International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 4, Issue 4, July - August 2015 **ISSN 2278-6856**

A Novel Clustering Algorithm for Color Image Segmentation using Rough set and Fuzzy set

Mr. Venkateswara Reddy Eluri[1], Dr. M.Ramesh[2] and Dr. Mare Jane Aragon[3]

[1]Shinas College of Technology, Shinas,
Ministry of Man Power, Sultanate of Oman, Oman

[2]R.V.R &J.C College of Engineering & Technology,
Chowdavaram, Guntur, AP, India.

[3]Shinas College of Technology, Shinas,
Ministry of Man Power, Sultanate of Oman, Oman

Abstract

Clustering is a process which partitions a given data set into homogeneous groups based on given features such that similar objects are kept in a group whereas dissimilar Objects are in different groups. This paper describes a proposed algorithm of Dynamic Histogram based Rough-Fuzzy C-Means (DHRFCM), in which identifying initial seed points by constructing dynamic histogram rather than randomly generated initial seed points for identifying similar objects. After that using rough set theory, we can reduce the seed point selection and then apply an un supervised clustering algorithm of Fuzzy C-Means(FCM) to segment the given color image. And also compared the proposed algorithm with various clustering algorithm such as K-means, Fuzzy C-Means, Rough-Fuzzy C-Means. To determine the best clustering algorithm we used various cluster validity indices such as Davies-Bouldin (DB) index, RAND index, Silhouette index and Jaccard index. Experimental results shows that the proposed method perform well and improve the segmentation results.

Keywords: Clustering, Rough set, Dynamic Histogram, Image Segmentation, Fuzzy C-Means.

1. INTRODUCTION

Clustering analysis has been an emerging research issue in data mining due to its variety of applications. Main problem with any un supervised learning clustering technique is that it can't be standardized. Some algorithms are best suitable to some particular linear data, but it can't able to segment the non-linear data. Before exploring various clustering algorithms in detail let's have a brief overview about what is the goal of clustering. The goal of clustering is to group similar objects in one cluster and dissimilar objects in different clusters.

In this paper we propose a novel algorithm, termed as Dynamic Histogram based Rough-Fuzzy C-Means (DHRFCM) using dynamic histogram, rough sets and fuzzy sets. The membership function of fuzzy sets enables efficient handling of overlapping partitions. The concept of lower and upper approximations of rough sets deals with uncertainty and incompleteness in class definition. Each partition is represented by a set of tree parameters namely a cluster seed point, a crisp lower approximation, and a fuzzy boundary.

The structure of the rest of this paper is as follows. Section 2 describes the Fuzzy C Means (FCM) traditional clustering algorithm. In section 3 we describe Rough Fuzzy C-Means (RFCM) clustering algorithm, section 4 describes proposed Dynamic Histogram based Rough-Fuzzy C-Means (DHRFCM) Clustering Algorithm. In section 5 we describe various cluster validity indices to measures the performance of clustering algorithm such as DB Index, XB Index, RAND index and Jaccard Index.

2. FUZZY- C MEANS (FCM) CLUSTERING ALGORITHM

Set your page as A4, width 210, height 297 and margins as follows:

The algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center clearly summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated according to the formula:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left[\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right]^{\frac{2}{m-1}}}$$

$$c_j = \frac{\sum_{i=1}^{n} \mu_{ij}^{m} x_i}{\sum_{i=1}^{n} \mu_{ij}^{m}}$$

Where 'n' is the number of data points
'm' is the fuzziness index
'c' represents the number of cluster center
'C_j' represents j^{th} cluster center
'X_i' represents i^{th} data item
'μ_{ij}' represents the membership of i^{th} data to j^{th} cluster center.

FCM Algorithm

Step 1: Initialize U= [μ_{ij}] matrix, u [0]
Step 2: At k-step: calculate the center vectors c(k) =[cj] with U(K).

$$c_j = \frac{\sum_{i=1}^{n} \mu_{ij}^m x_i}{\sum_{i=1}^{n} \mu_{ij}^m}$$

Step 3: Update U(K), U(k+1)

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left[\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right]^{\frac{2}{m-1}}}$$

Step 4: if $\|U(K+1) - U(K)\| <$ Threshold then STOP; otherwise return to step2.

3 Rough-Fuzzy C-Means (RFCM) Clustering Algorithm

Step1: Choose initial means z_i , $1 \le i \le k$ for the k clusters.

2. Compute u_{ij} by above equation for k clusters and n data objects.

3. Assign each data point x_j to the lower approximation $\underline{BU}_i$ or upper approximation $\overline{BU}_i$, $\overline{BU}_{i'}$ of cluster pairs U_i , $U_{i'}$ by computing the difference membership $u_{ij} - u_{i'j}$

4. Let u_{ij} be maximum and $u_{i'j}$ be the next to maximum. If $u_{ij} - u_{i'j}$ is less than some threshold then $x_j \in \overline{BU}_i$ and $x_j \in \overline{BU}_{i'}$ x_j can't be a member of any lower approximation else $x_j \in \underline{BU}_i$ such that the membership u_{ij} is maximum over the k clusters.

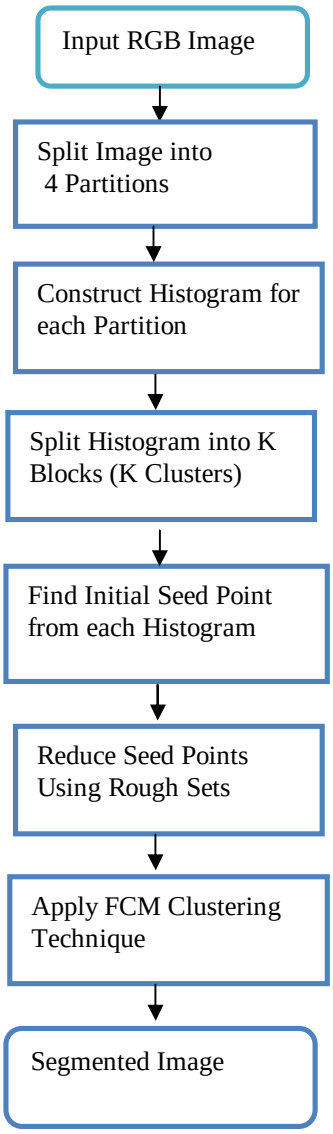
5. Compute the new mean z_i for each cluster as,

$$z_i = \begin{cases} \frac{\sum_{x_j \in (\overline{BU}_i - \underline{BU}_i)} \mu_{ij}^{\alpha} x_j}{\sum_{x_j \in (\overline{BU}_i - \underline{BU}_i)} \mu_{ij}^{\alpha}} \\ W_{low} \times \frac{\sum_{x_j \in \underline{BU}_i} \mu_{ij}^{\alpha} x_j}{\sum_{x_j \in \underline{BU}_i} \mu_{ij}^{\alpha}} + W_{up} \times \frac{\sum_{x_j \in (\overline{BU}_i - \underline{BU}_i)} \mu_{ij}^{\alpha} x_j}{\sum_{x_j \in (\overline{BU}_i - \underline{BU}_i)} \mu_{ij}^{\alpha}} \\ \frac{\sum_{x_j \in \underline{BU}_i} \mu_{ij}^{\alpha} x_j}{\sum_{x_j \in \underline{BU}_i} \mu_{ij}^{\alpha}} \end{cases}$$

6. Repeat steps 2-5 until convergence, i.e., there are no more new Assignments or upper limit on the number of iterations are reached.

4 Dynamic Histogram based Rough-Fuzzy C-Means (DHRFCM) Clustering Algorithm

4.1 Flow Chart



4.2 Proposed Algorithm

In the existed rough set based fuzzy c means algorithm, initial seed points will consider randomly. This will take time consuming and will not get more accuracy. Instead of selecting seed points randomly, we select initial seed points by constructing the histogram for three components and then split each component as 'C' groups (required no of clusters C). Consider peak values of every group as initial seed point. The proposed Method, Histogram based RFCM consists of the following steps:
Step1: Read an RGB image as input
Step2: Split Image into specified no of partitions.
Step3: Construct histogram for each partition, and then split it into 'C' blocks (C is no of clusters required).
Step4: Consider peak values of every block as initial seed point of that particular cluster.

Step5: Apply Rough sets concepts to reduce the seed points
Step6: and then use Fuzzy C Means (FCM) algorithm to segment the image
Step 7: Get Segmented image as output.
Step 8: Find the cluster validity indices such as DB index, Rand index, XB index, etc., to identify the best segmentation results.

5 RESULTS & DISCUSSION

The performance of proposed DHRFCM algorithm is compared with that of different clustering algorithms such as K-Means, Fuzzy C Means (FCM), and Rough-Fuzzy C Means (RFCM). The proposed work is carried out on a system with Pentium V Processor performing at a speed of 3.0 GHZ with windows platform and 80GB HDD. Though we used windows platform, the work can be integrated for other platforms also. Computers running with high speeds are preferred for high speed and reliable segmentation. Entire programming and experimentation work is carried out by using MATLAB 9.0.

Fig 5-a Original RGB Image of size 384 x 256 (Corel Data Set – 508)

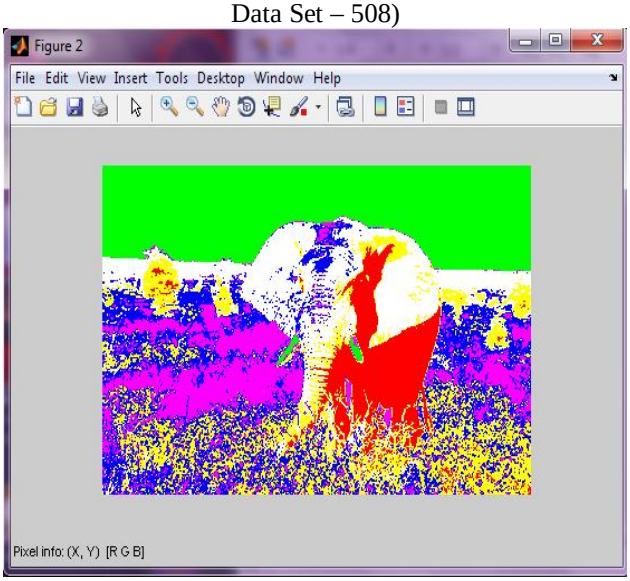


Fig 5-b Segmented Image Using Dynamic Histogram Based RFCM

Fig 5©Original RGB Image BSD-86000

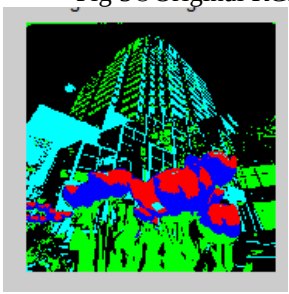


Fig 5(d)

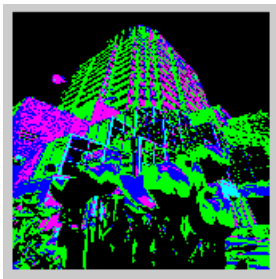


Fig 5 (e)

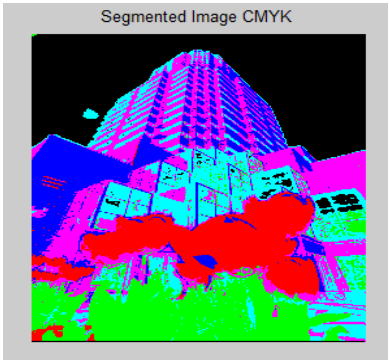
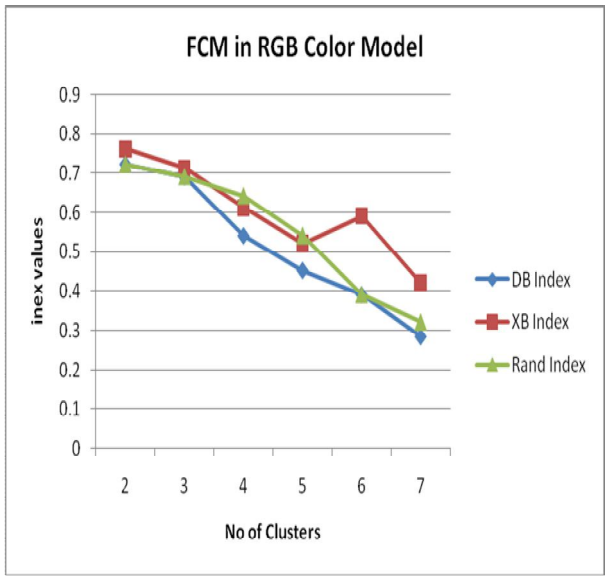


Fig 5 (f)

Fig 5(d) : segmented image using FCM
Fig 5(e) : segmented image using RFCM
Fig 5(f) : segmented image using DHRFCM

Table-1 DB, XB and Rand Index values of RGB Image Using Fuzzy C Means Clustering Algorithm

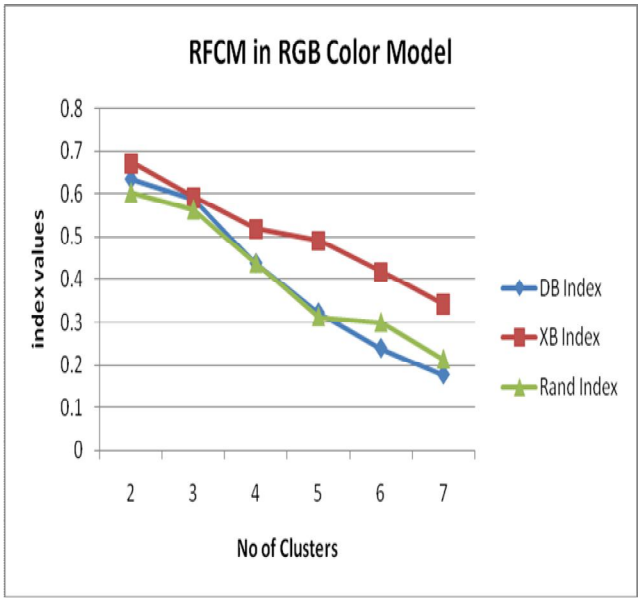
No of Clusters	DB Index	XB Index	Rand Index
2	0.7231	0.7621	0.7236
3	0.6912	0.7121	0.6924
4	0.5412	0.6125	0.6432
5	0.4532	0.5213	0.5421
6	0.3912	0.5912	0.3921
7	0.2845	0.4213	0.3215



Graph-1: Represents DB, XB and Rand Index Values of RGB Segmented Image using FCM

Table-2: DB, XB and Rand Index values of RGB Image Using Rough-Fuzzy C Means Clustering Algorithm

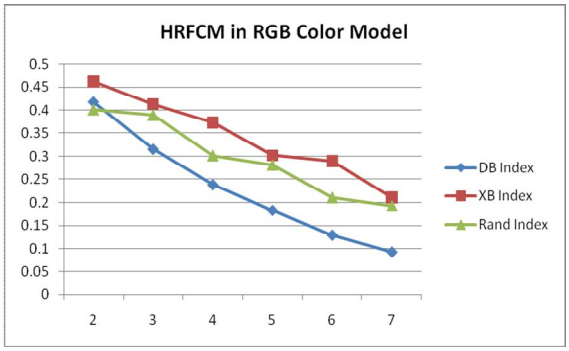
No of Clusters	DB Index	XB Index	Rand Index
2	0.0102	0.0105	0.0523
3	0.0127	0.0094	0.0456
4	0.0113	0.0083	0.0347
5	0.0083	0.0075	0.0134
6	0.0061	0.0065	0.0104
7	0.0042	0.0046	0.0087



Graph-2: Represents DB, XB and Rand Index Values of RGB Segmented Image using RFCM

Table-3 DB, XB and Rand Index values of RGB Image Using Dynamic Histogram based Rough-Fuzzy C Means Clustering Algorithm

No of Clusters	DB Index	XB Index	Rand Index
2	0.0102	0.0094	0.0314
3	0.0084	0.0083	0.0257
4	0.0073	0.0067	0.0115
5	0.0066	0.0043	0.0096
6	0.0058	0.0021	0.0078
7	0.0028	0.0015	0.0046



Graph-3: Represents DB, XB and Rand Index Values of RGB Segmented Image using DHRFCM

6.Conclusion

In this paper, we have proposed a new clustering algorithm called Dynamic Histogram based Fuzzy C Means (DHFCM). Instead of selecting seed points randomly, we calculated initial seed points by using dynamic histogram, and then using this initial seed points we applied Rough Fuzzy C Means Clustering Algorithm.

The results are obtained from two data sets, an RGB image (384 x 256) from Corel Data Set and IRIS data from BSDS data set. We have calculated validity indices of DB Index, XB Index and RAND Index to find the accuracy of segmentation.

The validity indices of DB, XB and RAND Indices for the Rough Fuzzy C Means Clustering Algorithm (RFCM) of an RGB Image are 0.0042, 0.0046 and 0.0087 respectively. The validity indices of DB, XB and RAND indexes for the proposed Dynamic Histogram based Rough Fuzzy C Means (RFCM) Clustering algorithm of an RGB Image are 0.0028, 0.0015 and 0.0046 respectively. The clustering algorithm that produces a least DB, XB or RAND index is considered the best algorithm. Therefore, the proposed Dynamic Histogram based Rough Fuzzy C Means (RFCM) Clustering algorithm is the best clustering algorithm compared to all the above said Clustering Algorithms.

References

[1] E. Venkateswara Reddy , Dr E.S.Reddy "Image Segmentation Using Rough Set Based Fuzzy K-Means Clustering algorithm" Global Journals Inc, USA, GJCST, Volume 13 Issue 6 Version 1.0 , May 2013, pp. 23-28.

[2] E.Venkateswara Reddy, Dr E.S.Reddy "Image Segmentation Using Rough Set Based Fuzzy C-Means Clustering algorithm" International Journal of Computer Applications , USA, (IJCA), Volume 74 - No.14 , July 2013, pp. 23-28.

[3] Mr.E.Venkateswara Reddy, Dr E.S.Reddy "A Comparative Study of Color Image Segmentation Using Hard, Fuzzy, Rough Set Based Clustering Techniques" Council for Innovative Research, IJCT, Volume-11, No-8 , November 2013, pp. 2873-2878.

[4] P. Dhanalakshmi & T.Kanimozhi: "Automatic Segmentation of Brain Tumor using K-Means Clustering and its Area Calculation", IEEE Transsactions on Fuzzy Systems, pp. 2278-2285, Apr. 2013.

[5] Arnau Oliver, Xavier Munoz, Joan Battle: "Improving Clustering Algorithms for Image Segmentation using Contour and Region Information", IEEE Transactions on Fuzzy Systems, 2009.

[6] Dhanalakshmi, Hannah Inbarani. H : "Fuzzy Soft Rough K-Means Clustering Approach For Gene Expression Data" IEEE Transactions on System, Man, and Cybernetics, pp. 1123-29, 2007.

[7] Bezdek, J. C., Pal, N. R.: "Some New Indexes for Cluster Validity", IEEE Transactions on System, Man, and Cybernetics, Part B, Cybernetics, 28, 1988, pp. 301–315.

[8] Dubois, D., H.Prade: "Rough Fuzzy Sets and Fuzzy Rough Sets", International journal of General Systems, 1990, pp. 191–209.

[9] Ali Salem Bin Samma and Rosalina Abdul Salam: "Adaptation of K-Means Algorithm for Image Segmentation", International Journal of Information and Communication Engineering, 2009, pp. 270 – 275.

[10] Chih-Tang Chang, Jim Z. C. Lai And Mu-Der Jeng : "A Fuzzy K-means Clustering Algorithm Using Cluster Centre Displacement", journal of information science and engineering , pp. 995-1009, 2011.

[11] Pradipta Maji and Shankar K. Pal, Fellow, IEEE: "Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices", IEEE transactions on systems, man, and cybernetics—part b: cybernetics, vol. 37, no. 6, December 2009.

[12] Subrajeet Mohapatra, Dipti Patra, and Kundan Kumar: "Unsupervised Leukocyte Image Segmentation Using Rough Fuzzy Clustering", International Scholarly Research Network ISRN Artificial Intelligence, Article ID 923946, pp.415-421, 2012.

AUTHOR

Venkateswara Reddy Eluri doing his PhD from Nagarjuna University and he done his M.E in Computer Science Engineering from Sathya bama University, India in 2003. He has 14 Years of teaching experience and currently working as a Lecturer in the Information Technology Department of Shinas College of Technology, Shinas, Ministry of Manpower, Sultanate of Oman, Oman. His research interests are in Image processing, Neural Networks and Data Mining. He is a life member for Computer Society of India (CSI). He has authored a good number of research Papers published in International Journals and Conference Proceedings like IEEE and Springer.

Dr. M.RAMESH received Ph.D from Acharya Nagarjuna University, India in 2014.
He received MCA in 1996 and M.Tech in 2004, both from Andhra University, India. He is currently working as an Associate Professor in Information Technology of RVR & JC College of Engineering, India. He has 16 years on teaching experience and worked in computer science and engineering, information technology and computer applications departments. His main research interests include image processing, data mining, machine learning and big data. He is life member of ISTE and CSI.

Dr. Mare Jane Aragon, currently working as a Head of the Department, Information Technology in Shinas College of Technology, Ministry of Manpower, Sultanate of Oman.