Available at: www.researchpublications.org

Brain Image Segmentation Algorithm using K-Means Clustering

Ms. Mamta K. Date
M.E. computer science and engineering,
PRMITR, Badnera
Email-mamta_date@rediffmail.com

Mr. S.P. Akarte
Assistant professor of computer science and engineering,
PRMITR,Badnera
Email-s_akarte25@rediffmail.com

***Abstract*-To segments the medical image using K-means clustering algorithm. To propose an algorithm that can be better for large datasets and to find initial centroid. To compare the performance. An algorithm is described for segmenting MR brain image into *K* different tissue types, which include gray, white matter and CSF, and maybe other abnormal tissues. MR images considered can be either scale- or multivalued. Each scale-valued image i s modeled as a collection of regions with slowly varying intensity plus a white Gaussian noise. The proposed algorithm is an adaptive K-means clustering algorithm for 3- dimensional and multi-valued images. Each iteration consists of two steps: estimate mean intensity at each location for each type, and estimate tissue types, Its performance is tested using patient data.**

I. INTRODUCTION

The advances in imaging technology, diagnostic imaging has become an indispensable tool in medicine today. X-ray angiography (XRA), magnetic resonance angiography (MRA), magnetic resonance imaging (MRI), computed tomography (CT), and other imaging modalities are heavily used in clinical practice. Such images provide complementary information about the patient. While increased size and volume in medical images required the automation of the diagnosis process, the latest advances in computer technology and reduced costs have made it possible to develop such systems. Blood vessel delineation on medical images forms an essential step in solving several practical applications such as diagnosis of the vessels (e.g. stenosis or malformations) and registration of patient images obtained at different times. Segmentation algorithms form the essence of medical image applications such as radiological diagnostic systems, multimodal image registration, creating anatomical atlases, visualization, and computer-aided surgery Vessel segmentation algorithms are the key components of automated radiological diagnostic systems. Segmentation methods vary depending on the imaging modality, application domain, method being automatic or semi-automatic, and other specific factors. There is no single segmentation method that can extract vasculature from every medical image modality. While some methods employ pure intensity based pattern recognition techniques such as thresholding followed by connected component analysis, some other methods apply explicit vessel models to extract the vessel contours. Depending on the image quality and the general image artifacts such as noise, some segmentation methods may require image preprocessing prior to the segmentation algorithm. On the other hand, some methods apply post-processing to overcome the problems arising from over segmentation. Vessel segmentation algorithms and techniques can be divided into six main categories, pattern recognition techniques, model-based approaches, tracking-based approaches, artificial intelligence-based approaches, neural network-based approaches, and miscellaneous tube-like object detection approaches. Pattern recognition techniques are further divided into seven categories, multi-scale approaches, skeleton-based approaches, region growing approaches, ridge-based approaches, differential geometry-based approaches, matching filters approaches, and mathematical morphology schemes. Clustering analysis plays an important role in scientific research and commercial application. This thesis provides a survey of current vessel segmentation methods using clustering approach and provides both early and recent literature related to vessel segmentation algorithms and techniques.

Existing system

Initial centroid value is assumed. Probability of getting a zero matrix. Fails in large data sets

Proposed system

The improved K-means algorithm is a solution to handle large scale data, which can select initial clustering center purposefully, reduce the sensitivity to isolated point, and avoid dissevering big cluster. By using this technique locating the initial seed point is easy and which will give more accurate and high-resolution result. By using various techniques we can study or compare the results and find out which technique gives higher resolution Initial centroid algorithm is useful to avoid the formation of empty clusters, as the centroid values are taken with respect to the intensity value of the image. Proposed algorithm is better for large datasets and to find initial centroid.

Module description:

User Login:

Allow only authenticated users. Restrict unauthorized access.

Available at: www.researchpublications.org

Table contains user information and also their brain reports.

Images Segmentation:
Extract 2-D images from the 3-D images.
Store that in different locations.
2-D image will be converted into 1-D image.

II.CLUSTERING ANALYSIS

K-Means clustering algorithm – similar to nearest neighbor techniques (memory-based-reasoning and collaborative filtering) – depends on a geometric interpretation of the data Organizing data into clusters shows internal structure of the data Ex. Clusty and clustering genes above Sometimes the partitioning is the goal Ex. Market segmentation Prepare for other AI techniques Ex. Summarize news (cluster and then find centroid) Techniques for clustering is useful in knowledge discovery in data

Block Diagram:

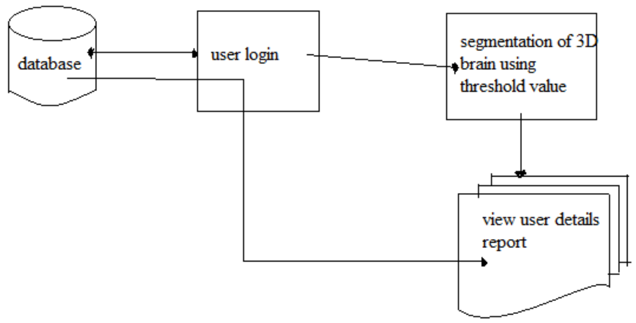
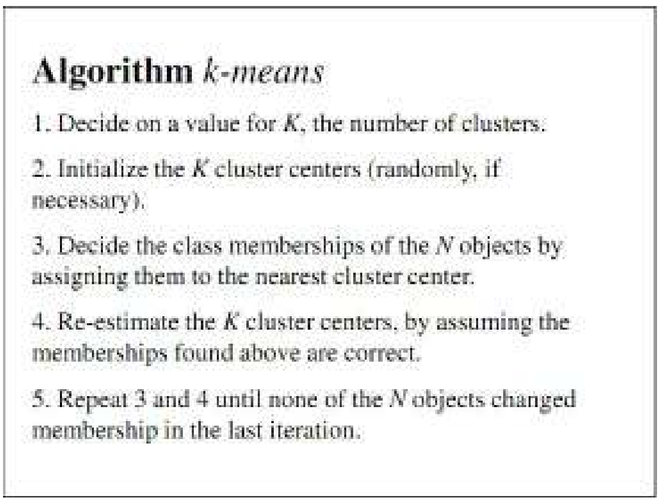


Figure 1:-Block diagram

*Block diagram-Description***:**

- Back end: Sql Server2008
- Contains different type of tables.
- Mainly for User details and Login purpose.
- Scanned image will be stored in file locations.

III. K-MEANS ALGORITHM



Algorithm ***k-means***

1. Decide on a value for *K*, the number of clusters.
2. Initialize the *K* cluster centers (randomly, if necessary).
3. Decide the class memberships of the *N* objects by assigning them to the nearest cluster center.
4. Re-estimate the *K* cluster centers, by assuming the memberships found above are correct.
5. Repeat 3 and 4 until none of the *N* objects changed membership in the last iteration.

Figure 2:-Algorithm for K-Means Algorithm

A) K-means overview

There are always K clusters. There is always at least one item in each cluster. The clusters are non-hierarchical and they do not overlap. Every member of a cluster is closer to its cluster than any other cluster because closeness does not always involve the 'centre' of clusters. *K*-means clustering in particular when using heuristics such as Lloyd's algorithm is rather easy to implement and apply even on large data sets. As such, it has been successfully used in various topics, ranging from market segmentation, computer vision and astronomy to agriculture. It often is used as a preprocessing step for other algorithms, for example to find a starting configuration. In statistics and data mining, *k*-means clustering is a method of cluster analysis which aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. Figure1 shows the flow chart of k-means algorithm which is relatively efficient and applicable only when mean is defined. Figure1 shows the flow chart of k-means algorithm which is relatively efficient and applicable only when mean is defined.

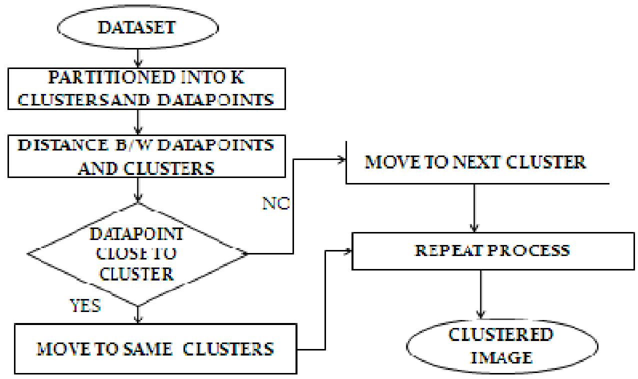


Figure 3:- Flow chart for K-Means Algorithm

B) K-means clustering

K-means (Macqueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group age is done. At this point we need to recalculate k new centroids as barycenters of the clusters resulting from the previous step.

After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop had been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move anymore. K-Means clustering generates a specific number of disjoint, flat clusters. K-Means method is numerical, unsupervised, non-deterministic and iterative. Hierarchical clustering is also widely employed for image segmentation. The most popular method for image segmentation is k-means clustering.

IV.FEATURE EXTRACTIONS

The issue of choosing the features to be extracted should be guided by the following concerns. The features should carry enough information about the image and should not require any domain-specific knowledge for their extraction. They should be easy to compute in order for the approach to be feasible for large image collection and rapid retrieval. An image is partitioned into 4x4 blocks, a size that provides a compromise between texture granularities, computation time and segmentation coarseness as a part of preprocessing; each 4x4 block is replaced by a single block containing the average value over the 4x4 block. To segment an image into objects, some features are extracted from each block. Texture features are extracted using Haar Wavelet Transform. After obtaining features from all pixels on the image, perform k-means clustering to group similar pixel together and form objects. Feature extraction has been done using MATLAB Image Processing tool. The advantage of k-means algorithm is that it works well when clusters are not well separated from each other, which is frequently encountered in images. However k-means requires the user to specify the initial cluster centers. Image clustering consists of two steps, the former is feature extraction and the second part is grouping. For each image in the database, a feature vector capturing certain essential properties of the image is computed and stored in a feature base. Clustering algorithm is applied over this extracted feature to form the group. In terms of performance the algorithm is not guaranteed to return a global optimum. The quality of the final solution depends largely on the initial set of clusters, and may, in practice, be much poorer than the global optimum.] Since the algorithm is extremely fast, a common method is to run the algorithm several times and return the best clustering found.

V. CLUSTER ALGORITHM

K-Means uses a two-phase iterative algorithm to minimize the sum of point-to-centroid distances, summed over all k clusters: The first phase uses *batch updates*, where each iteration consists of reassigning points to their nearest cluster centroid, all at once, followed by recalculation of cluster centroids. This phase occasionally does not converge to solution that is a local minimum, that is, a partition of the data where moving any single point to a different cluster increases the total sum of distances. This is more likely for small data sets. The batch phase is fast, but potentially only approximates a solution as a starting point for the second phase. The second phase uses *online updates*, where points are individually reassigned if doing so will reduce the sum of distances, and cluster centroids are recomputed after each reassignment. Each iteration during the second phase consists of one pass though all the points. The second phase will converge to a local minimum, although there may be other local minima with lower total sum of distances. The problem of finding the global minimum can only can be solved in general by an exhaustive (or clever, or lucky) choice of starting points, but using several replicates with random starting points typically results in a solution that is a global minimum.

K-means functions

K-means is a clustering algorithm, which partitions a data set into clusters according to some defined distance measure. Images are considered as one of the most important medium of conveying information. Understanding images and extracting the information from them such that the information can be used for other tasks is an important aspect of Machine learning. An example of the same would be the use of images for navigation of robots. One of the first steps in direction of understanding images is to segment them and find out different objects in them. To do this, we look at the algorithm namely K-means clustering. It has been assumed 1that the number of segments in the image is known and hence can be passed to the algorithm. K-Means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The functions of k-means are as follows. IDX = k-means(X, k) partitions the points in the n-by-p data matrix X into k clusters. This iterative partitioning minimizes the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid distances. Rows of X correspond to points, columns correspond to variables. K-means returns an n-by-1 vector IDX containing the cluster indices of each point. By default, k-means uses squared Euclidean distances [8,9]. When X is a vector, k-means treats it as an *n*-by-1 data matrix, regardless of its orientation.[IDX,C] = k-means(X,k) returns the k cluster centroid locations in the k-by-p matrix C.

VI. SIMULATION RESULTS

Figure 4:- Original image

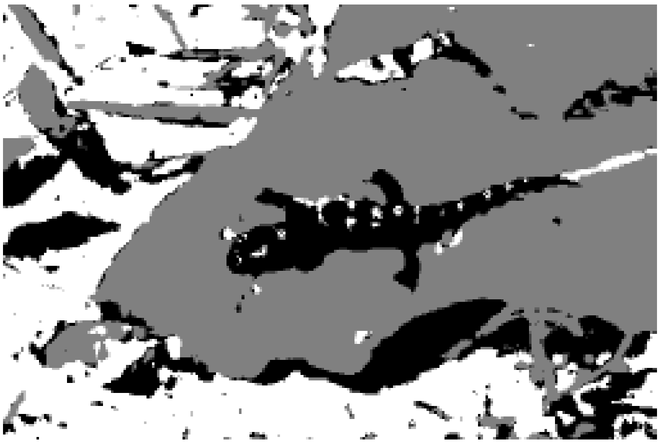


Figure 5:-image labeled by cluster index

Figure 6:-objects in cluster 1

Figure 7:-objects in cluster2

Figure 8:-objects in cluster 3

VI. CONCLUSION

Vessel segmentation methods have been a heavily researched area in recent years. Even though many promising techniques and algorithms have been developed, it is still an open area for more research. This algorithm does not require any user interaction, not even to identify a start point. Here seed points are selected randomly which determines the main branches of the vessel structure. Random selection of seed points does not yield accurate segmentation. Accuracy of the segmentation process is essential to achieve more precise and repeatable radiological diagnostic systems. Accuracy can be improved by incorporating a priori information on vessel anatomy and let high level knowledge guide the segmentation algorithm. K means algorithm is a popular clustering algorithm applied widely, but the standard algorithm which selects k objects randomly from population as initial centroids cannot always give a good and stable clustering. Experimental results show that selecting centroids by our algorithm can lead to a better clustering. Along with the fast development of database and network, the data scale clustering tasks involved in which becomes more and more large. K-means algorithm is a popular partition algorithm in cluster analysis, which has some limitations when there are some restrictions in computing resources and time, especially for huge size.

VII. REFERENCES:

[1] S. P. Lloyd, ―"Least squares quantization in PCM," *IEEE Trans. Inf.Theory*, vol. IT-28, no. 2, pp. 129–136, Mar.1982.

[2] J. Shi and J. Malik, ―Normalized cuts and image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug.2000.

[3] M. Mignott e, C. Collet, P. Pérez, and P. Bouthemy,"Sonar image segmentation using a hierarchical MRFmodel,"*IEEE Trans. Image Process.*, vol. 9, no.7, pp.1216–1231, Jul. 2000.

[4] F. Destrempes, J.-F. Angers, and M. Mignotte, "Fusion of hidden Markov random field models and its Bayesian estimation," *IEEE Trans. Image Process.*, vol. 15, no. 10,pp. 2920–2935, Oct. 2006.

[5] J.A Hartigan "Clustering Algorithms", New York Wiley 1975.

[6] Staib,L., Zeng,X., Schultz,R., and Duncan,J., Shape constraints in deformable models. Handbook of Medical Imaging, Bankman,I.,ed.,2000,pp.147-157

[7] Leventon,M., Faugeraus,O., Grimson,W., and Wells,W.,Level set based segmentation with intensity and curvature priors.Workshop on Mathematical Methods in Biomedical Image Analysis Proceedings,2000,pp.4-11.

[8] Paragios, Deriche R, Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Transaction on Pattern Analsis and Machine Intelligence, 2000,pp .266-280.

[9] Vese L A, Chan T F, A multiphase level set frame wor for image segmentation using the mumford and shah model. International Journal of Computer Vision,2002 ,pp.271-293.

[10] Shi Yanggang,Karl W C, Real-time tracking using level set, IEEE Computer Society Conference on Computer Vision and Pattern Recognition,2005,pp,34-42