection is increasing rapidly. Efficient image searching, browsing and retrieval tools are required by users from various domains, including remote sensing, fashion, crime prevention, publishing, medicine, architecture, etc. For this purpose, many general purpose image retrieval systems have been developed. There are two frameworks: text-based and content-based.[1].The commercial QBIC system is definitely the most well-known system. Another commercial system for image and video retrieval is Virage that has well known commercial customers such as CNN. Most of the available systems are, however from academia. It would be hard to name or compare them all but some well-known examples include Candid, Photo book and Netra that all use simple colour and texture characteristics to describe the image content.[2]

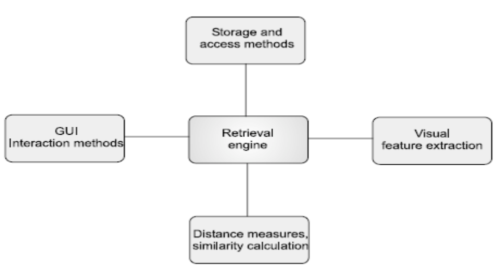

Fig. 1 The principal components of all content-based image retrieval systems.[2]

Visual Features Used

- Colour
- Texture
- Local and global features
- Segmentation and Shape Features

B. Image Segmentation

Image segmentation is an essential preliminary step in most automatic pictorial pattern-recognition and scene analysis problems. Segmentation subdivides an image into its constituent regions or objects, which ideally correspond to different real-world objects. The level to which subdivision is carried out depends upon the problem being solved. When an object of interest in an application has been isolated, segmentation should be stopped. Segmented images are now used routinely in a multitude of different applications, such as, diagnosis, treatment planning, in the robotics, localization of pathology, geology, study of anatomical structure, meteorology, computer-integrated surgery, among others.

1) *Segmentation Process:* Segmentation is the process of partitioning an image into regions i.e. group of connected pixels with similar properties such as gray levels, colors, textures, motion characteristics (motion vectors), edge continuity[6],[7]. There are two approaches to segmentation:

- Region segmentation
- Edge segmentation

.

C. Clustering

Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A cluster is therefore a collection of objects which are similar between them and are dissimilar to the objects belonging to other clusters.

1) *Distance-based clustering:* Two or more objects belong to the same cluster if they are "close" according to a given distance.
2) *Conceptual clustering:* Two or more objects belong to the same cluster if this one defines a concept common

to all that objects. In other words, objects are grouped according to their fit to descriptive concepts, not according to simple similarity measures.

3) *Requirements*: The main requirements that a clustering algorithm should satisfy are:

- scalability
- dealing with different types of attributes
- discovering clusters with arbitrary shape
- minimal requirements for domain knowledge to determine input parameters
- ability to deal with noise and outliers
- insensitivity to order of input records
- high dimensionality
- Interpretability and usability.

4) Problems: There are a number of problems with clustering. Among them:

- current clustering techniques do not address all the requirements adequately (and concurrently)
- dealing with large number of dimensions and large number of data items can be problematic because of time complexity
- the effectiveness of the method depends on the definition of "distance" (for distance-based clustering)
- if an obvious distance measure doesn't exist we must "define" it, which is not always easy, especially in multi-dimensional spaces
- The result of the clustering algorithm (that in many cases can be arbitrary itself) can be interpreted in different ways [8],[9].

II. K-MEANS CLUSTERING

K-means is one of the most popular methods of hard clustering. The standard k-means clustering algorithm does not take into account the spatial constraints of the image. Also it performs clustering by assuming a constant intensity distribution for the pixels within a cluster. That is it does not take into consideration of the intensity variations of pixels within the clusters. The success of this approach lies in the estimation and selection of an appropriate window size for the intensity calculation.[3] Another important criterion for the k-means clustering method is the selection of the optimal feature used for the clustering. It makes the algorithm more reliable. The ambiguity with this feature selection is dealt with in this proposed method that makes the process of feature selection automatic It has been found to be useful for noisy images like medical images where the intensity variations are also found to be very smooth.[4] The steps for a standard k-means clustering algorithm are as follows.

1. Consider a set of n data points (feature vectors) to be clustered. Here the data points are the image pixels.
2. Assume the no: of clusters as k, where 2≤ k < n.
3. Randomly select k initial cluster center locations.
4. All data points are assigned to a partition defined by nearest cluster centers. This is determined by using a distance measure to check the closeness of the pixels to the chosen cluster centers.
5. After the partitioning, the cluster centers are moved to the geometric centroid of their data points in their respective partitions.
6. Repeat steps 4 & 5 until the overall objective function is smaller than a given tolerance, or until the calculated cluster centers don't move to new points. The aim of k-means is to minimize an objective function, which can be a squared error function, as considered here. It depends on what distance metric chosen for clustering. The objective function for the case considered is,

$$J = \sum_{j=1}^{k} \sum_{i=1}^{m} \left\| x_i^{(j)} - c_j \right\|^2$$

Where $\left\| x_i^{(j)} - c_j \right\|$ is a chosen distance measure between a set of data points $x_i^{(j)}$ and the cluster centers c_j . It is an indicator of the distance of the m data points from their respective cluster centers and k is the number of clusters chosen.

III. OUR ALGORITHM

A. Proposed Method of segmentation:

1. In the first step, the mean of feature vector is computed.

The mean value of the image is calculated using the 3-tuple-feature vector formed as before. . Calculation is done as follows:

$$m^{(k)} = (1/n_k)\sum_{i=1}^{n_k} x_i^{(k)}$$

where, m^k are the mean values for the k clusters or segments. , x_i are either the intensity values or the location values of the image pixels within a cluster, n_k are the number of pixels within the k clusters considered.

2. Next, the splitting of mean is done to obtain the initial seed vectors or cluster centers for clustering using Binary split vector quantization.
3. Perform K-means clustering on the image pixels
4. This step includes the recalculation of the cluster centers for the newly formed clusters.
5. Repeat steps 2,3 & 4 until the stopping criteria have reached. The following stopping criteria used in this paper.

a. The no. of pixels in each cluster formed < $N_{threshold}$

Let the no: of pixels in the image be N. Then the threshold value for the

Number of pixels is taken as $N/2^i$. Where, i is the step number of clustering.

If the no: of pixels within a cluster formed is less than this value, the segmentation can be stopped, as the clusters will become insignificant with number of pixels lesser than the threshold values.

b. The variance of the pixel values in a cluster < $V_{threshold}$

The variance of the intensity values within a cluster is also used as stopping criterion. Once the variance value becomes less than this, segmentation can be stopped to avoid the formation of insignificant clusters. For an 8-bit image the value is found to be in the range 0.00001 to 0.0001. Beyond these values it leads to insignificant cluster formation. Changing the values of the variance threshold will result in varying of the number of segments formed for the image.

6. Labelling of the Clusters is performed in this final step

B. Proposed Algorithm

1. Convert the image to Greyscale.
2. Perform the image enhancement using pre-processing Tools to Normalize the Image.
3. Initialize the population respective to the Genetic algorithm.
4. Perform the Image Segmentation using Fuzzy C Means.
5. Define the initial Fitness Function.
6. For i=1 to Max Iteration
   [Repeat Steps 7 to 11]
7. Perform the Selection on this Training Dataset.
8. Perform the Crossover on selected parents and generate the next level child.
9. Perform the Mutation to neglect the values that does not support fitness function.
10. Recombine the generated child with existing population to generate new Population Set.
11. Apply the Fuzzy C Means on this new population Set.
12. Generate the Mean of this Image.
13. Compare the image pixel with this obtained clustered Threshold value, and derive the result image.
14. Present the tumor detected image.
   }

As we can see the algorithm is divided in three main phases.

a) Preprocessing
b) Clustering
c) Genetics based classification

IV. EXPERIMENTAL RESULTS

Original Image

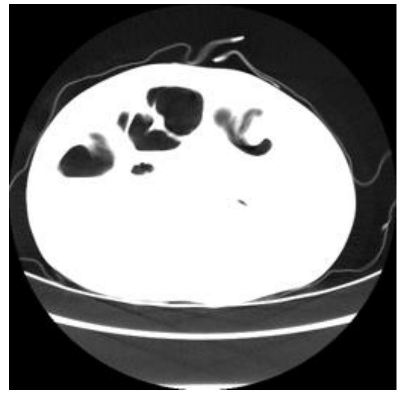

Fig 2.Original Medical Image

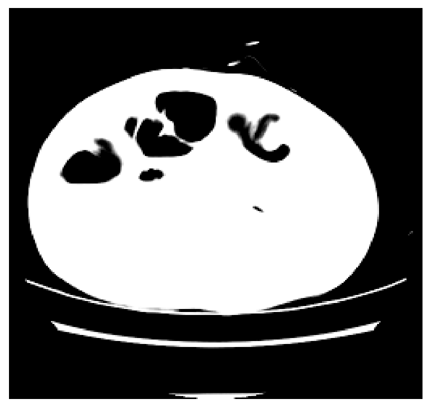

Fig 3. Extracted image

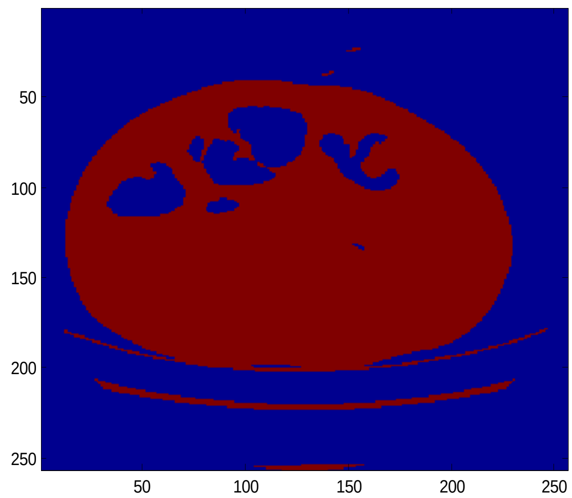

Fig 4. Colored Medical Image

V. CONCLUSIONS & FUTURE SCOPE

In this work the soft computing approach is been implemented to detect the tumor from the medical brain image. The wok can be extended in different ways.

- In this work, the tumour detection is performed on 2 D images, the work can be implanted for the 3D images.
- The enhancement in the work can be done by performing the implementation of sequence images.
- The work can be performed on some other image formats
- The variation can be performed in terms of clustering algorithm as well the stages of the genetic approach.

REFERENCES

[1] Ying Liu, Dengsheng Zhang, Guojun Lu,Wei-Ying Ma(2007),“A survey of content based image retrieval with high level semantics ”,The Journal of the pattern recognition society, vol. 40,pp. 262-282

[2] Henning Müller, Nicolas Michoux, David Bandon, Antoine Geissbuhler (2003) , “ A review of content-based image retrieval systems in medical applications——clinical benefits and future directions” International Journal of Medical Informatics (2004) 73, 1—23.

[3] Armato, S G., M B. Altman, and P J. La Riviere. "Automated Detection of Lung Nodules in CT Scans: Effect of Image Reconstruction Algorithm." Medical Physics 30 (2003): 461 472.

[4] Aggarwal Preeti, Sardana H.K., Vig R(2012). Journal of Advances in Information Technology, Volume 3, Number 4, ISSN 1798-2340

[5]Garima Garg and Sonia Juneja. Brain Tumor Segmentation using Genetic Algorithm and FCM Clustering Approach. International Journal of Computer Applications 49(2):24-27, July 2012. Foundation of Computer Science, New York, USA.

[6]N. Nandha Gopal , Dr. M. Karnan, “Diagnose Brain Tumor Through MRI Using Image Processing Clustering Algorithms Such As Fuzzy C Means Along With Intelligent Optimization Techniques”, 978-1-4244-5967-4/10 ©2010 IEEE.

[7] Armato, S G., M B. Altman, and P J. La Riviere. "Automated Detection of Lung Nodules in CT Scans: Effect of Image Reconstruction Algorithm." Medical Physics 30 (2003): 461 472.

[8]Aggarwal Preeti, Sardana H.K., Vig R(2012). Journal of Advances in Information Technology, Volume 3, Number 4, ISSN 1798-2340.

[9] Pham et al. [2010] current methods in medical image segmentation.Annual Review of Biomedical Engineering Vol. 2: 315-337.

[10] Ehab F. Badran, Esraa Galal Mahmoud, and Nadder Hamdy (2010). An algorithm for detecting brain tumors in MRI images. Computer Engineering and Systems (ICCES), 2010 International Conference, pp: 368 – 373, ISBN: 978-1-4244-7040-2, Cairo.