

Detection of Tumors in MRI Images Using Genetic and Firefly C Mean and K Mean Clustering

Bandi Nissy¹, P. MallikarjunaRao²

¹Department of Electronics and Communication Engineering, Andhra University college of Engineering, Visakhapatnam, Andhra Pradesh, India

²Professor, Andhra University College of Engineering, Visakhapatnam, Andhra Pradesh, India

Abstract - Brain tumour extraction and its analysis are challenging tasks in medical image processing because brain image and its structure are complicated that can be analysed only by expert radiologists. Segmentation plays an important role in the processing of medical images. MRI (magnetic resonance imaging) has become a particularly useful medical diagnostic tool for diagnosis of brain and other medical images. This paper presents a comparative study of three segmentation methods implemented for tumour detection. The methods include k-means clustering with watershed segmentation algorithm, optimized k-means clustering with genetic algorithm and [1][3] optimized c- means clustering with genetic algorithm. Traditional k-means algorithm is sensitive to the initial cluster centres. [3] Genetic c-means and k-means clustering techniques are used to detect tumour in MRI of brain images. At the end of process, the tumour is extracted from the MR image and its [2] exact position and the shape are determined. The experimental results indicate that genetic c-means not only eliminate the over segmentation problem, but also provide fast and efficient clustering results.

Index Terms - MRI Images, Genetic, Firefly, C Mean Cluster, K Mean Cluster

1.INTRODUCTION

1.1 Image Segmentation

In image segmentation, one challenge is how to deal with the nonlinearity of real data distribution, which often makes segmentation methods need more human interactions and make unsatisfied segmentation results. Medical image segmentation plays an instrumental role in clinical diagnosis. An ideal medical image segmentation scheme should possess some preferred properties such as minimum user interaction, fast computation, and accurate and robust segmentation results.

Image segmentation is an image analysis process that aims at partitioning an image into several regions according to a homogeneity criterion. Image segmentation is a very complex task, which benefits from computer assistance, and yet no general algorithm exists. It has been a research field in computer science for more than 40 years now, and the early hope to find general algorithms that would achieve perfect segmentations independently from the type of input data has been replaced by the active development of a wide range of very specialized techniques. Most of the existing segmentation algorithms are highly specific to a certain type of data, and some research is pursued to develop generic frameworks integrating these techniques.

Segmentation can be a fully automatic process, but it achieves its best results with semi-automatic algorithms, i.e., algorithms that are guided by a human operator. This concept of semi-automatic process naturally involves an environment in which the human operator will interact with the algorithms and the data in order to produce optimal segmentations. The simplest example of the need of a human intervention during the task of segmentation results from the specificity of the existing algorithms: depending on the type of input data, the operator will have to carefully pick the best adapted algorithm, which most of the time cannot be done in an automatic way. The subjective point of view of the human is required.

^[5] Image segmentation has become an essential tool in the medical field with the generalization of diagnosis using Magnetic Resonance Imaging (referred to as MRI), image segmentation is often required to allow doctors and surgeons to analyse the patient's data, e.g. prior to surgery to determine ^[2] the exact location of an organ or a tumour. MRI, also known as Nuclear Magnetic Resonance, is a method used mainly (but not

only) in medical applications to visualize the insides of a patient in a harmless fashion. It relies on the relaxation properties of excited hydrogen nuclei in water after the body or part of the body to image has been placed in a powerful and uniform magnetic field. The obvious benefit of this technique is its harmless character, compared to other techniques such as CT scans and X-rays in which the patient is exposed to ionizing radiations. 3D and 4D (3D + time) MRI is increasingly used in diagnosis and therapy.

The data acquired from MRI is generally presented as a volumetric image that can be viewed as a series of slices following one of the 3 axes: sagittal (following the X axis), coronal (following the Y axis) or axial (following the Z axis). The aim of the segmentation process is to locate with the highest accuracy possible the boundaries of a special part of the image (an organ, a tumour), thus allowing the use of 3D information for planning of treatment, navigation or visualization. Once this task is achieved, simulation, ^[5]pre-computations or training can be performed to prepare the real operation.

In this project trials were conducted to develop a segmentation algorithm for abnormal MRI images using optimized k mean and optimized C-means clustering technique using genetic algorithm and also to compare the results. Abnormal brain images from four classes metastases, meningioma, glioma and astrocytoma are being used in this work.

1.2 Background

A mathematical definition of segmentation

The following is a very general definition of image segmentation. It uses a homogeneity predicate $P(R)$ that helps formalizing the notion of homogeneity in an image: a region R is homogeneous if and only if $P(R) = True$. Therefore, the homogeneity can be defined in infinity of different ways: on the grey levels, on the textures or even on non-obvious properties of the image.

Definition 1 (segmentation): Let I be the set of pixels (the input image) and $P()$ the homogeneity predicate defined on groups of connected pixels.

A segmentation S of I is a partitioning set of image regions $\{R_1, R_2, \dots, R_n\}$ such that

$$\bigcup_{i=1}^n R_i = I \text{ and } R_i \cap R_j = \emptyset \forall i \neq j$$

Is a mathematical definition of a partition: the union of all the regions forms the whole image and all the regions are distinct?

$$P(R_i) = True \forall i$$

Signifies that the homogeneity predicate is valid on every region signifies that the union of two adjacent regions cannot satisfy the homogeneity predicate, i.e. two adjacent regions must be distinct regarding the homogeneity predicate.

$$(R_i \subset R_j) \wedge (R_i \neq \emptyset) \wedge (P(R_j) = True) \Rightarrow (P(R_i) = True)$$

Signifies that the homogeneity predicate is valid on any sub-region of a region where it is verified.

[2] Region Growing:

Region growing is a technique for extracting an image region that is connected based on some predefined criteria. These criteria can be based on intensity information and/or edges in the image. In its simplest form, region growing requires a seed point that is manually selected by an operator and extracts all pixels connected to the initial seed based on some predefined criteria. For example, one possible criterion might be to grow the region until an edge in the image is met. Like thresholding, region growing is seldom used alone but usually within a set of image-processing operations, particularly for the delineation of small, simple structures such as tumours and lesions.

Region growing is a simple region-based image segmentation method. It is also classified as a pixel-based image segmentation method since it involves the selection of initial seed points.

This approach to segmentation examines neighbouring pixels of initial "seed points" and determines whether the pixel neighbours should be added to the region. The process is iterated on, in the same manner as general data [3] clustering algorithms. A general discussion of the region growing algorithm is described below.

2.REVIEW OF LITERATURE:

CLUSTERING METHOD:

^[1] Clustering can be considered the most important unsupervised learning problem because no information is provided about the "right answer" for

any of the objects. It classifies a set of observations in the data, and it finds a reasonable structure in the data set. Here priori information about classes is not required, i.e., neither the number of clusters nor the rules of assignment into clusters are known. They have to be discovered exclusively from the given data set without any reference to a training set. Cluster analysis allows many choices about the nature of the algorithm for combining groups. There are two basic approaches to clustering, which we call supervised and unsupervised. In the case of unsupervised classification on clustering, we do not have labels. If we know the labels of our input data, the problem is considered supervised, or otherwise it is called unsupervised.

Clustering Definition: Clustering is a grouping of data with similar characteristics. To divide the data into several groups the similarity of objects is used, here the distance functions are being used to find the similarity of two objects in the data set.

Cluster analysis:

^[1] Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense.

Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics.

Besides the term clustering, there are a number of terms with similar meanings, including automatic classification, numerical taxonomy, bryology and typological analysis.

Types of clustering:

Hierarchical algorithms find successive clusters using previously established clusters. These algorithms can be either agglomerative ("bottom-up") or divisive ("top-down").^[3] Agglomerative algorithms begin with each element as a separate cluster and merge them into successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters.

^[3] Partitioned algorithms typically determine all clusters at once, but can also be used as divisive algorithms in the hierarchical clustering.

Density-based clustering algorithms are devised to discover arbitrary-shaped clusters. In this approach, a cluster is regarded as a region in which the density of data objects exceeds a threshold. DBSCAN and OPTICS are two typical algorithms of this kind.

Two-way clustering, co-clustering or bi-clustering are clustering methods where not only the objects are clustered but also the features of the objects, i.e., if the data is represented in a data matrix, the rows and columns are clustered simultaneously.

Many clustering algorithms require specification of the number of clusters to produce in the input data set, prior to execution of the algorithm. Barring knowledge of the proper value beforehand, the appropriate value must be determined, a problem for which a number of techniques have been developed.

^[1] Hierarchical clustering:

Hierarchical clustering creates a hierarchy of clusters which may be represented in a tree structure called a dendrogram. The root of the tree consists of a single cluster containing all observations, and the leaves correspond to individual observations.

Algorithms for hierarchical clustering are generally either agglomerative, in which one starts at the leaves and successively merges clusters together; or divisive, in which one starts at the root and recursively splits the clusters. Any valid metric may be used as a measure of similarity between pairs of observations. The choice of which clusters to merge or split is determined by a linkage criterion, which is a function of the pair-wise distances between observations. Cutting the tree at a given height will give a clustering at a selected precision. In the following example, cutting after the second row will yield clusters {a} {b c} {d e} {f}. Cutting after the third row will yield clusters {a} {b c} {d e f}, which is a coarser clustering, with a smaller number of larger clusters.

^{[1][3]} Agglomerative hierarchical clustering:

For example, suppose this data is to be clustered, and the Euclidean distance is the distance metric.

3.METHODOLOGY

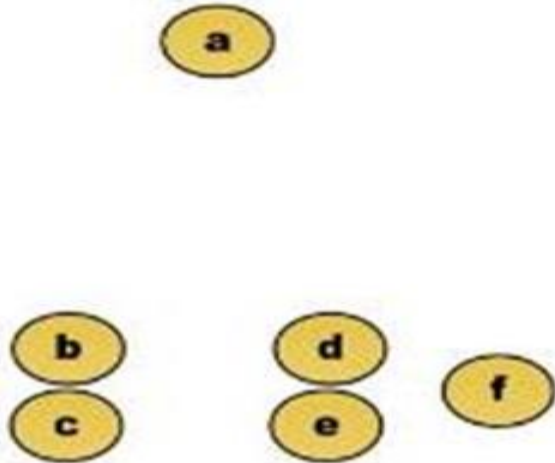


Figure: Raw data

The hierarchical clustering dendrogram would be as such:

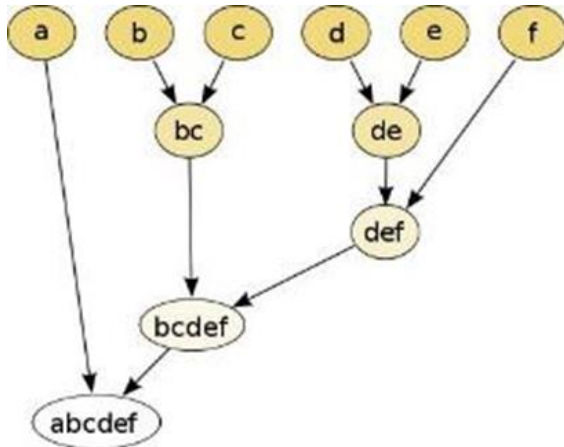


Figure: Traditional representation

This method builds the hierarchy from the individual elements by progressively merging clusters. In our example, we have six elements {a} {b} {c} {d} {e} and {f}. The first step is to determine which elements to merge in a cluster. Usually, we want to take the two closest elements, according to the chosen distance.

Optionally, one can also construct a distance matrix at this stage, where the number in the i-th row j-th column is the distance between the i-th and j-th elements. Then, as clustering progresses, rows and columns are merged as the clusters are merged and the distances updated. This is a common way to implement this type of clustering and has the benefit of caching distances between clusters. A simple agglomerative clustering algorithm is described in the single-linkage clustering page; it can easily be adapted to different types of linkage.

This section presents three important methods of image segmentation for extraction of tumour in the MRI images. The methods include [2]k-means clustering with watershed segmentation algorithm, optimized k-means clustering with genetic algorithm and optimized c-means clustering with genetic algorithm. The Basic approach of the comparison is to implement different algorithm for segmentation of tumour area of MRI images.

[2] Segmentation of brain MRI images for Tumour Extraction

Segmentation of brain MRI images is implemented by combining k-means clustering and watershed algorithm. K-means clustering method helps to segment the brain tumour image and the second method improves the primary results of segmentation of tumour. The method is implemented using following steps:

1. Original image of brain tumour is acquired.
2. The original image is converted into grey scale image since colour image requires lot of
3. time.
4. Suitable pre-processing is applied such as reformatting the data. For example, some commands of MATLAB do not work with unsigned integer and hence the data needs to be converted into double.
5. Histogram is obtained and two limits are determined. Generally LOW limit indicates for black pixel and HIGH limit corresponds to white pixel.
6. The intensity values of image are complemented and adjusted so that the histogram is displayed whose number of bins can be specified. For grey scale images, 256 bins are used as default value but for 2 bins for binary image. Flow of procedures for k means and watershed algorithm can be seen.

[4]Optimized k-means Clustering using Genetic Algorithm (GA)

Brain tumour detection is achieved with the help of k-means clustering which involves a number of post processes such as visualization and analysis. The k means algorithm is very effective; however, traditional k-means algorithm is sensitive to the initial cluster

centres. If these clusters fluctuate with different initial input, then problem arises. Over segmentation and sensitivity to false edges are other difficulties in ordinary k-means method.

Determination of exact location and area of brain tumour using k-means method becomes very difficult and hence use of genetic algorithm is suggested. GAs with the modification of mutation operations improves the speed of convergence and computing time is reduced also. Several experimental results suggest that genetic algorithm-based k-means algorithm not only eliminates the over-segmentation problem but provides fast and efficient clustering also.

Various major steps of optimized k-means algorithm for tumour detection are:

1. Step 1: Determine the parameters of GA and generate initial population randomly.
2. Step 2: Determine upper and lower bounds of parameters.
3. Step 3: Evaluate the fitness function of individuals of the initial population before and after optimization.
4. Step 4: Create new set of populations by using selection, crossover and mutation operators.
5. Step 5: Evaluate the fitness of individuals of the new population and repeat Step 4 until fitness requirements are met.
6. Step 6: Obtain clustering results and apply filtering over clustering results.
7. Step 7: Apply morphological operation and then apply watershed algorithm to extract the tumour area.

Using the above steps, brain tumour extraction and detection is achieved using following steps:

1. Step 1: Get an original image and convert into grey scale image.
2. Step 2: Apply some pre-processing methods such as image de-noising.
3. Step 3: Initialize the population for GA.
4. Step 4: Apply the clustering algorithm.
5. Step 5: Assume any suitable fitness function initially. Now, for $i=1$ to maximum value of iterations: repeat Step vi to Step x
6. Step 6: Perform the Selection operation on training dataset.
7. Step 7: Determine the Crossover of selected parents; and generate the next level.

8. Step 8: Perform the Mutation for ignoring the values not following fitness function.
9. Step 9: Recombine the generated child with existing population; and generate new population.
10. Step 10: Apply clustering algorithm over new set.
11. Step 11: Generate mean of the resulting image; compare the result with suitable threshold value and obtain tumour detection resulted in final image.

^[3] Optimized c-means Clustering

MRI images of brain are segmented, and tumour can be efficiently extracted and detected is c-means clustering is optimized. The c-means clustering method has been implemented and its performance can be improved by using optimization with the use of genetic algorithm. The combined method results an improvement in segmentation efficiency and higher area of affected region extraction and detection. The aim of optimized clustering and its analysis is to divide a given set of data into a number of clusters which should follow the properties given below:

- i. Homogeneity inside clusters: The information or the data of a cluster is as similar as possible.
- ii. Heterogeneity between the clusters: Here the data belongs to different clusters are different.

A clustering technique is used to obtain a partition of N objects using a suitable measure such as resemblance function as a distance measure's'. The method involves following steps.

- i. Step 1: Choose a suitable number of clusters (say 'k').
- ii. Step 2: Set initial centres of clusters as c_1, c_2, \dots, c_k .
- iii. Step 3: Classify each vector x into the closest centre c_i by using Euclidean distance:
 - a. $\|x_i - c_i\| = \min \|x_i - c_i\|$
- iv. Step 4: Recomputed the estimated cluster centres.
- v. Step 5: If no any cluster centres changes then go to Step 3.

^[4] Genetic Algorithm (GA)

The term Genetic is derived from Greek word "genesis" which means "to grow" or "to become", and therefore the algorithm makes a function grow. This algorithm was introduced by John Holland on the basis of a heuristic method. The method grows in search of "survival of the fittest. Since fittest is searched by the algorithm and hence used in optimization tasks, the

implementation of genetic algorithm begins with an initial population of chromosomes which are randomly selected. A chromosome is a long thread of DNA (deoxyribonucleic acid). Particulars traits determine the hereditary of an individual where each trait is coded by some combination of DNA bases. The four main bases of DNA are A (Adenine), C (Cytosine), T(Thymine) and G (Guanine). Just like English alphabet, the combinations of various letters give some meaningful information; GA also follows the same concept. Flow diagram of implementation of GA for an application is shown.

Firefly Algorithm

Algorithm:

The firefly algorithm is implemented in this paper and the steps are stated below:

Step 1. Initialization

Step 2. The population of fireflies are given as, $\{x_1, x_2, \dots, x_n\}$.

Step 3. Calculate brightness value using cost function for assigned firefly.

Step 4. Fireflies intensity are given as $\{I_1, I_2, \dots, I_n\}$.

Step 5. Update the step of each firefly.

Step 6. Rank the fireflies and find the current best.

Step 7. Move each firefly i towards other brighter fireflies.

Step 8. Update the solution set.

Step 9. Stop when criterion is fulfilled; otherwise go to Step 2.

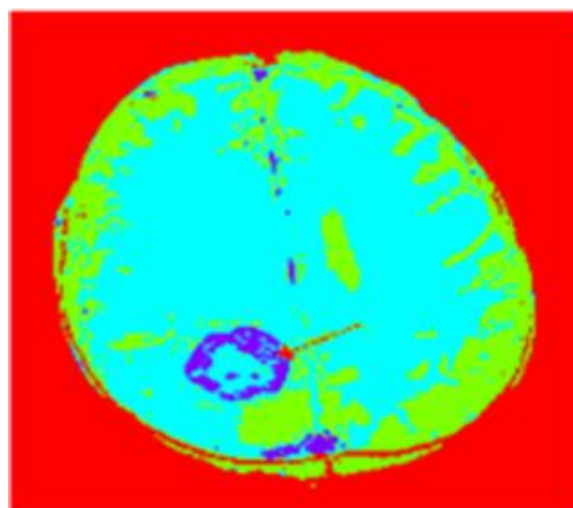
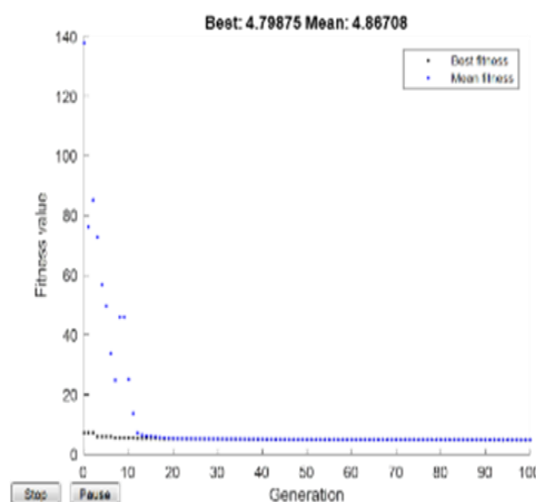
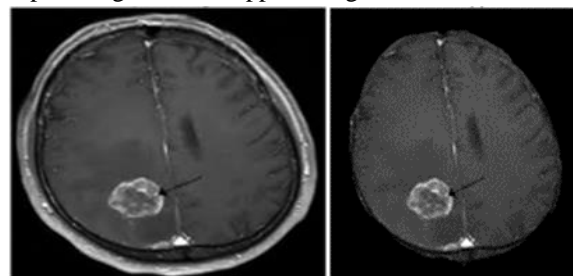
4.RESULTS AND DISCUSSION

The searching time and area of tumour region were considered as comparison parameters for comparison of various methods. The c-means clustering produced good results and performed better than other optimized clustering methods. A number of experimental results were obtained applying existing algorithms. In this work, around 50 MRI images of brain were collected as real time images. This was done with the help of radiologists of diagnosis centre.

Some brain MRI images are shown. These images are subjected to different methods for image segmentation and detection of tumours present in the images. [6] MRI images were segmented using k-means clustering and Watershed algorithm. The method is implemented using process of two stages. The first stage of the

process uses k-means clustering and primary segmentation results are produced for the brain MRI images. Second stage of the process is applied as [6] watershed segmentation algorithm to improve the results of the primary segmentation; and the results obtained are final results.

Input Image Skull Stripped Image





GA K-Means

Image after removing small Connected components using GA



Image after removing small Connected components using FA

| Image name | K-mean Clustering Area (Pixel Value) | GA K-Mean Area (Pixel Value) | FA K-Mean Area (Pixel Value) |
|------------|--------------------------------------|------------------------------|------------------------------|
| MRI | 689 | 777 | 821 |

5.CONCLUSION

The segmentation of brain MRI images using k-means, c-means and optimized methods were implemented. Following are the outcomes of the work:

1. Segmentation was achieved for all the proposed methods tumour detection was done.
2. The k- means clustering with watershed segmentation algorithm, optimized k –means clustering with genetic algorithm and optimized

c- means clustering with genetic algorithm were the main methods.

3. A comparison was also made in terms of tumor region and search time.
4. The c-means clustering after optimization was found better than other methods.
5. The problem of over segmentation was also addressed.
6. The genetic algorithm improves the convergence and computing time is reduced.
7. The firefly algorithm improves the convergence even better compared to genetic algorithm.

The further research scope of present work is to implement the methods for 3D images of brain tumour MRI. Other parameters such as volume and growth rate of tumour can be extracted to produce better segmentation results and their performance.

REFERENCE

- [1] P.Tamije;V. Palanisamy; T. Purusothaman: "Performance Analysis of Clustering Algorithms in Brain Tumor Detection of MR Images" European Journal of Scientific Research, ISSN1450-216X Vol.62 No.3 (2011), pp. 321-330.
- [2] Ratan, Rajeev, Sanjay Sharma, and S. K. Sharma. "Brain tumor detection based on multi-parameter MRI image analysis." International Journal on Graphics, Vision and Image Processing vol9.no.3, pp.9-17,2009.
- [3] Gopal, N.N.; Karnan, M., "Diagnose brain tumor through MRI using image processing clustering algorithms such as Fuzzy C Means along with intelligent optimization techniques," IEEE International Conference on Computational Intelligence and Computing Research (ICIC), vol.2, no.3, pp.1-4, 2010.
- [4] Amanpreet Kaur; Gagan Jindal "Tumor Detection Using Genetic Algorithm" International Journal on Computer Science and Technology, vol. 4, no.1, pp. 423-427 2013.
- [5] S. Datta; M. Chakraborty. "Brain Tumor Detection from Pre- Processed MR Images using Segmentation Techniques". Special Issue on 2nd National Conference-Computing, Communication and Sensor Network (CCSN) Published by Foundation of Computer Science, New York, USA. vol.2, pp.1-5, 2011.

- [6] Wenli Yang; Zhiyuan Zeng; Sizhe Zhang:
“Application of Combining Watershed and Fast Clustering Method in Image Segmentation”,
Computer Modeling and Simulation, 2010.
ICCMS '10. Second International Conference on,
vol.3, no., pp.170-174,22-24, Jan.2010 doi:
10.1109/ICCM2010.407.