

Segmentation of Brain Tumour from MRI image – Analysis of K-means and DBSCAN Clustering

Prof. Samir Kumar Bandyopadhyay
Department of Computer Science and Engineering
University of Calcutta

Tuhin Utsab Paul
Department of Electronics and Communication Engineering, IEM
West Bengal University of Technology

Abstract : We have applied K-means clustering and DBSCAN to the problem of clustering and segmenting Brain Tumour from Magnetic Resonance images of human brain. For our experiments we used a large number of MR images from the “radiopaedia image database”, which is freely available collection of Radio images of different human anatomies specifically for research purpose. We ran the algorithms with varying values on their parameters and studied the effects and did a comparative study of the two algorithms. The resulting cluster were analysed and for K-means we found three types of cluster. Although the results from K-means were quite noisy, many of the clusters could be used for further analysis. The results from DBSCAN proved to be more satisfactory as the clustering method is density based and the density of clusters in input data are highly specially concentrated. Further study of other clustering algorithms applied to the same data would shed light on the actual effectiveness of the algorithms studied here.

I. Introduction

Today, Cluster Analysis is important for many fields of study, for example biology, psychology and other social sciences[1]. More importantly for this paper it provides an effective way of extracting knowledge from large sets of data, i.e. Data Mining[2]. Specifically, Cluster Analysis is a method for partitioning data into groups where members of a group is similar in some way and dissimilar to members outside the group. These groups are referred to as clusters[1].

In computer vision, segmentation refers to the process of partitioning a digital image into multiple regions (set of pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images.

Segmentation is usually the first task of any image analysis process, and thus, subsequent tasks rely heavily on the quality of segmentation. For this reason, a considerable care is taken to improve the probability of a successful segmentation.

Image segmentation has taken a central place in numerous applications, including, but not limited to, multimedia databases, color image and video transmission over the Internet, digital broadcasting, interactive TV, video-on-demand, computer-based training, distance education, video-conferencing and tele-medicine, and, with the development of the hardware and communications infrastructure, to support visual applications. The field has become a principal area of research, not only in electrical engineering, but also in other academic disciplines, such as computer science, geography, medical imaging, criminal justice, and remote sensing.

The result of image segmentation is a set of region that collectively covers the entire image, or a set of contour extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics.

In this paper we had segmented Magnetic Resonance images of human brain using two different segmentation and clustering algorithms. Firstly, we had used K means algorithm which is a unsupervised learning, iterative convergence based segmentation algorithm. The second algorithm used is DBSCAN which is a density based segmentation algorithm. The main advantage of DBSCAN over K-means is that DBSCAN is efficient to handle noise points as outlier points but k means cannot handle noise with such high efficiency. In the case of brain MR images, the presence of such noise is high as the cerebrospinal fluid, which is the main constituent of brain tumour is also present scattered on the upper strata to the brain. So it becomes increasingly difficult to

differentiate between small coagulation of cerebrospinal fluids and the tumour. In these condition DBSCAN gives much satisfactory results.

II. Recent Works

The process of separation of image areas based on different attribute such as texture, grey value range etc is known as Image segmentation. The main objective in image processing applications is extraction of important image features from image data which will eventually lead to automatic computerized description, interpretation and analysis of the scene. Segmentation by the medical experts manually from the magnetic resonance images of the brain tumour is very much time-consuming task, tiresome, susceptible to error,. Several segmentation methods had been proposed by the digital image processing community, many of which are ad - hoc [6]. Four of the most common methods are: 1) amplitude thresholding, 2) texture segmentation 3) template matching, and 4) region-growing segmentation. This is very much important for detecting necrotic tissues, edema and tumours.

Various algorithms for segmentation [16:17:18:19:20:21:22] had been suggested by several authors. Siyal et al described a new method on "Fuzzy C-means for segmentation purpose" [7]. Phillips, W.E et al described "Application of fuzzy C-Means Segmentation Technique for tissue Differentiation in MR Images of a hemorrhagic Glioblastoma Multiforme"[15]. S. Murugavalli et al, proposed "A high speed parallel fuzzy c-mean algorithm for brain tumor segmentation" [14]. S. Murugavalli, proposed "An Improved Implementation of Brain Tumor Detection Using Segmentation Based on Neuro Fuzzy Technique" [13], Vaidyanathan M et al described "Comparison of Supervised MRI Segmentation methods for Tumor Volume Determination during Therapy"[12]. Jayaram K et al described "Fuzzy Connectedness and Image Segmentation"[11]. Kannan et al describe "Segmentation of MRI Using New Unsupervised Fuzzy C mean Algorithm"[10] Ruspini, E Described "Numerical methods for fuzzy clustering"[9]. Dunn, J.C., described "A fuzzy relative of the ISODATA process and its use in detecting compact, well Separated clusters"[6]. Bezdek, J.C., described "Cluster validity with fuzzy sets"[8]. Some other methods such as Learning vector quantization, Watershed, Hybrid SOM or graph cut based approach had also been proposed in different literatures. [32]

III. Proposed Work

Cluster Analysis:

Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait - often proximity according to some defined distance measure. Data clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. The computational task of classifying the data set into k clusters is often referred to as

k-clustering.

Using this concept we have worked on two algorithms. We have implemented those algorithms in satellite images for our concerned problem and have done a comparative study on them.

K-MEANS ALGORITHM:

One the most famous partitional clustering algorithm is K-means algorithm. K-means algorithm is an iterative technique which classifies objects based on attributes or features into K number of group (where k is a positive integer). The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. [34][35]

The K-means algorithm assigns each point to the cluster whose center (also called centroid) is nearest. The center is the average of all the points in the cluster — that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster.

Example: The data set has three dimensions and the cluster has two points: $X = (x_1, x_2, x_3)$ and $Y = (y_1, y_2, y_3)$. Then the centroid Z becomes $Z = (z_1, z_2, z_3)$, where $z_1 = (x_1 + y_1)/2$ and $z_2 = (x_2 + y_2)/2$ and $z_3 = (x_3 + y_3)/2$. [34][35]

Steps of K-means clustering algorithm are as below.

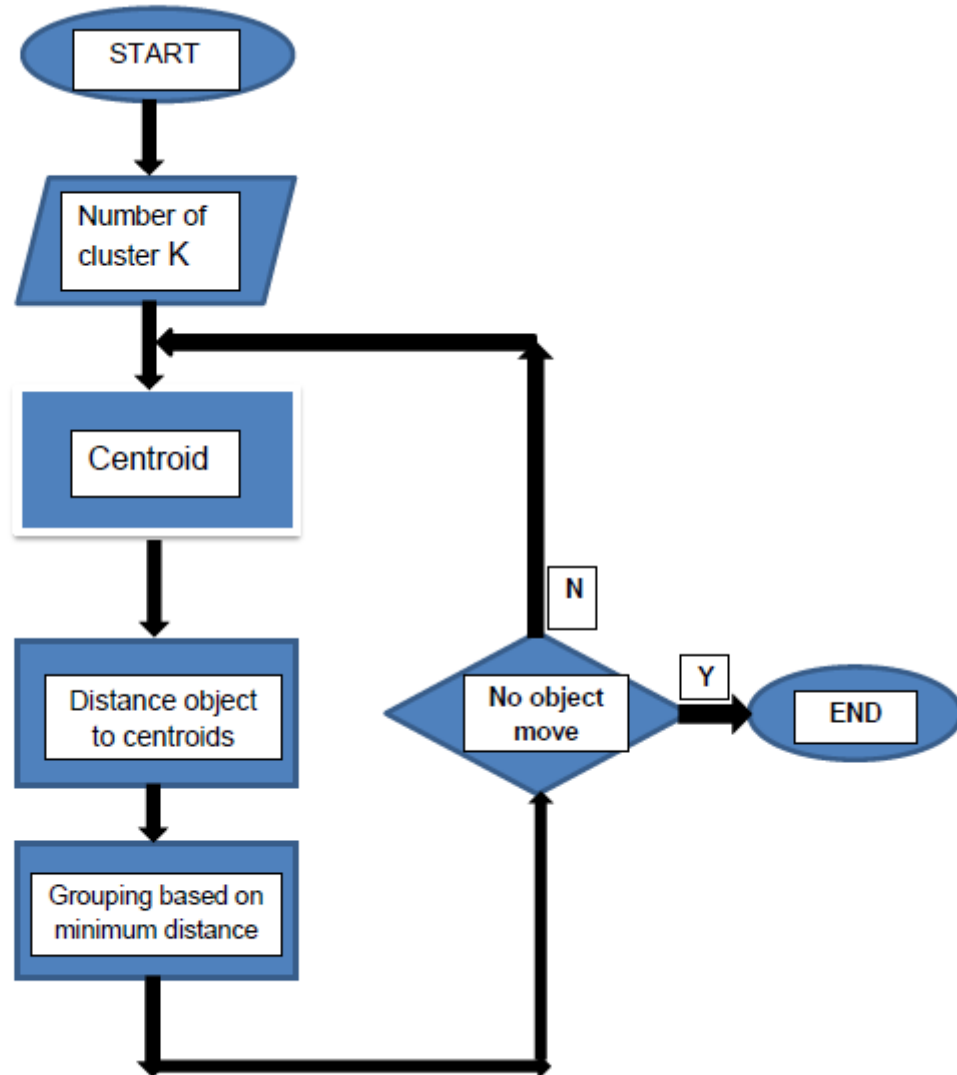
Step1. Choose number of cluster, k .

Step2. Randomly generate k clusters and determine the cluster centers, or directly generate k random points as cluster centers.

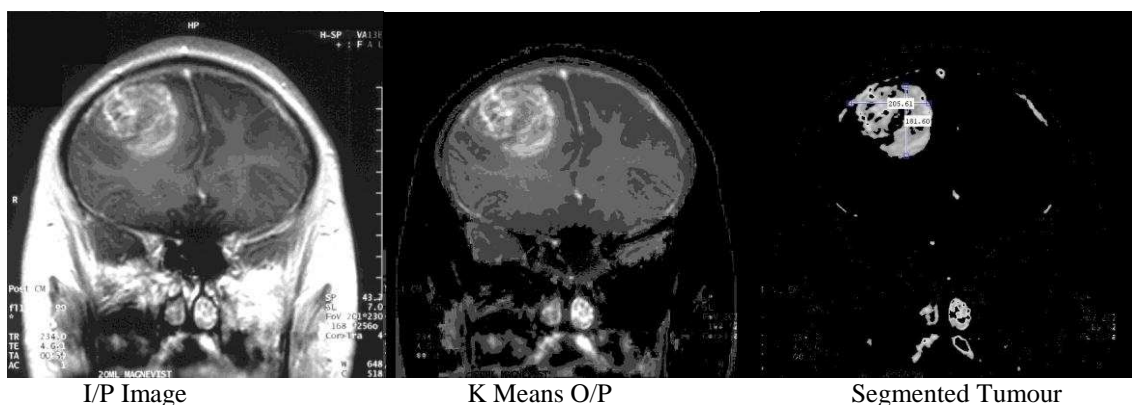
Step3. Assign each point to the nearest cluster center.

Step4. Recompute the new cluster centers.

Step5. Repeat the two previous steps until some convergence criterion is met (usually that the assignment hasn't changed).



Results of K – Means:



IV. DBSCAN

The DBSCAN algorithm [33] can identify clusters in large spatial data sets by looking at the local density of database elements, using only one input parameter. Furthermore, the user gets a suggestion on which parameter value that would be suitable. Therefore, minimal knowledge of the domain is required. The DBSCAN can also determine what information should be classified as noise or outliers. In spite of this, its working process

is quick and scales very well with the size of the database– almost linearly. By using the density distribution of nodes in the database, DBSCAN can categorize these nodes into separate clusters that define the different classes. DBSCAN can find clusters of arbitrary shape, as can be seen in figure [1]. However, clusters that lie close to each other tend to belong to the same class.

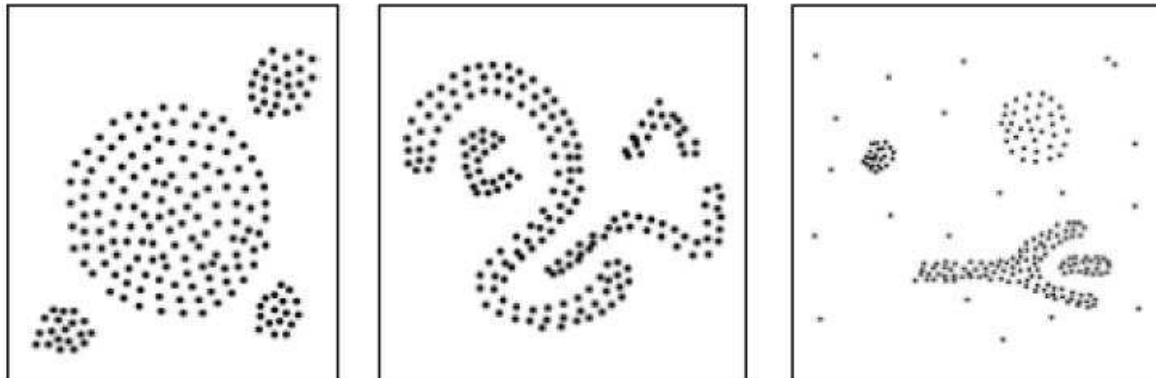


Figure: 1. The node distribution of three different databases, taken from SEQUOIA 2000 benchmark database.

The following section will describe further how the DBSCAN algorithm works. Its computing process is based on six rules or definitions, creating two arguments.

Definition 1: (The *Eps*-neighborhood of a point)

$$NEps(p) = \{q \in D \mid \text{dist}(p, q) < Eps\}$$

For a point to belong to a cluster it needs to have at least one other point that lies closer to it than the distance *Eps*. [33]

Definition 2: (Directly density-reachable)

There are two kinds of points belonging to a cluster; there are border points and core points, as can be seen in figure 2.

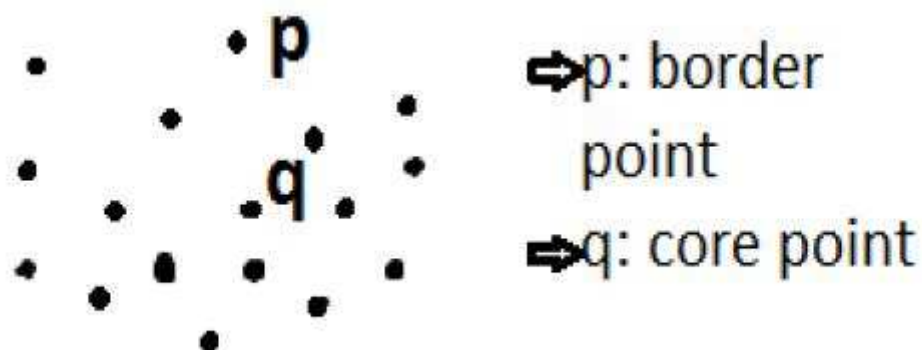


Fig: 2. Border & Core Point

The *Eps*-neighborhood of a border point tends to have significantly less points than the *Eps* neighborhood of a core point. The border points will still be a part of the cluster and in order to include these points, they must belong to the *Eps*-neighborhood of a core point *q* as seen in figure 3.[33]

1) $p \in NEps(q)$

In order for point *q* to be a core point it needs to have a minimum number of points within its *Eps* neighborhood.

2) $|NEps(q)| \geq \text{MinPts}$ (core point condition)

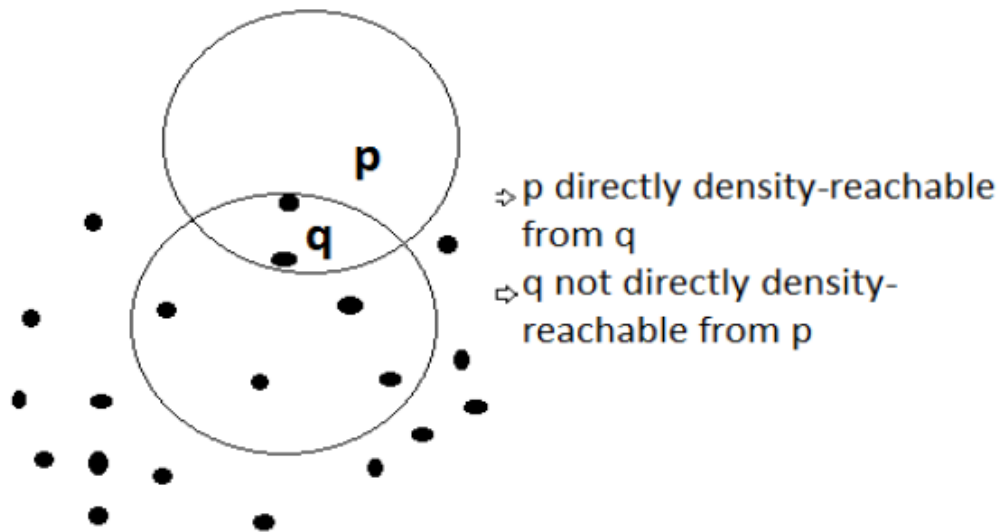


Figure: 3. Point p is directly density-reachable from point q but not vice versa

Definition 3: (Density-reachable)

A point p is *density-reachable* from a point q with respect to Eps and $MinPts$ if there is a chain of points p_1, \dots, p_n , $p_1 = q$, $p_n = p$ such that p_{i+1} is *directly density-reachable* from p_i . Figure [4] shows an illustration of a density-reachable point. [33]

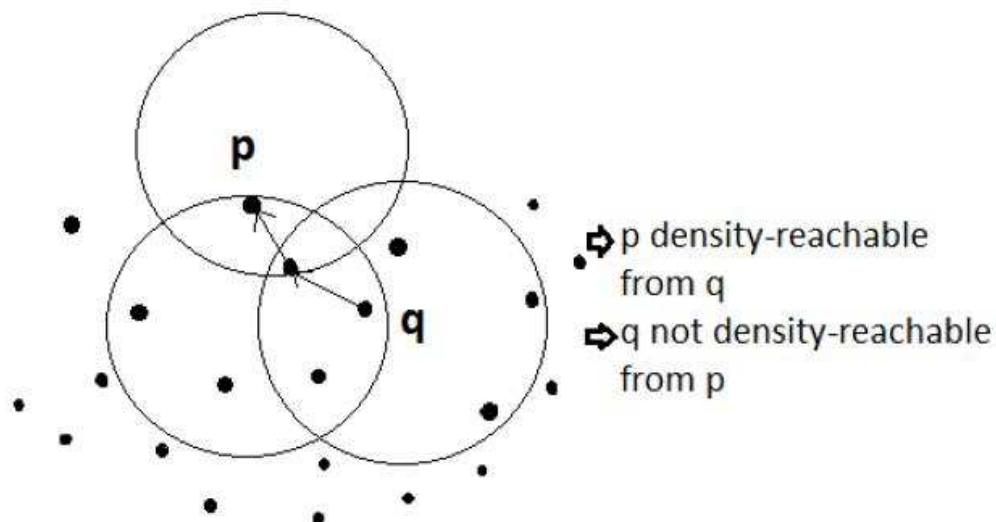


Figure: 4. Point p is density-reachable from point q and not vice versa

Definition 4: (Density-connected)

There are cases when two border points will belong to the same cluster but where the two border points don't share a specific core point. In these situations the points will not be *density-reachable* from each other. There must however be a core point q from which they are both *density-reachable*. Figure [5] shows how density connectivity works. [33]

A point p is *density-connected* to a point q with respect to Eps and $MinPts$ if there is a point o such that both, p and q are density-reachable from o with respect to Eps and $MinPts$.

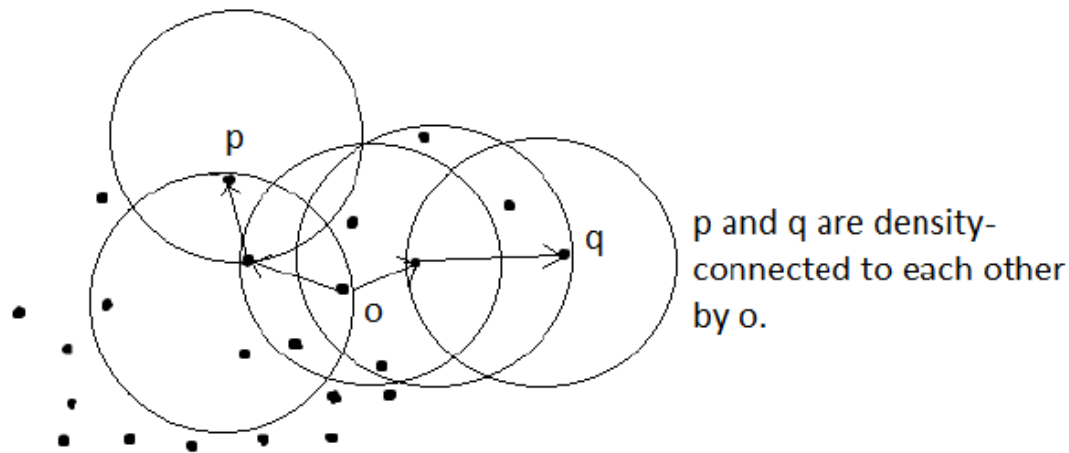


Figure: 5. Density connectivity.

Definition 5: (cluster)

If point p is a part of a cluster C and point q is *density-reachable* from point p with respect to a given distance and a minimum number of points within that distance, then q is also a part of cluster C .

1) " $\square \square p, q$ ": if $p \in C$ and q is *density-reachable* from p with respect to Eps and $MinPts$, then $q \in C$.

Two points belongs to the same cluster C , is the same as saying that p is density-connected to q with respect to the given distance and the number of points within that given distance.

2) " $\square \square p, q \in C$ ": p is *density-connected* to q with respect to Eps and $MinPts$. [33]

Definition 6: (noise)

Noise is the set of points, in the database, that don't belong to any of the clusters.

Argument 1:

A cluster can be formed from any of its core points and will always have the same shape.

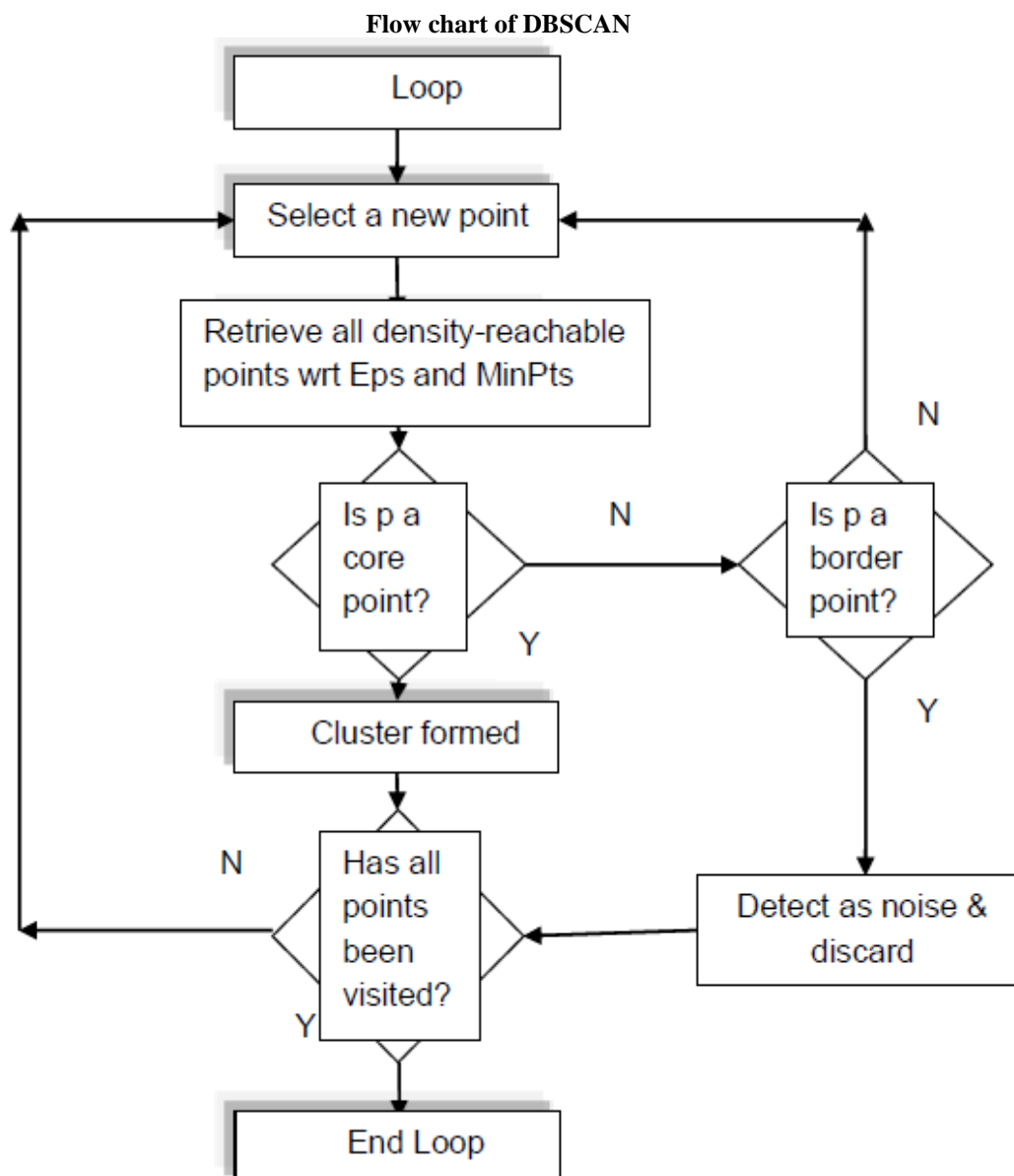
Argument 2:

Let p be a core point in cluster C with a given minimum distance (Eps) and a minimum number of points within that distance ($MinPts$). If the set O is *density-reachable* from p with respect to the same Eps and $MinPts$, then C is equal to the set $_O$.

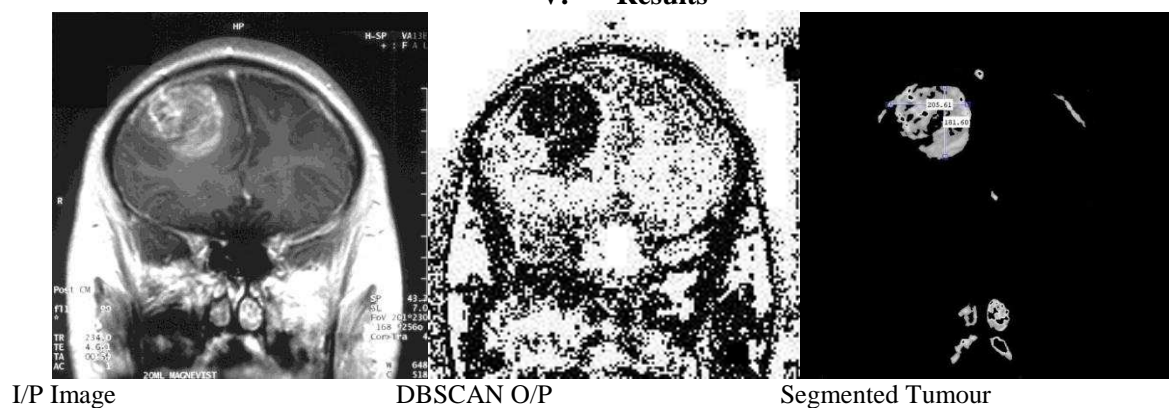
—To find a cluster, DBSCAN starts with an arbitrary point p and retrieves all points density-reachable from p with respect to Eps and $MinPts$. If p is a core point, this procedure yields a cluster with respect to Eps and $MinPts$ (see Argument: 2). If p is a border point then no points are density-reachable from p and DBSCAN visits the next point of the database.

DBSCAN: The Algorithm

1. Arbitrary select a point p
2. Retrieve all the points density-reachable from with respect to Eps and $MinPts$.
3. If p is a core point, a cluster is formed.
4. If p is a border point no points are density reachable from and DBSCAN visits the next points of the image.
5. Continue the process until all the points have processed.



V. Results



VI. Comparison of between k-means & DBSCAN

In our project, to segment GIS images we have used two algorithms. One is a standard one which is already in use and known as K-means and another one is DBSCAN algorithm which we have implemented for the first time in the image processing domain to segment GIS images, which is our point of interest.

Logical Comparison:

I. **Family of Clusters:** K-means is Centre (centroid) based Clustering approach. DBSCAN is density based approach. In case K-means algorithm we have seen that the algorithm is based on the initial value of K and formation of k number of centroid. Whereas in case of DBSCAN, it is completely based on density reachability of a point to others.

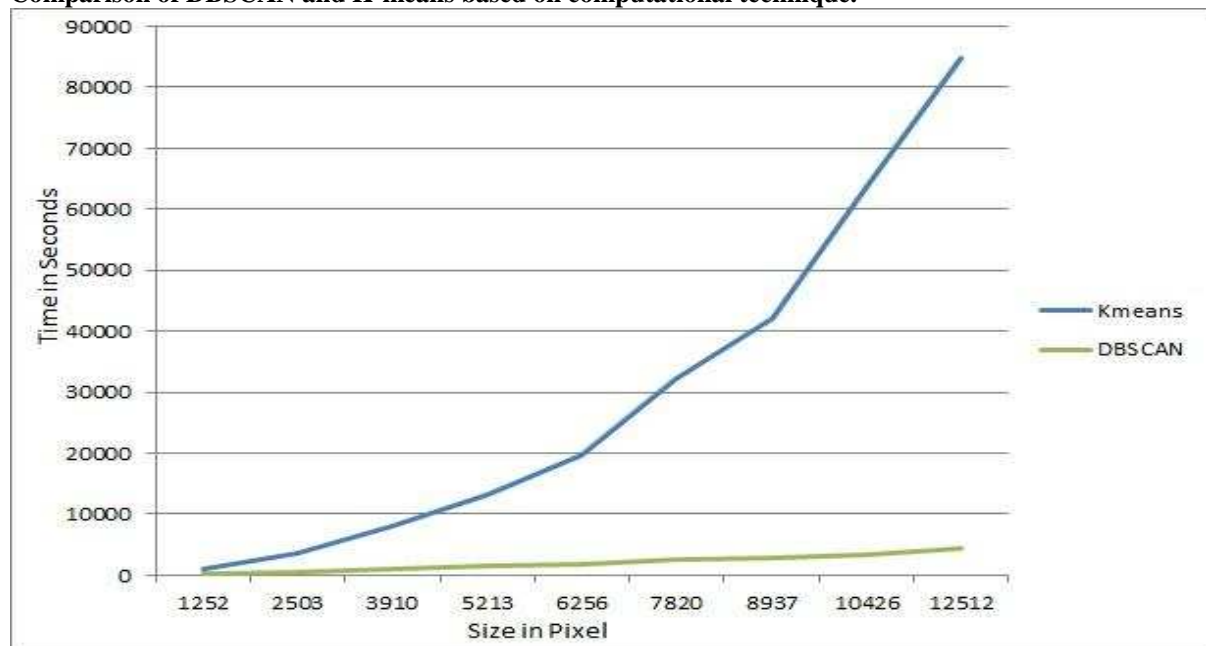
II. **Cluster Shapes:** K means cannot build Non-convex shaped cluster but there is no such constrain in DBSCAN. Even in case of K-means we are never going to know the real cluster, using the same data, if it is inputted in a different way may produce different cluster if the number of data is a few. So, we can have more proper segmented result is case of DBSCAN. DBSCAN can find arbitrarily shaped clusters. It can even find a cluster completely surrounded by (but not connected to) a different cluster. Due to the *MinPts* parameter, the so-called single-link effect (different clusters being connected by a thin line of points) is reduced.

III. **Predefined Cluster Numbers:** In K-means the number of cluster needs to be predefined (as K). Even we need to have some knowledge about the image we are using to initiate an effective k value. DBSCAN does not require one to specify the number of clusters in the data a priori, as opposed to k-means.

IV. **Density difference:** K-means algorithm does not depend on density. Whereas DBSCAN cannot cluster data sets well with large differences in densities, since the *MinPts* combination cannot then be chosen appropriately for all clusters.

V. **Outliers handling:** K means is sensitive to noise as small number of such data can substantially influence the mean value. But DBSCAN efficiently handles this problem. It has a notion of noise. In DBSCAN Noise is the set of points, in the database, that don't belong to any of the clusters.

Comparison of DBSCAN and K-means based on computational technique.



Graph1: run time in sec.

To summarize the table, the DBSCAN has an almost linear increase in computing time, relative to the number of points in the database. The K-means gain, however, is exponential and gets outperformed with a factor of 250 to 1900 in this test. The factor will continue to get bigger as the size of the database increases.

VII. CONCLUSION

This project has been successfully completed using various segmentation techniques. Sample Brain MRI images has been taken as test subjects and segmented using DBSCAN and K-Means algorithm. Our project targets were achieved successfully at theoretical and MATLAB Simulation level. The purpose of this paper is not to provide a definition of Brain MRI segmentation, however interesting debate may be, instead we examined the properties of a couple of algorithms when they are applied to this specific kind of images.

Reference

- [1] Tan, Pang-Ning. Steinbach, Michael. Kumar, Vipin. Introduction to Data Mining. Pearson Education, 2005. p. 487- 490.
- [2] Zhang Y et al (2007) A novel medical image segmentation method using dynamic programming. In: International conference on medical information visualisation-bioMedical visualisation, pp 69–74
- [3] Hall LO, Bensaid AM, Clarke LP, Velthuizen RP, Silbiger MS, Bezdek J (1992) A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *IEEE Trans Neural Netw* 3:672–682
- [4] Tian D, Fan L (2007) A brain MR images segmentation method based on SOM neural network. In: The 1st international conference on bioinformatics and biomedical engineering, pp 686–689
- [5] Han X, Fischl B (2007) Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE Trans Med Imaging* 26(4):479–486
- [6] Dunn, J.C., 1973.” A fuzzy relative of the ISODATA process and its use in detecting compact, well Separated clusters”,*Journal of Cybernetics*, 3: 32-15.
- [7] sai .C, Manjunath B.S,Jagadeesan.R.,”Automated Segmentation of brain MR Images”, Pergamon, Pattern Recognition, Vol 28, No 12, March 1995.
- [8] Bezdek, J.C., 1974. “Cluster validity with fuzzysets”, *Cybernetics*, 3: 58-73.
- [9] Ruspini, E., 1970. “Numerical methods for fuzzy clustering”, *Information Sciences*, 2: 319-350.
- [10] Kannan S.R.:”Segmentation of MRI Using New Unsupervised Fuzzy C mean Algorithm”,*ICGST*,2005.
- [11] Jayaram K.Udupa,Punam K.Saha,:”Fuzzy Connectedness and Image Segmentation”, *Proceedings of the IEEE*,vol.91,No 10,Oct 2003.
- [12] Vaidyanathan M, Clarke L.P, Velthuizen R.P, Phuphanich S, Bensaid A.M, Hall L.O, Bezdek J.C, Greenberg H, Trotti A, Silbiger ,:”Comparison of Supervised MRI Segmentation methods for Tumor Volume Determination During Therapy”,*Pergamon,Magnetic Resonance Imaging*,vol.13,no.5,pp,719-728,1995.
- [13] S. Murugavalli , V. Rajamani,” “A high speed parallel fuzzy c-mean algorithm for brain tumor segmentation”, *BIME Journal*, Volume (06), Issue (1), Dec., 2006.
- [14] S. Murugavalli , V. Rajamani,” An Improved Implementation of Brain Tumor Detection Using Segmentation Based on Neuro Fuzzy Technique” *Journal of Computer Science* 3 (11): 841-846, 2007
- [15] Parra, C.A., K. Iftekharuddin and R. Kozma, 2003.”Automated Brain Tumor segmentation and pattern recognition using ANN”, *Computational Intelligence Robotics and Autonomous Systems*.
- [16] Aidyanathan M, Clarke L.P, Velthuizen R.P, Phuphanich S, Bensaid A.M, Hall L.O, Bezdek J.C, Greenberg H, Trotti A, Silbiger M.:”Comparison of Supervised MRI Segmentation methods for Tumor Volume Determination During Therapy”,*Pergamon,Magnetic Resonance Imaging*,vol.13,no.5,pp,719-728,1995.
- [17] Chunyan Jiang,Xinhua Zhang,Wanjun Huang,Christoph Meinel.:”Segmentation and Quantification of Brain Tumor”,*IEEE International conference on Virtual Environment,Human-Computer interfaces and Measurement Systems*, USA, 12-14, July 2008.
- [18] HanShen,WilliamSandham,MalcolmGranet,Annette Sterr.:”MRI Fuzzy Segmentation of Brain Tissue Using Neighbourhood Attraction with Neural-Network Optimization”,*IEEE Transactions on Information Technology in Biomedicine*Vol.9,No.3,Sep 2005.
- [19] Hillips,W.E, Velthuizen R.P, Phuphanich S, L.O, Clarke L.P, Silbiger ,:”Application of fuzzy C-Means Segmentation Technique for tissue Differentlation in MR Images of a hemorrhagic Glioblastoma Multiforme “, *Pergamon,Megnetic Resonance Imaging*, Vol.13, 1995.
- [20] Kohonen..T “The self-organizing map”. *Proceedings of the IEEE*, (78(9)):1464–1480, 1990.
- [21] Pal, N.R. and S.K. Pal, 1993.” A review on image segmentation techniques”, *Pattern Recognition* 26(9): 1277-1294.
- [22] Phillips,W.E, Velthuizen R.P, Phuphanich S, L.O, Clarke L.P, Silbiger ,:”Application of fuzzy C-Means Segmentation Techniquefor tissue Differentlation in MR Images of a hemorrhagic Glioblastoma Multiforme “, *Pergamon,Megnetic Resonance Imaging*, Vol.13, 1995.
- [23] Squire LF, Novelline RA (1997). *Squire’s fundamentals of radiology* (5th ed.) Harvard University Press.
- [24] Linda G. Shapiro and George C. Stockman (2001): “Computer Vision”, pp 279-325, New Jersey, Prentice-Hall.
- [25] S. Wareld, J. Dengler, J. Zaers, C. Guttman, W. Gil, J. Ettinger, J. Hiller, and R. Kikinis. “Automatic identification of grey matter structures from mri to improve the segmentation of white matter lesions”. *J. of Image Guided Surgery*, 1(6):326{338}, 1995.

- [26] M. Sezgin and B. Sankur (2004). "Survey over image thresholding techniques and quantitative performance evaluation". *Journal of Electronic Imaging* 13 (1): 146–165.
- [27] Nobuyuki Otsu (1979). "A threshold selection method from gray-level histograms". *IEEE Trans. Sys., Man., Cyber.* 9 (1): 62–66.
- [28] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297.
- [29] Test case images from <http://radiopaedia.org/>.
- [30] Lin, W., E. Tsao and C. Chen, 1991. "Constraint satisfaction neural networks for image segmentation", In: T.Kohonen, K. Mäkelä, O. Simula and J. Kangas (eds.), *Artificial Neural Networks* (Elsevier Science Publishers), pp: 1087-1090.
- [31] Milan Sonka, Satish K. Tadikonda, Steve M. Collins, Knowledge-Based Interpretation of MR Brain Images, *IEEE Transaction on Medical Imaging*, Vol. 15, No. 4, 1996, pp. 443-452.
- [32] "Segmentation of Brain MRI Image—A Review" SK Bandhyopadhyay, *TU Paul IJARCSSE* 2 (3), 2012, pp. 409 - 413
- [33] "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96),
- [34] "Segmentation of Brain Tumor from Brain MRI Images Reintroducing K–Means with advanced Dual Localization Method", *TU Paul, SK Bandhyopadhyay, IJERA* 2 (3), 226 – 231
- [35] "Automatic Segmentation of Brain Tumour from Multiple Images of Brain MRI", SK Bandhyopadhyay, *TU Paul, IJAIEEM* 2 (2)