Color Image Segmentation using Rough Set based K-Means Algorithm

Amiya Halder
Department of CSE
St.Thomas' College of Engg. and Technology
Kolkata, West Bengal, India

Avijit Dasgupta
Department of ECE
St.Thomas' College of Engg. and Technology
Kolkata, West Bengal, India

ABSTRACT
This paper describes a rough set approach for color image segmentation that can automatically segment an image to its constituents parts. The aim of the proposed method is to produce an efficient segmentation of color images using intensity information along with neighborhood relationships. The proposed method mainly consists of spatial segmentation; the spatial segmentation divides each image into different regions with similar properties. Proposed algorithm is based on a modified K-means clustering using rough set theory (RST) for image segmentation, which is further divided into two parts. Initially the cluster centers are determined and then in the next phase they are reduced using RST. K-means clustering algorithm is then applied on the reduced and optimized set of cluster centers with the purpose of segmentation of the color (R,G,B components) images. The existing clustering algorithms namely the K-means and the Fuzzy C-Means (FCM) requires initialization of cluster centers whereas the proposed scheme does not require any such prior information to partition the exact regions. This rough set based image segmentation scheme results in satisfactory segmented image and Validity Index (VI) which is better than other state-of-the-art image segmentation.

General Terms
Image Processing, Pattern Recognition, Algorithms.

Keywords
Image Segmentation, RGB images, Rough Set, K-means algorithm.

1. INTRODUCTION
Image segmentation is one of the most challenging tasks in image analysis and a one of the very important pre-processing steps in low level computer vision as well as high level computer vision. It is also useful in the field of pattern recognition.

In image segmentation an image is divided into different regions with similar features. There are many different types of approaches of image segmentation. Edge-based method, region-based techniques and threshold-based techniques and so on. Threshold-based techniques depend on grayscale information of pixels in each plane of color (R,G,B plane) image. They do not need any priori information of the image. This technique has one limitation that they do not consider any spatial information. Edge-based methods distinguish different regions in image or objects through boundary detection. In region based methods, image is partitioned into connected regions by grouping neighboring pixels of similar intensity levels in each plane of a image. Images are partitioned according to their global feature distribution by clustering based image segmentation methods. In this paper, a color image segmentation method based on K-means using rough set theory is proposed, in which pixels are clustered according to the intensity and spatial features and then clusters are combined to get the results of final segmentation.

K-means clustering [1, 2] is an elegant method. It requires less computation. So this method has got certain advantages. But it has also got some disadvantages. The main disadvantage that we improved by using rough set theory is that it needs some initial cluster center set. So the no. of cluster centers and cluster center set is required. However if these are chosen incorrectly K-means algorithm may not converge. So we introduced rough set theory. By using rough set theory we optimized the no. of clusters and the value of the cluster centers which ultimately results in better segmentation of a color image.

Rough set theory proposed by Pawlak is a mathematical tool to analyze vagueness and uncertainty inherent making decision [3,4]. It does not rely on accessional information data set, and it analyzes and discovers reliant relation among data just from the point of view of data's discreditable attribute, just based on the concept of an upper and lower approximation set, as well as approximation space and models of sets. For the sake of this reason, it can provide more convenience for knowledge reduction than other methods and also superiority to other method in analyzing vagueness and uncertainty inherent making decision.

Based on the above theory we introduced Rough set theory into K-means algorithm for color image segmentation. By using rough set theory the redundant cluster center set is optimized and then the optimal cluster center set is given to the K-means algorithm for segmentation of image. Each plane is segmented by Rough set based K-means algorithm and then they are combined to get the final segmented image. This proposed method demonstrated efficiency and effectiveness.

This proposed algorithm is an elegant and simple scheme to identify the spatial segmentation of color image. This paper presents in fact a possible extension of the previously presented concept of RST based Image Segmentation [13] of gray level images.

The rest of the paper is organized as follows. In section 2 Literature review is given. In section 3 Rough set is described. In section 4 rough set based K-means algorithm is proposed. In section 5 we have shown the experimental results and in section 6 some conclusions have been made.

2. LITERATURE REVIEW
There are different types of image segmentation methods that have been proposed over the years. We will be giving some overview of some of these methods in this section. Advantages and disadvantages are also discussed in this section.

Many thresholding based image segmentation methods can be found in the literature. One such technique proposed by Suzuki and Toriwaki in [5] which proposes a knowledge guided thresholding technique for brain tumor segmentation. But the problem with this method is associated with the selecting threshold value for better segmentation of an object. Selection of wrong threshold can ignore the region of interest (ROI) completely. So this kind of techniques is not reliable.

An improved cluster region generation was proposed by different algorithms like K-means, Fuzzy C-Means etc. However determining the initial cluster center is a serious drawback of these region based segmentation algorithms. Few studies have been employed over the year to overcome this drawback like genetic algorithm based clustering [7,8], particle swarm optimization based clustering [6] etc. In recent year clustering techniques has been incorporated in image segmentation for detection of region of interest (ROI), object detection, skin segmentation etc. As a general purpose segmentation method, feature space clustering method has the advantage that it is straight forward classification. But the drawback of the standard algorithms is the predetermination of initial clusters. There is large no of segmentation algorithms present in literature, but there is no single segmentation algorithm that can be considered good for every image. Algorithm developed for a class may always_not be good for other class of images. One of the most commonly used technique for image segmentation in K-means algorithm [1]. This method is very simple and efficient for region based clustering. But the main drawback of this method is that if initial clusters are chosen wrongly this algorithm will not converge. As this is the simplest algorithm among the segmentation algorithms so K-means algorithm was chosen.

In this paper a hybrid color image segmentation algorithm using rough set and K-means algorithm for color image segmentation has been proposed. However, this rough set theory is already used in Medical Image Segmentation [14-18]. This proposed algorithm gives better results in terms of validity index which shows the superiority of our algorithm over others. The advantage of K-means is that for large number of variables, it may be computationally faster than hierarchical small and K-means produces tighter clusters than hierarchical clustering. However issues like incompleteness can be efficiently handled by rough set.

3. ROUGH SET

In this section we will introduce some basics of Rough Set theory which are relevant to this paper. For details one may refer to [3] and [4].

According to the definition given by Pawlak [3], an information system is pair $S=<U,A,V,f>$ or a function $f: U \times A \rightarrow V$ where U is a non-empty finite set of N objects $\{x_1, x_2, \ldots\ldots x_n\}$ called the universe and A is a non empty set of finite attributes and V is the value set such that $a: U \rightarrow V_a$ for every $a \in A$. the set V_a is the set values of attribute a, called the domain of a. A subset of attributes defines an equivalence relation named an indiscernibility relation on U. this relationship is defined as

$$IND(B) = \{(x,y) \in U \times U : for\ every\ a \in B, a(x) = a(y)\} \quad (1)$$

Given any subset of attributes B, any concept $X \subseteq U$ can be defined approximately by employing two exact sets called lower and upper approximations. The lower and upper approximations can be defined as follows:

$$\bar{B}X = \cup \{Y \in U | IND(B): Y \cap X = \phi\} \quad (2)$$

$$\underline{B}X = \cup \{Y \in U | IND(B): Y \subseteq X\} \quad (3)$$

The set $\bar{B}X$ also known as upper approximation of X whereas $\underline{B}X$ is known as lower approximation of X. Upper approximation is the set of elements of U which can possibly be classified as the elements of X , employing knowledge B . Lower Approximation is the set of all elements of U which can be classified as elements of X with certainty, in the knowledge B. Obviously zthe difference set yields the set of elements which lie around the boundary.

The set $BN_R(X)$is called B-borderline of X where

$$BN_R(X) = \bar{B}X - \underline{B}X \quad (4)$$

4. ROUG SET BASED K-MEANS ALGORITHM FOR COLOUR IMAGE SEGMENTATION

4.1 K-means algorithm for Color Image Segmentation

The algorithm composed of several steps as described below: For each plane (Red, Green and Blue) perform the following steps and store the results in three different matrices.

- **Step 1:** Read the image and select k random pixel values of the image, these are the initial assumed centroids.
- **Step 3:** For each and every pixel in the image, place it into group i, such that its distance is least from the i^{th} centroid.

$$\text{Distance}= \left\| x_i^j - c_j \right\|^2 \quad (5)$$

- **Step 4:** Re-calculation of centroid values. For each group so formed, calculate the new centroid as the average of the pixel values of all the pixels of that group.
- **Step 5:** Check if all the new centroids are within a predefined tolerance deviation (say +/-1) of the corresponding previous centroid value. If not repeat from step 3.
- **Step 6:** End.

Now merge the three resultant matrices to get the original color segmented image.

4.2 Reducing the redundant Cluster Centers using Rough Set Theory

In the following section, the method proposed in this paper is described in detail.

4.2.1 Obtain each feature's membership value

Initially, randomly choose the cluster centers $\{C_1, C_2, \ldots C_c\}$ from an image data point set. Where $c \in [c_{\min}, c_{\max}], c_{\min} = 2, c_{\max} = n = \sqrt{M \times N}$ (M×N is the total number of pixels). Each cluster centers C_i is represented by n numeric image features $\{\xi_i,\ for\ i = 1,2,\ldots. n\}$. Then each feature is described by in terms of its fuzzy membership values corresponding to three fuzzy sets, namely, Low(L), Medium(M), and High(H), which characterized by respectively by a π-membership function:

$$\mu(\xi_i) = \begin{cases} 2\left(1-\dfrac{|\xi_i-\beta|}{\gamma}\right)^2 & for\ \dfrac{\gamma}{2} \le |\xi_i-\beta| \le \gamma \\ 1-2\left(\dfrac{|\xi_i-\beta|}{\gamma}\right)^2 & for\ 0 \le |\xi_i-\beta| \le \dfrac{\gamma}{2} \\ 0 & Otherwise \end{cases} \quad (5)$$

Here, γ is the radius of π-membership function with β as the central point. These values of center β and radius γ are chosen so those for these three fuzzy sets have overlapping nature as described in [20].

4.2.2 Constitute a decision table for the initial cluster centers set

In a same cluster centers set, if a cluster center has a same similarity value to an-other one, then they are called redundant cluster center each other [9].

If C_i and C_j are redundant cluster centers each other, C_j and C_k are redundant cluster centers each other, then C_i, C_j and C_k belong to a same redundant cluster center, i.e.

$$C_i \leftrightarrow C_j, C_j \leftrightarrow C_k \Rightarrow C_i \leftrightarrow C_j \leftrightarrow C_k \quad (6)$$

Now, consider the initial cluster centers as objects and cluster centers feature ξ_i, the central point β and the radius γ as conditional attributes and computing π -membership function value as decision attribute

4.2.3 Eliminating redundant cluster centers from the initial cluster centers set

Reducing the redundant cluster center from the initial cluster centers set depend on the following four decision rules:

1. $if\ \xi_i = \xi_j\ and\ \mu(\xi_i) = \mu(\xi_j)\ then\ f(C_i) = f(C_j)$

Two centers C_i and C_j are redundant.

2. $if\ \xi_i = \xi_j\ and\ \mu(\xi_i) \ne \mu(\xi_j)\ then\ f(C_i) = f(C_j)$

Two centers C_i and C_j are redundant.

3. $if\ \xi_i \ne \xi_j\ and\ \mu(\xi_i) = \mu(\xi_j)\ then\ f(C_i) = f(C_j)$

Two centers C_i and C_j are redundant.

4. $if\ \xi_i \ne \xi_j\ and\ \mu(\xi_i) \ne \mu(\xi_j)\ then\ f(C_i) \ne f(C_j)$

Two centers C_i and C_j are not redundant.

Where *f( )* denotes the decision attribute. Eliminate the redundant decision items from an initial cluster center. For each condition attribute carry out the process mentioned above until condition attribute set does not change

4.2.4 Validity Index

As soon as redundant initial cluster centers in the initial cluster set is eliminated, a reduced and optimized cluster center set is used as the K-means initial input for image segmentation. To evaluate cluster quality Validity index is used. The Validity index is expressed as follows:-

The cluster validity measure used in the paper is the one proposed by Turi [8]. It aims at minimizing the validity index given by the function,

$$V = y \times \frac{intra}{inter} \quad (7)$$

The term intra is the average of all the distances between each pixel x(r,g,b) and its cluster centroid $z_i(r,g,b)$ which is defined as

$$intra = \frac{1}{N}\sum_{i=1}^{K}\sum_{x\in C_i}\| x(r,g,b) - z_i(r,g,b) \|^2 \quad (8)$$

Where $\|x(r,g,b) - z_i(r,g,b)\|$ means the Euclidean distance, which is calculated as

$$\sqrt{\left((x_{red}-Z_{ired})^2 + \left(x_{green}-Z_{igreen}\right)^2 + (x_{blue}-Z_{iblue})^2\right)}$$

where N is the total number of pixels, C_i is the cluster number, z_i is the centroids of cluster C_i, K is the total number of clusters. Intra cluster dependency is the sum of square of Euclidean distance of every element from the centroids of the cluster to which it belongs.

On the other hand, inter is the inter cluster dependency which gives the idea about the extent to which each clusters are related. The higher this value the better clustering is obtained. It is represented as

$$inter = \min(\| z_i(r,g,b) - z_j(r,g,b) \|^2),\ \ where$$
$$i = 1,2,...K-1,\ j = i+1,...,K \quad (9)$$

and Z_i and Z_j are the centroids. Intra cluster dependency is the minimum of the square of Euclidean distances of each centroids from the other.

Lastly, y is given as

$$y = c \times N(2,1) + 1 \quad (10)$$

Where c is a constant (selected value is 25), N(2,1) is a Gaussian distribution function with mean 2 and standard deviation 1, where the variable is the cluster number and is given as

$$N(\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(k-\mu)^2}{2\sigma^2}} \quad (11)$$

Where K is the cluster number and μ=2 and σ=1 as per Turi's thesis on clustering. This is done to negate the occurrence of lower number of clusters 2 or 3.

This validity measure serves the dual purpose of

- minimizing the intra-cluster spread, and
- maximizing the inter-cluster distance.

Moreover it overcomes the tendency to select a smaller number of clusters (2 or 3) as optimal, which is an inherent limitation of other validity measures such as the Davies-Bouldin index or Dunne's index.

Now the procedure for Rough Set based K-means image segmentation method is shown in Fig 1.

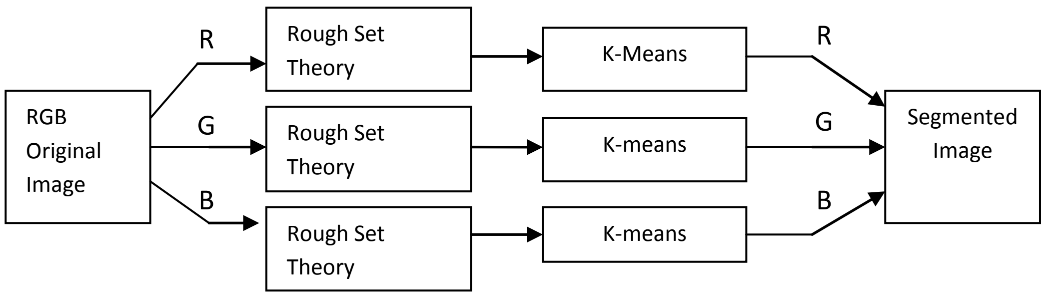
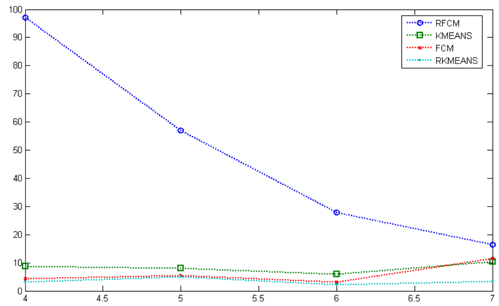


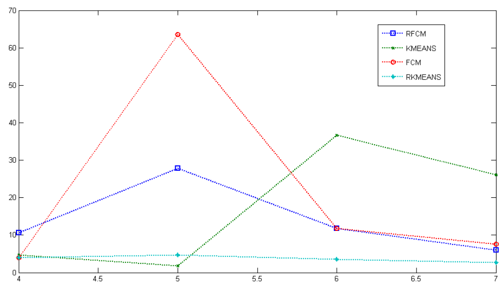
Fig 1: Block Diagram of the Proposed Algorithm

5. EXPERIMENTAL RESULTS

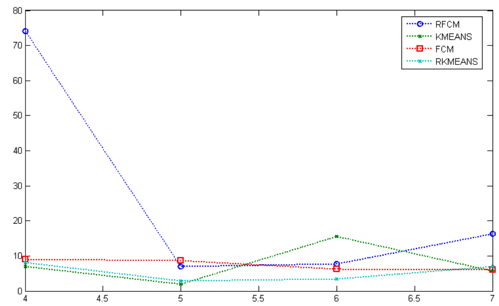
Proposed algorithm applied on well known natural images shown below. This similarity based image segmentation method partitions into different regions exactly. Visually as well as theoretically our method gives better results other than state of the art methods like K-Means, FCM, Rough set based FCM. VI index for each image is shown in Fig 2. The results tabulated here for each image is the mean of 10 simulations. VI index measured for the compactness of each image regions. Compared to the results of the VI index with state of the art image segmentation method gives the better results which are shown below.



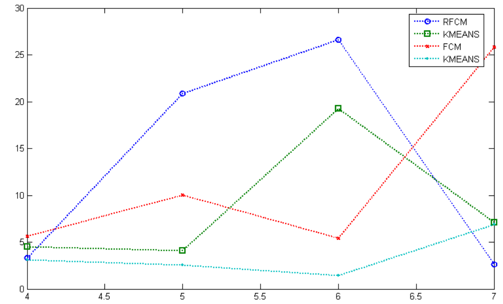
(b)



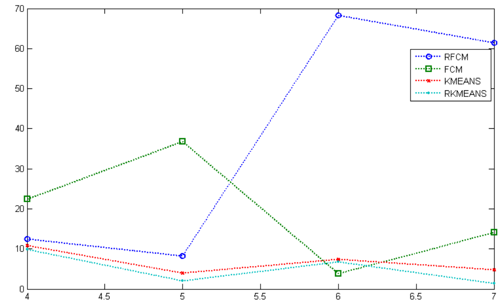
(a)



(c)



(d)



(e)

Fig 2: (a), (b),(c),(d),(e) represent the No. of clusters Vs. Validity Index graph for Lena, Mandrill, Sea Beach, Parrot and Peppers Images.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Fig3: (a) & (e), (b) & (f), (c) & (g), (d) & (h) represent the segmented image of Sea beach and Peppers image using Rough set based FCM, K-means, FCM and Proposed method.

6. CONCLUSIONS

This paper presents a new unsupervised color image segmentation algorithm that can automatically separate the different regions in each image. The proposed method does not require any priori information. This rough set based K-means color image segmentation technique confirmed the superiority and more stable (entropy) than other state-of-the-art-image segmentation techniques. However there are still many things that need to be improved, including a faster speed for real time applications. Rough set based this technique can be applied further to change detection for remote sensing images, medical imaging system and object detection.

7. REFERENCES

[1] Hartigan, J.A., Wong, M.A.: Algorithm AS136: A K-Means Clustering Algorithm. Applied Statistics 28, pp. 100–108 (1979).

[2] JA Hartigan, Clustering Al gorithms, John Wiley & Sons, New York, 1975.

[3] Pawlak Z.: Rough set, International Journal of Computer and Information Science 11 (1982) pp.341-356.

[4] Skowron A., Stepaniuk, J.: Information granules in distributed environment. In: Zhong, N., Skowron, A., Ohsuga, S. (eds.) RSFDGr C 1999.LNCS (LNAI) Vol. 1711, pp. 357–365. Springer, Heidelberg (1999).

[5] Suzuki, H. and J. Toriwaki, 1991. Automatic Segmentation of Head MRI Images by Knowledge Guided Thresholding. Computer Med. Imaging Graph: The official J. Computerized Imaging Society, 15 (4): 233-240.

[6] Mahamed G. H. Omran, Andries P Engelbrecht and Ayed Salman, Dynamic Clustering using Particle Swarm Optimization with Application in Unsupervised Image Classification, PWASET Volume 9, 2005.

[7] Amiya Halder, Nilavra Pathak, "An Evolutionary Dynamic Clustering Based Colour Image Segmentation", International Journal of Image Processing (IJIP), Volume (4): Issue (6), pp.549-556, January 2011.

[8] Amiya Halder, Soumajit Pramanik, Arindam Kar, "Dynamic Image Segmentation using Fuzzy C-Means based Genetic Algorithm", International Journal of Computer Applications Volume 28– No.6, pp. 15-20, August 2011.

[9] Sankar K. Pal, Pabitra Mitra: "Case Generation Using Rough Sets with Fuzzy Representation", IEEE Transactions on Knowledge and Data Engineering, pp.-292-300,2004.

[10] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, Pearson Education, 2002.

[11] Zhenghao Shi,Yuyan Chao,Lif eng He,Tsuyoshi Nakamura and Hidemori Itoh, "Rough set based FCM algorithm for Image Segmentation". International Journal of Computational Science, Global Information Publisher 2007, Vol. 1, No1, 58-68.

[12] R. H. Turi, "Clustering-Based Color Image Segmentation", PhD Thesis, Monash University, Australia, 2001.

[13] Amiya Halder and Avijit Dasgupta, "Image Segmentation using Rough Set Based K-Means Algorithm", CUBE 2012, September 3–5, 2012, Pune, Maharashtra, India.

[14] Wei Zhang, Yu-zhu Zhang, Cheng Li, "A new hybrid algorithm for image segmentation based on rough sets and enhanced fuzzy c-means clustering", IEEE, International Conference on Automation and Logistics, August 2009.

[15] Milinda M. Mushrif, Ajoy K. Ray, "Color image segmentation: Rough-set theoretic approach", Pattern Recognition Letters, Elsevier, 2008.

[16] Ran Li, "Medical Image Segmentation Based on Watershed Transformation and Rough Sets", 4th International Conference on Bioinformatics and Biomedical Engineering, pp. 1-5, June 2010.

[17] S Mohapatra, D Patra and Kundan Kumar, "Blood microscopic image segmentation using Rough Sets", IEEE, ICIIP 2011.

[18] S Mohapatra, D Patra, "Unsupervised Leukocyte Image Segmentation Using Rough Fuzzy Clustering", ISRN Artificial Intelligence, 2012.