ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

DENSITY BASED FUZZY C MEANS (DBFCM) IMAGE SEGMENTATION

Arpan Garai[1], Kalyani Mali[2]

Computer Science Engineering, Kalyani Govt. Engineering College, Kalyani, India [1]

Computer Science Engineering, Kalyani University, Kalyani, India [2]

Abstract: The fuzzy c means image segmentation algorithm is mainly implemented taking the attribute as intensity. Often some of the relevant segments in different density are missed as they have same intensity which is not desired. But if the segmentation is done in density domain as well as intensity domain then far better result can be obtained. Here In this paper density based fuzzy c means (DBFCM) clustering is presented. It is divided into two steps first different dense region is found using k^{th} nearest neighbour then fuzzy c means segmentation is done on each dense region. The DBFCM is implemented upon some of the satellite images to get the segments as experimental results. Then it is compared with the conventional fuzzy c means approach.

Keywords: Density based segmentation; Image segmentation; k^{th} nearest neighbour distance; fuzzy c means; satellite image segmentation

I. INTRODUCTION

Image segmentation is very important and challenging problem and a necessary role in image analysis as well as in high-level image interpretation and understanding such as robot vision, object recognition, and medical imaging [1]. The goal of image segmentation is to partition an image into a set of disjoint regions with uniform and homogeneous attributes such as intensity, color, tone or texture, etc. Many different segmentation techniques have been developed till now. In this paper, a clustering based method for image segmentation will be considered.

Classification can be of two types i.e. supervised and unsupervised. Unsupervised classification is known as clustering. In supervised classification we need some prior information about the classification. But in unsupervised classification no prior information about the classification is needed i.e. It automatically generates the clusters.

Clustering is a process by which we can group together the objects such that the objects belongs to the same cluster will have same property but objects belongs to different cluster will have different property. There are mainly two types of clustering namely partitioning and hierarchical. At present other technique has been developed. Many of them are hybrid in nature. Nevertheless, based on the basic architecture clustering can be classified as density-based, grid based, model-based, sample-based etc. Density-based cluster methods are done based on density. It is believed that density-based cluster methods have the potential to reveal the structure of a spatial data set in which different point processes overlap. Ester et al. (1996) and Sander et al. (1998) introduced the approaches of DBSCAN and GDBSCAN to address the detection of clusters in a spatial database according to a difference in density [2].

In this paper an experimental study on density-based image clustering is done. Then we estimated the parameters of density-based clustering and hence deduced an algorithm to find the segment by setting property both densities as well as intensity. After that the method is applied into several satellite images. Fuzzy c means algorithm as a standard unsupervised algorithm also applied into those satellite images. The results of those two algorithms are compared. From that it can be found that fuzzy c means clustering fails to find any relevant cluster if it is covered by some different dense cluster whereas density-based clustering can easily find that. In case of density based fuzzy c means segmentation we don't need any external noise removal algorithms i.e. it can remove noise by itself. Whereas fuzzy c means algorithm can’t do it.

II. LITERATURE REVIEW

The aim of clustering is to group data into meaningful sub-classes (clusters). Previously it is being said that any clustering algorithms can be roughly classified into two categories namely hierarchical clustering and partition clustering. The partition clustering methods obtain a partition of objects into clusters such that the objects in a cluster are more similar to each other than to objects in other clusters. The hierarchical cluster method is a nested sequence of partitions; it starts by placing each object in its own cluster and then merges these atomic clusters into larger clusters until some termination condition is satisfied [3]. But in density-based methods differs from the partition and hierarchical methods. In case of density based classification method we concentrate on classifying dense objects into homogenous clusters with arbitrary

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

shape and size and remove noise using a density criterion. There are mainly two strategies, i.e. the grid-based and the distance-based, have been adopted for finding density homogeneous clusters in density-based methods. Density-based cluster methods are characterized by aggregating mechanisms based on density [9]. It is believed that density-based cluster methods have the potential to reveal the structure of a spatial data set in which different point processes overlap. The approaches of DBSCAN and GDBSCAN were introduced to address the detection of clusters in a spatial database according to a difference in density [2]. Since then, many modifications have been published. However, such methods have some drawbacks. In DBSCAN and its modifications it is required to define their parameters (for example, *Eps*, the distance used to separate clusters of different densities) in an interactive way. The parameters, estimated in this fashion, may lead to classification errors in terms of class type number and membership of each point, especially in complex scenarios.

Due to its power in estimating local densities, a kernel function is used within grid-based clustering methods to determine statistically significant clustering [5] [11] [12] [13]. Despite the computational speed, the grid-based methods suffer from a major drawback: the clustering results are sensitive to the grid partition scheme, i.e. the cell size in a grid. Choice of cell size may significantly affect the outcome of the analysis in terms of size, shape and significance of clusters.

In density-based cluster methods the clusters are defined by distance between a point and its neighbour, which reflects the density of the local area. The density based methods can avoid the computation of local density. Due to this advantage, many approaches have been proposed in density based clustering methods. Most of them are often based on the k^{th} nearest neighbour distance [5] [13] [14] [15].

The Fuzzy C-Means (FCM) clustering algorithm is one of the most known supervised clustering algorithms. It was first introduced by Dunn and later was extended by Bezdek. The algorithm is based on an iterative method of finding clusters.

III. Presentation of the Main Contribution of This Paper

There are several density-based algorithms in segmentation. But very few of them are experimented upon an image. Here in this paper a density based fuzzy c means image segmentation is presented. Then an experimental study is done of this density based fuzzy c means segmentation algorithm upon image and compared with fuzzy c means. The novelty of the density based fuzzy c means segmentation lies upon estimating the parameters i.e. the threshold point between high dense and low dense regions. And then apply FCM for each of the dense regions.

IV. Methodology

Now, in this section the density based fuzzy c means algorithm is presented. In which, the clusters and the noise based on density as well as intensity is found.

Basically in this method density-based segmentation is done first i.e. the various dense regions are found. Then for each dense region clustering is done based on intensity using fuzzy c means image segmentation.

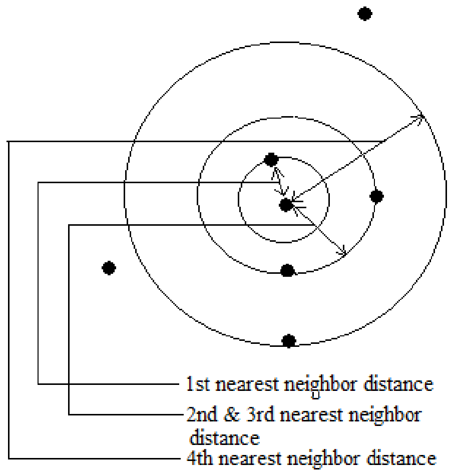

Fig. 1 k^{th} nearest neighbour distance

The very first step to find the various dense regions is to find the k^{th} nearest neighbour distances for each pixel. The k^{th} nearest neighbour distance is the basically the distance from the point to its k^{th} nearest neighbour as shown in Fig 1. So, clearly it can be said that the more the k^{th} nearest neighbour distance for the point is then the less in the dense region it belong to. Following that rule the density is found using k^{th} nearest neighbour distance and then the density based segments are calculated. Then after finding the k^{th} nearest neighbour distance it is sorted which may look like Fig. 2 for any real life images such as any satellite image, medical image etc. Here the sorted k^{th} nearest neighbour distance curve in Fig. 2 found after calculating k^{th} nearest neighbour distance of histogram equalised IRS-1A Calcutta image (band 4) shown in Fig. 4. It can be clearly seen from the pattern of the curve that after a certain point the slope is very high i.e. k^{th} nearest neighbour distance is very high.

In the next steps of the algorithm the valleys of the curve and the point after which k^{th} nearest neighbour distance is very high have to found out. So, meet that goal the slopes of each point of the curve have to be calculated. Then the maximum slope of that point has to be traced. That point will the point after which k^{th} nearest neighbour distance is

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

very high i.e. from that point to the last point of the curve is said as noise. From the pattern of the curve it is clearly predicted that the next maximum slope will be very near to current maximum. So, the next maximum point will be calculated from beginning to some distance from maximum slope point. The next maximum slope point to the previous slope point is said to be one valley. That way the algorithm continues until all the valleys are found. Every valley is said to be one dense region.

Now, after getting all the dense regions the fuzzy c means algorithm is applied to find all the segments in intensity domain. FCM is an iterative algorithm. In every iteration it produces an optimal c partition by minimizing the weighted within group sum of squared error objective function J_{FCM}.

$$J_{FCM} = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q \, d^2(x_k, v_i)$$

where c is the number of clusters and $X = \{x_1, x_2, \cdots, x_n\} \subseteq R_p$ is the data set in the p-dimensional vector space, n is the number of data items, c is the number of clusters with $2 \leq c < n$, u_{ik} is the degree of membership of x_k in the i^{th} cluster, q is a weighting exponent on each fuzzy membership, v_i is the prototype of the centre of cluster i, d2 (x_k , v_i) is a distance (Y. Yang, S. Huang) measure between object x_k and cluster centre v_i. The entire algorithm is as follows:

1. Take the image as input.
2. Make that image histogram equalized.
3. Initialize k, end_th, j = 1.
4. Find the K-nearest neighbour distance of all the pixels and put it in an array A[0,n]
5. Sort A in ascending order.
6. Calculate the Slope of array A i.e. Slope_A[0,n].
7. The maximum slope of Slope_A = max_index.
8. The point from the maximum slope point to last of A is said to be the noise.
9. max_index_prev = max_index
10. Find the maximum A[0, max_index_prev – end_th] put the index into max_index_next.
11. The point from max_index_next up to max_index_prev previous of A is said to be a dense region cj, j=j+1.
12. max_index_prev = max_index_next
13. Go to step 10 untill max_index_next = 0
14. For each cj, j= 1 to m, where m is the number dense region

(a) Set values for c, q and eps.

Where c is the number of clusters, q is the weighting exponent.

(b) Initialize the fuzzy partition matrix $U = [u_{ik}]$.

(c) Set the loop counter b = 0.

(d) Calculate the c cluster centers $\{v_i^{(b)}\}$ with $U^{(b)}$

$$v_i^{(b)} = \frac{\sum_{k=1}^{n} (u_{ik}^{(b)})^q \, x_k}{\sum_{k=1}^{n} (u_{ik}b)^q}$$

(e) Calculate the membership $U^{(b+1)}$. For k = 1 to n, calculate the following:

$I_k = \{ i \mid 1 \leq i \leq c, d_{ik} = x_k - v_i = 0 \}, I;$

for the k^{th} column of the matrix, compute new membership values:

i. if $I_k = \varphi$, then

$$u_{ik}(b+1) = \frac{1}{\sum_{j=1}^{c} (d_{ik} / d_{jk})^{2/(q-1)}}$$

(f) If $U^{(b)} - U^{(b+1)} <$ eps , stop; otherwise, set b = b + 1 and go to step (d).

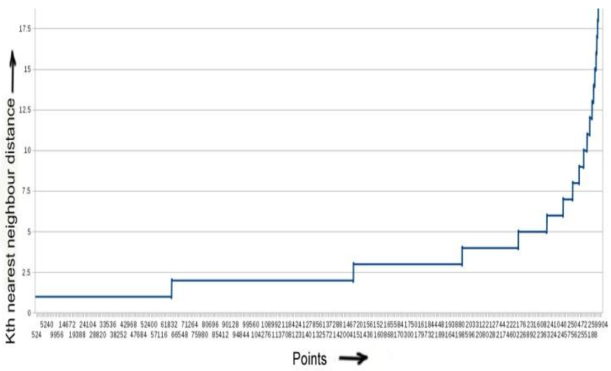

Fig. 2 Kth nearest distance curve

V. FINDINGS

In this section the results of the density based fuzzy c means algorithm is shown and compared with the fuzzy c means experimentally.

Here the segmentation results of the density based segmentation on IRS-1A (band 4) Calcutta image Fig.3 is presented. The poor illumination led to a degraded visibility of the actual object classes present. Therefore we display in Fig.4 the histogram-equalized enhanced versions of the calcutta image in band 4, highlighting the different object regions for the convenience of the readers. These regions correspond to water, concrete, vegetation, habitation, open space, and roads (including bridges).

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

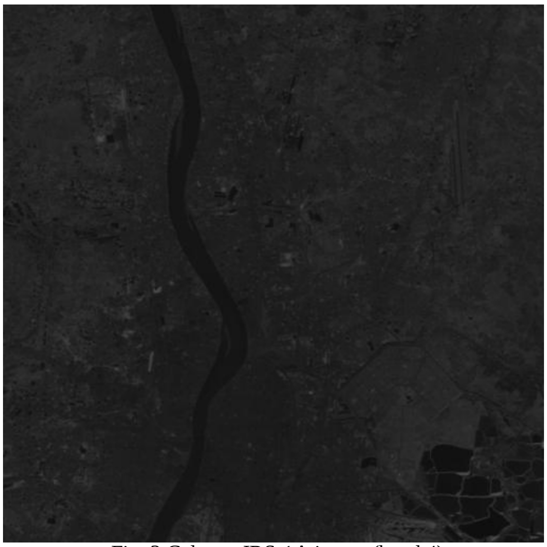

Fig. 3 Calcutta IRS-1A image (band 4)

Fig. 4 Histogram equalized image

The segmented results, however, were not enhanced and correspond to a processing involving all the bands at the input. The outcomes of the density based fuzzy c means image segmentation of calcutta image is shown in Fig.6 and Fig. 7. The result of FCM algorithm is shown in Fig. 5.

Here k and end_th taken as 4, (total points/2) i.e. 131072 respectively. Then k-nearest neighbour distance for each pixel is found out. Then all the kth nearest neighbour distance is sorted. It looked like Fig. 2. Then slope of each point is calculated. It can be seen that the first maximum slop of that curve is at the point 262140. So, the points with index value from 262140 to 262143 are said to very low dense region or noise. Now, next maximum point according to the algorithm is at 62947. So, the points with index value 62947 to 262140 belong to next higher dense. And finally 0 to 62946 belongs to the highest dense region. Now in these regions the clustering is done based on intensity using fuzzy c means algorithm.

Firstly in the noise part no farther segmentation is needed so it is simply discarded. Then into the points in the next high dense regions the fuzzy c means is implemented initializing c, q and eps as 5, 2 and 0.001 respectively. After applying fuzzy c means in the various dense regions output can be found in low dense and high dense regions as shown in Fig 6 and Fig. 7 respectively.

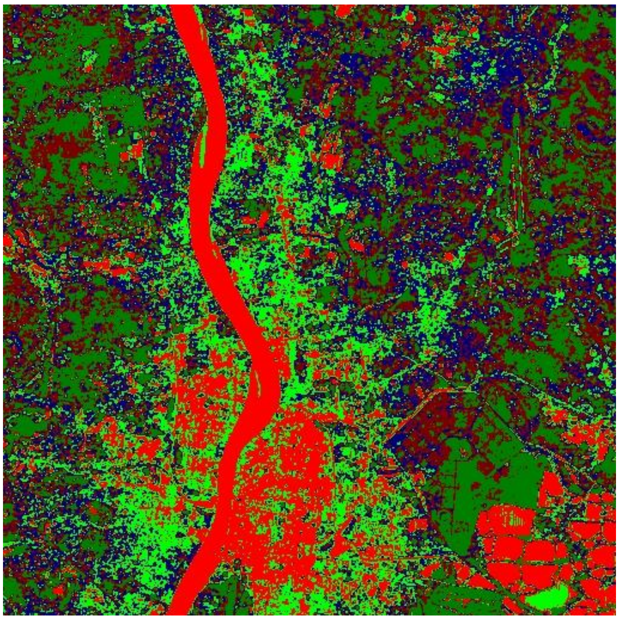

Fig. 5 output of fuzzy c means (5 clusters)

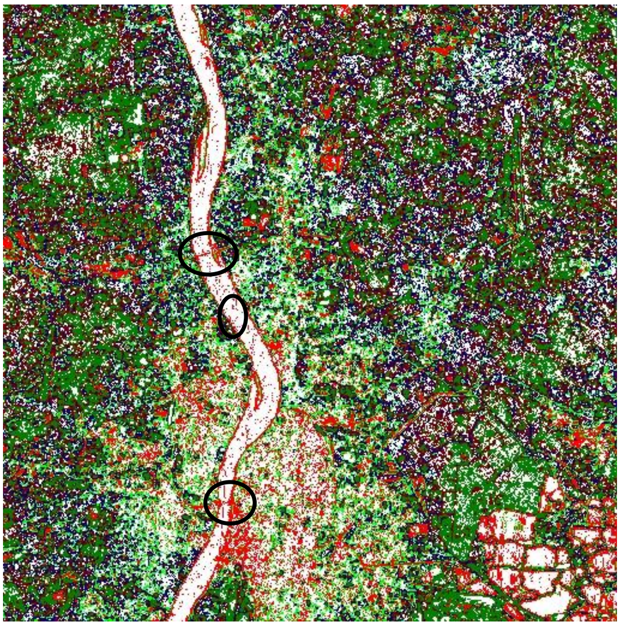

Fig. 6 Low dense output of DBFCM (5 clusters)

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

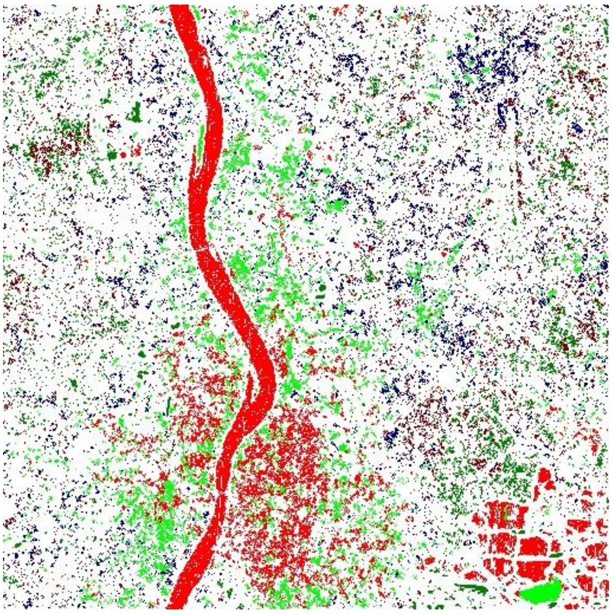

Fig. 7 High dense output of DBFCM (5 clusters)

VI. DISCUSSION

In the Fig.6 and Fig.7 those which are in the coloured regions are segmented and those are in the white are the points belongs to different dense regions or noise. The fig.5 image is the lower dense segment is pointed out. Here the bridge is perfectly segmented as marked in the lower dense segment i.e. Fig.6 where as in FCM [Fig.5] the bridge is lost. In case of density based FCM segmentation [Fig.6] some small sand islands in the river are traced where as conventional FCM [fig.5] fails to do that. The Davies–Bouldin is a very accepted index to identify how well the clustering is done. So, here in this index gives 1.53 and 1.54 for low dense and high dense (i.e. Fig.6 & Fig.7) respectively where as in case of FCM Fig. 5 gives value 1.54. So, it can be clearly said that the result of density based segmentation much better that result of FCM.

VII. LIMITATIONS

The time complexity of density based segmentation part is quite more. So, further work in this algorithm may be to make the running time lesser. If the algorithm can be done in a divide and conquer manner then we may get better results. If it is done in parallel then better result as well as better running time can be obtained. Here, some user defined parameters is taken. The some more experimental study has to be done to get better results. As, here segmentation is done in different dense regions it can be used in other situations.

VIII. CONCLUSIONS

The density based fuzzy c means segmentation algorithm is presented here and also a comparative study is done with well known unsupervised segmentation algorithm i.e. fuzzy c means algorithm. So as a result what can be clearly seen after that study is that it is very helpful in segmenting various dense objects in an image which so far was undone by the segmentation algorithms developed till now. But on the other hand it is also seen that the density based segmentation is time consuming. But it can give better results if done in divide and conquer manner and done in parallel. Farther work in this domain can give better results also.

REFERENCES

[1] Yong Yang and Shuying Huang "IMAGE SEGMENTATION BY FUZZY C-MEANS CLUSTERING ALGORITHM WITH A NOVEL PENALTY TERM" in Computing and Informatics, Vol. 26, 2007, 17–31
[2] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" in Institute for Computer Science, University of Munich Oettingenstr. 67, D-80538 München, Germany
[3] Kaufman L and Rousseeuw, "Finding Groups in Data: an Introduction to Cluster Analysi "in John Wiley & Sons 1990.
[4] Sushmita Mitra, Partha Pratim Kundu, "Satellite image segmentation with Shadowed C-Means in international conference of Elsevier doi:10.1016/j.ins.2011.04.027 april 21, 2011
[5] Tao Pei, Ajay Jasra, David J. Hand, Xing Zhu, Chenghu Zhou "DECODE: a new method for discovering clusters of different densities in spatial data", in Data Min Knowl Disc (2009) 18:337–369 DOI 10.1007/s10618-008-0120-3
[6] Sushmita Mitra , Witold Pedrycz and Bishal Barman, "Shadowed c-means: Integrating fuzzy and rough clustering" in Pattern Recognition 43 (2010) 1282–1291
[7] Raymond T. Ng and Jiawei Han, "CLARANS: A Method for Clustering Objects for Spatial Data Mining" in IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 14, NO. 5, SEPTEMBER/OCTOBER 2002
[8] Anindya Halder, Susmita Ghosh and Ashish Ghosh, "Ant based supervised and unsupervised land use map generation from remotely sensed images" in international conference of IEEE p.p 158 – 163, Aug, 2009
[9] Han JW, Kamber M, Tung AKH, "Spatial clustering methods in data mining" in Miller HJ, Han JW (eds) Geographic data mining and knowledge discovery Taylor & Francis, London, pp 188–217
[10] Dr. Malay Kumar Pakhira, "Digital image processing"
[11] Diggle PJ, "A kernel method for smoothing point process data" in Appl Stat 34:138–147. doi:10.2307/2347366
[12] Hinneburg A and Keim DA," An efficient approach to clustering in large multimedia databases with noise" In Proceedings of the knowledge discovery and data mining, pp 58–65
[13] Tran TN, Wehrensa R, Lutgarde MCB (2006) KNN-kernel density-based clustering for high-dimensionalmultivariate data. Comput Stat Data Anal 51:513–525. doi:10.1016/j.csda.2005.10.001
[14] Markus MB, Kriegel H-P, Raymond TN, Sander J," identifying density-based local outliers",.In Proceedings of ACM SIGMOD of 2000 international conference on management of data, vol 29, pp 93–104
[15] Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proc. 2nd int. conf. on knowledge discovery and data mining, Portland, OR, pp 226–231

BIOGRAPHY

Kalyani Mali received B.Tech. and M.Tech. Degree in 1987 and 1989 from University of Calcutta, India, and Ph.D. Degree in Computer Science and Engineering from Jadavpur University, Kolkata, India in 2005. She joined the faculty of Department of Computer Science and Engineering, University of Kalyani in 1992. She currently holds the rank of Associate Professor. Her current research interest includes pattern recognition, image processing, data mining and soft computing.

Arpan Garai is presently working as Project Linked person in ECSU department ISI, Kolkata. He has received M Tech degree from Kalyani Govt. Engineering College, West Bengal University of Technology in year 2013. Before that he done his BE degree from University Institute of Technology, Burdwan University in year 2011. His main areas of interest are image segmentation, data mining and soft computing.