A Novel Color Image Segmentation Approach Based On K-Means Clustering with Proper Determination of the Number of Clusters and Suitable Distance Metric

Anil Kumar Gupta, Dibya Jyoti Bora
Department of Computer Science & Applications
Barkatullah University, Bhopal (India)
Corresponding author: research4dibya@gmail.com

Abstract— ***K-Means algorithm is the most commonly chosen technique for color image segmentation task. Although this algorithm is famous for its low complexity and easy implementation, but usually it is seen that the segmentation results are suffering from noises and over-segmentation, which mislead the final image analysis process. This is because of the inappropriate selection of the number of clusters in K-Means. Also, choosing a proper distance metric for this algorithm is very important as it impacts on the final segmentation results. This paper deals with these problems and presents a novel approach to solving the same. As per color image segmentation concerned, so choosing a suitable color space is the first mandatory condition. HSV color space is selected for the proposed research work. The input RGB image is first converted to HSV one. Here, the number of clusters is determined from the input image beforehand. For this, it is considered that number of regions in an image is equivalent to the number of clusters that can be formed through clustering the pixels of the image. The number of regions of an image is determined with Meyer's Watershed algorithm with a proper preprocessing technique. Generally, mere applying Watershed algorithm results in over segmentation. So, we have introduced an improved Sobel filter based on multiple directional edge detection to deal with this problem. The V channel of the HSV converted image is filtered by the proposed improved Sobel filter first and then the filtered image is sent as input to the watershed algorithm. The watershed algorithm analyses the regions of the image through local minima calculations and the total number of regions hereby found is assigned to K. Then with the predetermined K, the pixels of the HSV converted image are clustered with K-Means Algorithm. "Cosine Distance Metric" is chosen for the distance based calculations involved in the K-Means algorithm. By properly labeling the different clusters, the final segmented image is obtained. The experimental results proved the better performance of the proposed approach in comparison to K-Means algorithm. Also, when applied to satellite color images, it is found that the proposed approach succeeds to form clear and distinct segments of the same and hence establishes a good framework for satellite color image segmentation.***

Keywords-Color Image Segmentation, Cosine Distance Metric, HSV Color Space, K-Means Algorithm, Satellite Image, Satellite Color Image Segmentation, Sobel Filter, Watershed Algorithm.

I. INTRODUCTION:

Image Segmentation is the most important and critical part of any image analysis process. This is a process by which specific boundaries of interesting objects are highlighted [1]. The main objective of image segmentation is to simplify the representation of pictures into regions of meaningful information. This is very necessary for image analysis process because in an image there may be several targets. Through image segmentation, these targets are located and isolated. Once these targets are isolated, then they can be measured and classified. Say, I be the given image. Then, by image segmentation, it will be partitioned into 'n' disjoint partitions R_i (i=1,2,..,n) so that the following properties will be satisfied[2][3]:

(i) $\bigcup_{i=1}^{n} R_i = R$

(ii) $R_i \bigcap R_j = \Phi$

(iii) $H(R_i) = \text{TRUE } \forall i$

$$\text{(iv)}\, H(R_i \cup R_j) = \text{FALSE}\ \forall R_i \ \&\ R_j \text{adjacent.}$$

Here, H(R) denotes the homogeneity attributes of pixels over region R on the basis of which the whole segmentation process is carried out [3]. These homogeneity criteria may be color, intensity, or texture [4]. From (iii), it is obvious that pixels within a region must share same featured components. And the property (iv) implies that if pixels belong to two different regions then their featured components must also be different from each other. Image segmentation may be gray or color. Color image segmentation is sometimes found more beneficial than grayscale segmentation because of the reason that human eyes are more adjustable to brightness, so, can identify thousands of color at any point of a complex image, while only a dozens of gray scale are possible to be identified at the same time [5]. Different techniques are available for color image segmentation among which "clustering based ones" are always given preference because of the resemblance between clustering and segmentation-both are the process of a grouping of similar elements on the basis of some similarity criteria. Clustering is an unsupervised study means here we are unaware of the input data means there is no previous labeling on the data to be grouped [6]. Same way, in image segmentation we are trying to make a grouping of pixels without previous labeling on them. Clustering can be either hard clustering or soft clustering [7]. In the case of hard clustering, data items are clustered hardly means if a datum belongs to a definite cluster, it could not be included in another cluster[6][7]. While in the case of soft clustering, data autumn are grouped on the basis of some membership degree, depending on the value of which a data item may belong to more than one cluster[6][7]. There exist different clustering techniques, both hard and soft, for image segmentation [8]. K-Means algorithm is a famous hard clustering algorithm frequently adopted for the same purpose [3][9].But K-Means algorithm is often observed to suffer from noisy image segmentation and over segmentation[3]. The main reason for this is the improper value of K(the number of clusters). In K-Means algorithm, we need to provide a proper value for K beforehand. If K value is assigned randomly without proper calculations, then it is found producing ineffective and poor results and also sometimes irrelevant results [10]. So, with this proposed work an effort has been made to determine the value of K beforehand. Also, the distance metric employed plays an important role in the performance of K-Means algorithm [11]. As in our case, data items considered are pixels, so "Cosine Distance Metric" is more relevant for a pixel to centroid distance calculations. Since color image segmentation is dependent on choosing a proper color space and HSV color space is performing better in comparison to other existing color spaces, so, we have chosen this for color attributes distribution.

The later portion of the paper is organized is as follows: In section II, a review of the previous work done is given. In Section III, the steps involved in the proposed algorithm are mentioned. Then, in the corresponding Section IV, V, VI, VII and VIII, the topics involved in the proposed algorithm are discussed thoroughly. Then experiments and results are presented in Section IX. This section also shows the application of the proposed approach for satellite color image segmentation. Finally, we come to the conclusion in Section X and after that, we have the reference section.

II. LITERATURE REVIEW:

In [12], the authors proposed an efficient and fast approach for color image segmentation where a new quantization technique for HSV color space is implemented to generate a color histogram and a gray histogram for K-Means clustering, which operates across different dimensions in HSV color space. Here, the initialization of centroids and the number of the clusters are automatically estimated. Also, a filter for post-processing is introduced to effectively eliminate small spatial regions. Experiments show that the proposed segmentation algorithm achieves high computational speed, and salient regions of images can be effectively extracted. Moreover, the segmentation results are close to human perceptions. In [13], the authors first applied partial stretching enhancement to improve the quality of the image. The subtractive cluster is used to generate the initial centers and these centers are used in K-Means algorithm for the segmentation of the image. Finally, a median filter is applied to the segmented image to remove any unwanted region from the image. In [14], an approach to color image segmentation using K-Means Classification on RGB histogram is proposed. The proposed approach has solved the problem of missing locality information and presented the image in distinct colors that clearly identifies the objects of the image. First, the image is read and adjusted to a standard size. Then, pixels are divided into different clusters based on their color, texture, and region. The cluster values are calculated using the K-Means clustering algorithm. If no pixel remains, all the clusters are combined and finally the image is presented in the form of a segmented image. In [15], the authors proposed an efficient technique for MRI brain image segmentation. The proposed technique is an HSV colorspace involved approach for classification of brain magnetic resonance images (MRI) based on color converted K- means clustering segmentation algorithm. In [16], a novel approach for clustering based image segmentation is proposed. The input image is first converted from RGB to LAB and then segmented by K-Means algorithm with cosine distance measure. Then, the segmented image is filtered by Sobel filter and further analyzed by Watershed algorithm. Lastly, the output image from Watershed algorithm is filtered by Median filter to obtain the final segmented image of the input image. The results are found very efficient as per the analyzed MSE and PSNR values.

In [17], the authors proposed a special constrained K-Means approach to color image segmentation. In this proposed approach K-Means clustering is employed in feature space first. Then, in the image plane, the spatial constraints are adopted into the hierarchical K-means clusters on each level. The two processes are carried out alternatively and iteratively. With this, an effective region merging method is proposed to handle the over-segmentation. The experimental results proved that the proposed approach is fast and generic and hence practically applicable.

III. STEPS INVOLVED IN THE PROPOSED ALGORITHM:

Following are the steps involved in the proposed algorithm:

Step1. The Input image is converted from RGB to HSV.

Step2. The V channel of the HSV converted image is filtered by an Improved Sobel filter based on multiple directional edge detection.

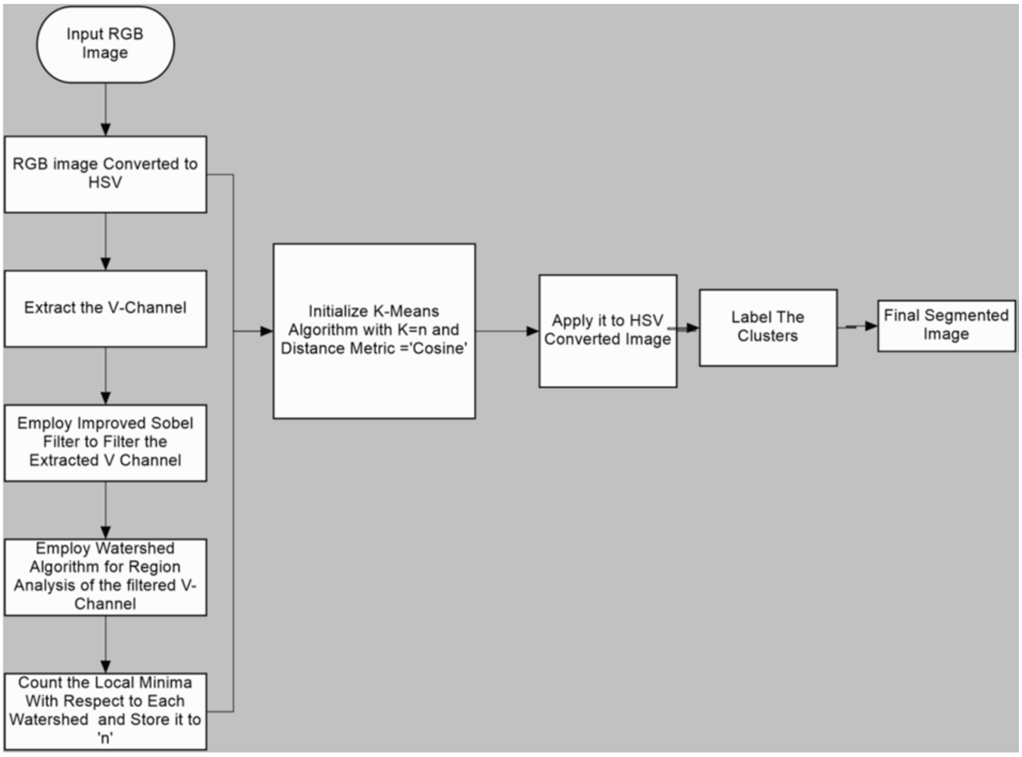
Step3. The filtered image is sent as input to the Watershed algorithm for the purpose of region analysis. The regions are counted based on the number of local minima detected in association to every watershed.

Step4. The number of regions counted at step 3 is stored into n (an integer variable).

Step5. The original HSV converted image obtained at step1 is now sent as input to K-Means algorithm with the following parameters: K value (number of clusters) as n (obtained at step4) and distance measure as "Cosine Distance Metric".

Step6. The clusters of pixels obtained as output from K-Means algorithm at step5 are labeled properly to obtain the final segmented image.

Flowchart 1: The Proposed Approach



IV. HSV COLOR SPACE:

HSV color space is a frequently chosen color space for color image segmentation. This color space can be represented by a hexacone in three dimensions where the central vertical axis represents intensity [18]. Here 'H' stands for 'Hue' which is an angle in the range $[0,2\pi]$ relative to the red axis with red at angle 0, green at $2\pi/3$, blue at $4\pi/3$ and red again at 2π[18][19]. Here 'S' stands for 'Saturation' and the saturation describes how pure the hue is with respect to a white reference. And 'V' stands for 'Value' and value is a percentage that goes from 0 to 100. This percentage value can be thought as the amount of light illuminating a color [20]. A diagrammatic view of the HSV color solid cylinder can be found in [21]:

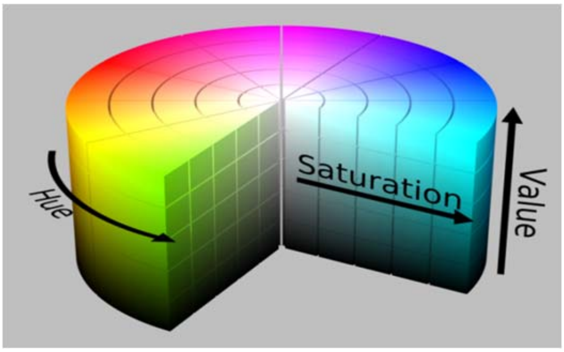


Figure 1: *HSV Color Space.*

V. IMPROVED SOBEL FILTER:

Sobel filter is a discrete differentiation operator which is most popularly adopted for detecting edges [23]. This computes an approximation of the gradient of the image intensity function. With each image, the Sobel operator involves two kernels Gx and G_y , where, G_x is the gradient estimation in x-direction while G_y the gradient estimation in the y-direction.

Then the absolute gradient magnitude will be given by:

$$|\mathbf{G}| = \sqrt{(\mathbf{G_x}^2 + \mathbf{G_y}^2)}$$

Although, this value is often approximated with:

$$|\mathbf{G}| = |\mathbf{G_x}|+|\mathbf{G_y}|$$

The directions θ of the gradient can be obtained as:

$$\theta = a\tan\left(\frac{G_y}{G_x}\right)$$

Sobel operator has the capacity of detecting edges for noisy images also. This is due to the reason that here every image is differentially separated by two rows and columns, thereby enhancing the edge elements on both sides, which in turn results in a very bright and thick look of the edges. But, this operator is confined to only two directions, i.e., horizontal and vertical. So, the performance of this operator can be increased if multiple directions are considered for gradient estimation [24]. So, using this concept an improved Sobel operator is adopted for this paper [24] where all the eight directions are considered for gradient estimation. These eight directions are 45 °, 90°, 135 °, 180 °, 225 °, 270 °, 315 °, and 360 °.

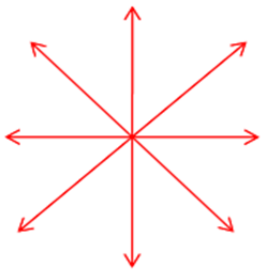


Figure 2: Eight Different Directions That Can Be Adopted For Edge Detection by Sobel Filter

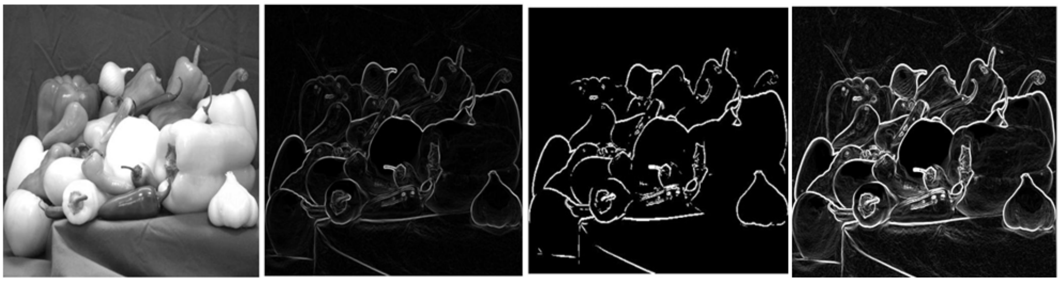
Consider a general template of 3x3 image mask:

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

Here, say p is the concerned pixel. Now, for eight directional edge detection we need to select four parameters [25]:

$$G_H = \begin{pmatrix} a & b & a \\ 0 & 0 & 0 \\ -a & -b & -a \end{pmatrix}, G_v = \begin{pmatrix} -a & 0 & a \\ -b & 0 & b \\ -a & 0 & a \end{pmatrix}, \ G_{dl} = \begin{pmatrix} 0 & a & b \\ -a & 0 & a \\ -b & -a & 0 \end{pmatrix} \text{ and } G_{dr} = \begin{pmatrix} b & a & 0 \\ a & 0 & -a \\ 0 & -a & -b \end{pmatrix}$$

We have implemented the proposed improved Sobel filter in Matlab R2013a. A few experiments for Sobel edge detection are performed on 'pepper's' image. For this, we first convert the RGB pepper's image to HSV and V channel has been extracted from the same. The extracted V channel has been used for the required edge detection. First, edges are detected with two directional Sobel filters; second, by four directional Sobel filter and lastly, by the proposed eight directional improved Sobel filter. Results are shown below:



(a) (b) (c) (d)

Figure 3: (a) V channel of the HSV converted Peppers' Image; (b) Filtered by Two Directional Sobel Filter ;

(c) Filtered By Four Directional Sobel Filter; (d) Filtered By Proposed Eight Directional Sobel Filter.

So, it is clearly seen from the above figures, that edges detected by the proposed eight directional Sobel filter is more distinct and thicker. This is due to the reason that we have considered every possible direction (with respect to eight neighbors of the concerned pixel) for the calculation of edges. This improved version of Sobel filter will be the most appropriate preprocessing technique for Watershed algorithm for region analysis since, in this case, the proposed technique will remove the possibility of over segmentation up to a satisfactory level. The proposed color image segmentation approach in this paper is developed by keeping in mind its application area towards satellite image segmentation. The main issue to focus on satellite image segmentation is to tackle unnecessary noises so that analysis of sensitive information will not be misled by the same.

VI. COSINE DISTANCE METRIC:

Cosine distance metric is a measure of similarity between two vectors of n dimensions by finding the cosine of the angle between them. This distance measure is mostly related to the measurement of operation, but not magnitude [16][26]. In our case, pixels orientation is given prior importance, which is same as in the case of cosine distance measure. That's why we have chosen this distance measure to be used in K-Means algorithm. Mathematically, we can define cosine distance measure as follows: Given an *m*-by-*n* data matrix X, which is treated as *m* (1-by-*n*) row vectors $x_1, x_2, ..., x_m$, the cosine distances between the vector x_s and x_t are defined as follows[27] :

$$d_{st} = 1 - \frac{x_s x_t'}{\sqrt{(x_s x_s')(x_t x_t')}}$$

VII.K –MEANS ALGORITHM:

K-Means algorithm is a famous hard clustering algorithm with comparatively less time complexity, is also known as Lloyd's algorithm in computer science community [28]. The basic steps involved in K-Means algorithm are [29] [30]:

[a] At first, k initial centroids are selected (randomly chosen).

[b] Then, all the objects are assigned to their closest clusters on the basis of the distance calculated through a proper distance measure.

[c] Recompute the centroid of each cluster

[d] Repeat [b] and [c] until the centroids do not change or memberships finalize.

Throughout the algorithm, it is tried to minimize the following objective function known as the squared error function [29] [30]:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \| x_i^{(j)} - c_j \|^2$$

where, $\| x_i^{(j)} - c_j \|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and a cluster center c_j. In our case, this distance measure is the cosine distance measure.

Also, random choosing K, the number of clusters may lead to improper segmentation of the image. So, in our proposed approach, we have introduced a technique for finding out K beforehand by analyzing the image with Watershed algorithm where we are counting the local minima (associated to each watershed) from the concerned image and assigning the total number as the required number of clusters(number of regions).

VIII. MEYER'S WATERSHED ALGORITHM:

This is a very powerful morphological tool for region-based image segmentation. The basic steps involved in this algorithm are [31]:

1. Say, Ip is the current pixel. Add neighbors to Q_p, where Q_p is the priority queue sorted by value.
2. Find local minima and assign them as the region seeds.
3. Consider the top priority pixel from queue
    1. If all labeled neighbors have the same label, assign to the pixel.
    2. Add all non-marked neighbors.
4. Repeat step 3 until finished.

But the main problem associated with this algorithm is the formation of "over segmentation". A few steps of pre-processing of the input image are required to overcome this problem [32]. In our case, we have introduced an improved Sobel filter for this purpose. As a result of the combination of the improved Sobel filter and Meyer's Watershed Algorithm, the over-segmentation problem can be overcome up to a satisfactory level so that the regions present in an image are clearly identified. The local minima (associated to every watershed) counted from the watershed algorithm can be treated as the number of regions in this case as the proposed improved Sobel filter removes the noises up to the needed level and hence redundant local minima are easily avoided in this case. The number of regions found is assigned to K (the number of clusters) in K-Means algorithm.

IX. EXPERIMENTS AND RESULTS:

We have implemented our proposed algorithm in Matlab R2013a. The experiments are conducted on a system with an Intel Core i5 processor running at 2.30 GHz and 4 GB RAM running 64-bit operating system Windows 10. For analysis and comparisons of the results found, we have adopted the following measurement metrics: I> Visual perspective; II> AMBE; III>NAE; IV> MSE and PSNR; and V> Borsotti's Evaluation Function.

(I) Visual Perspective:

By visual perspective, we mean how the human eye can detect the objects in an image [33]. In our case, visual perspective is taken as a measurement tool to determine whether segments are clearly visible to bare eye or not.

(II) Absolute Mean Brightness Error (AMBE):

AMBE is an objective measurement to rate the performance in preserving the original brightness. This can be defined as the absolute difference between the mean of the input and the output images and is proposed to rate the performance in preserving the original brightness [34, 35, 36]. AMBE is calculated using the formula:

$$AMBE = |E(\boldsymbol{X}) - E(\boldsymbol{Y})|$$

Here, X and Y denotes the input and output image, respectively, and E (.) denotes the expected value, i.e., the statistical mean. The above equation clearly shows that AMBE is designed to detect one of the distortions– excessive brightness changes [34] [35]. A Lower AMBE indicates the better brightness preservation of the image.

(III) Normalized Absolute Error (NAE):

NAE is defined as follows [37]:

$$NAE = \sum_{j=1}^{M}\sum_{k=1}^{N} | x_{j,k} - x'_{j,k} | \Bigg/ \sum_{j=1}^{M}\sum_{k=1}^{N} | x_{j,k} |$$

A large value of NAE implies the image is of poor quality.

(IV) Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR):

The MSE (Mean Squared Error) is the cumulative squared error between the compressed and the original image, whereas PSNR(Peak Signal to Noise Ratio) is the peak error[38]. MSE can be computed using the following formula [37] [38] is:

$$\text{MSE} = \sum_{y=1}^{M} \sum_{x=1}^{N} [I(x,y) - I'(x,y)]^2$$

where, I(x,y) is the original image, I'(x,y) is its noisy approximated version (which is actually the decompressed image) and M,N are the dimensions of the images value for MSE implies lesser error.

The formula for PSNR [38] is:

$\text{PSNR} = 10\log_{10}(MAXi^2/MSE)$

Where, *MAXi* is the maximum possible pixel value of the image. A higher value of PSNR is always preferred as it implies the ratio of Signal to Noise will be higher. 'Signal' here is the original image, and the 'noise' is the error in reconstruction.

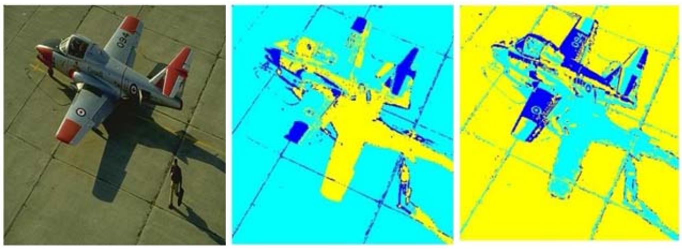
(V) Borsotti's Evaluation Function (Q): This is an empirical goodness evaluation metric which addresses the uniformity feature within segmented regions (color deviation). A low value of Q indicates better segmentation result. The formula for calculating Q is [39]:

$$Q = \frac{1}{1000(N*M)}\sqrt{R}\sum_{i=1}^{R}\left[\frac{e_i^2}{1+\log A_i}+\left(\frac{R(A_i)}{A_i}\right)^2\right]$$

Where, N*M is the size of the image I, e_i is the color error of the region i and R(A) is the number of regions of the size A.

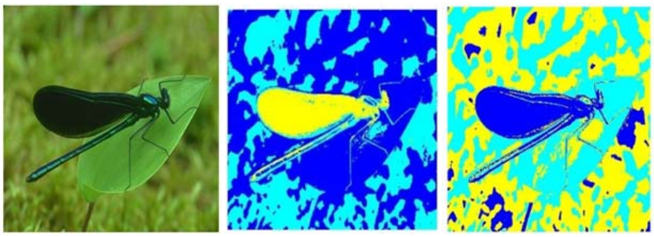
The experiments are performed on the images collected from Berkeley Image Segmentation Dataset [40]. Although the proposed algorithm has been applied to around 20 different images, we are presenting results for following six images:

Image1:



(a) (b) (c)

Image2:



(a) (b) (c)

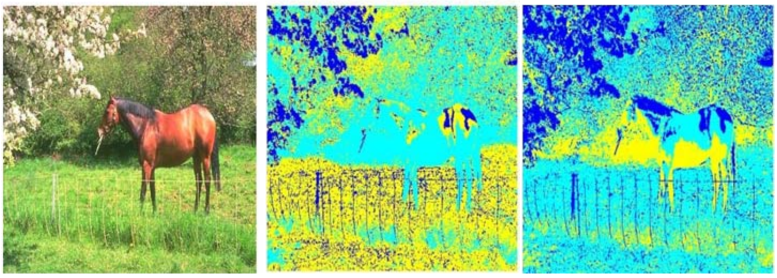
Image3:

(a) (b) (c)

Image4:

(a) (b) (c)

Image5:



(a) (b) (c)

Image6:

(a) (b) (c)

Figure 4: *(a) The original image , (b) The image obtained by K- Means Segmentation with Random Initialization of Cluster Numbers and Euclidean Distance Metric; and (c) The image obtained after segmentation by the proposed approach.*

So, from the above figures, if we consider the visual perspective point of view, then clearly the color image segmentation obtained through the proposed approach is much better than the same obtained through the K-Means algorithm, because in the former case, the regions of an image are properly segmented with better isolation of the region of interest (ROI) through significant boundaries.

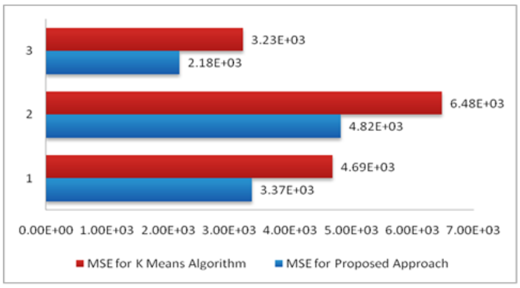
TABLE 1: MSE, PSNR, NAE and AMBE values for the Proposed Approach

	MSE	**PSNR**	**NAE**	**AMBE**
Image 1	MSE(:,:,1) =3.3686e+03 MSE(:,:,2) =4.8221e+03 MSE(:,:,3) =2.1798e+03	PSNR(:,:,1) =12.8563 PSNR(:,:,2) =11.2985 PSNR(:,:,3) =14.7466	0.4612	47.4039
Image 2	MSE(:,:,1) =2.6317e+03 MSE(:,:,2) =2.1373e+03 MSE(:,:,3) =4.0421e+03	PSNR(:,:,1) =13.9284 PSNR(:,:,2) =14.8321 PSNR(:,:,3) =12.0647	0.2568	40.6702
Image 3	MSE(:,:,1) =4.8288e+03 MSE(:,:,2) =2.8520e+03 MSE(:,:,3) =1.4205e+03	PSNR(:,:,1) =11.2924 PSNR(:,:,2) =13.5793 PSNR(:,:,3) =16.6064	.2822	2.2975
Image 4	MSE(:,:,1) =3.9141e+03 MSE(:,:,2) =3.2763e+03 MSE(:,:,3) =2.2502e+03	PSNR(:,:,1) =12.2045 PSNR(:,:,2) =12.9770 PSNR(:,:,3) =14.6086	.3241	22.7107
Image 5	MSE(:,:,1) =3.8313e+03 MSE(:,:,2) =2.9780e+03 MSE(:,:,3) =3.0463e+03	PSNR(:,:,1) =12.2974 PSNR(:,:,2) =13.3916 PSNR(:,:,3) =13.2931	.2537	12.6166
Image 6	MSE(:,:,1) =2.8252e+03 MSE(:,:,2) =1.7323e+03 MSE(:,:,3) =3.1737e+03	PSNR(:,:,1) =13.6204 PSNR(:,:,2) =15.7446 PSNR(:,:,3) =13.1151	.2433	32.6250

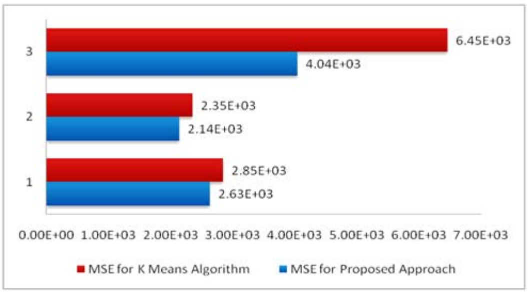
TABLE 2: MSE, PSNR, NAE and AMBE values for the K-Means Algorithm

	MSE	PSNR	NAE	AMBE
Image 1	MSE(:,:,1) =4.6872e+03 MSE(:,:,2) =6.4823e+03 MSE(:,:,3)=3.2270e+03	PSNR(:,:,1) =11.4217 PSNR(:,:,2) =10.0135 PSNR(:,:,3) =13.0428	.6323	53.3214
Image 2	MSE(:,:,1) =2.8473e+03 MSE(:,:,2) =2.3465e+03 MSE(:,:,3) =6.4523e+03	PSNR(:,:,1) =13.5865 PSNR(:,:,2) = 14.4266 PSNR(:,:,3) = 10.0337	.3125	40.8321
Image 3	MSE(:,:,1) =5.6723e+03 MSE(:,:,2) =3.8531e+03 MSE(:,:,3) =2.5123e+03	PSNR(:,:,1) =10.5932 PSNR(:,:,2) =12.2727 PSNR(:,:,3) =14.1301	.3523	3.2562
Image 4	MSE(:,:,1) =4.1521e+03 MSE(:,:,2) =4.0518e+03 MSE(:,:,3) =2.9125e+03	PSNR(:,:,1) =11.9481 PSNR(:,:,2) =12.0543 PSNR(:,:,3) =13.4881	.4501	23.5231
Image 5	MSE(:,:,1) =4.2135e+03 MSE(:,:,2) =3.9121e+03 MSE(:,:,3) =4.0156e+03	PSNR(:,:,1) =11.8844 PSNR(:,:,2) =12.2067 PSNR(:,:,3) =12.0933	.3153	15.3214
Image 6	MSE(:,:,1) =3.8351e+03 MSE(:,:,2) =2.0125e+03 MSE(:,:,3) =4.1723e+03	PSNR(:,:,1) =12.2930 PSNR(:,:,2) =15.0934 PSNR(:,:,3) =11.9270	.5623	35.3211

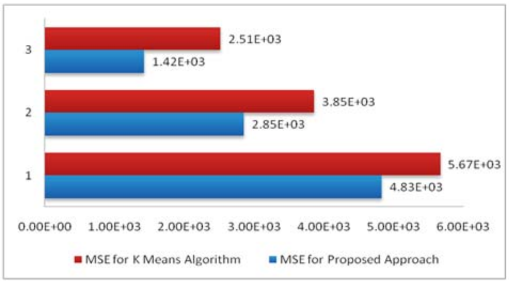
CHART 1: Comparisons of MSE values between Proposed Approach and K-Means Algorithm



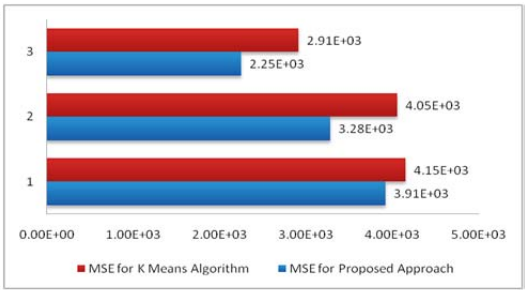
(a)For Image1



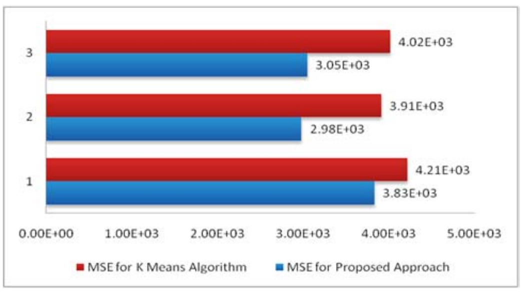
(b) For Image2



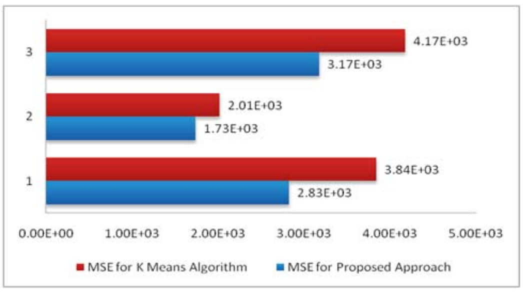
(c) For Image3



(d) For Image4

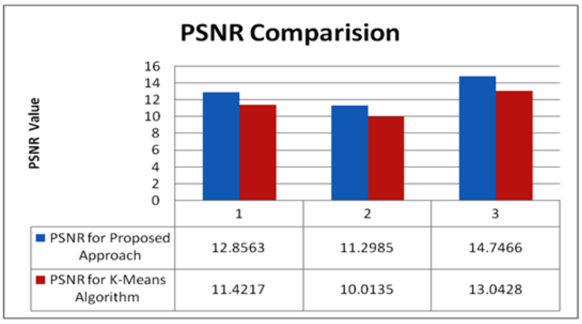


(e) For Image5

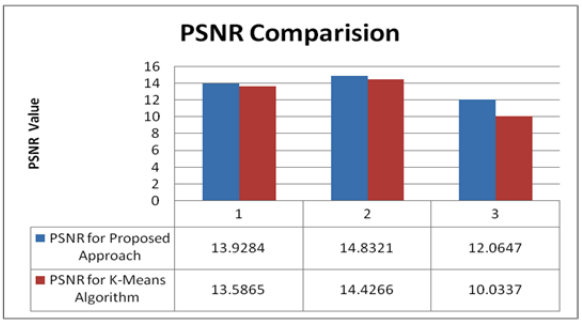


(f) For Image6

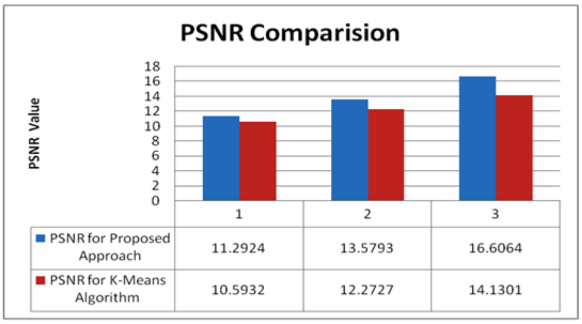
CHART 2: Comparisons of PSNR values between Proposed Approach and K-Means Algorithm



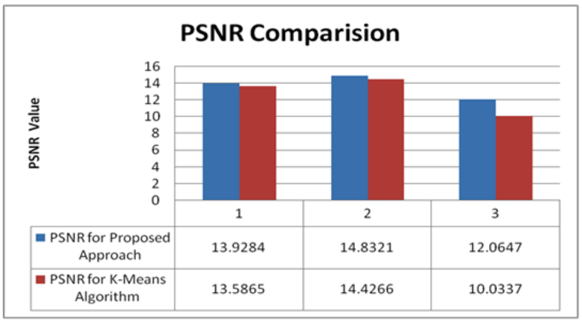
(a) For Image1



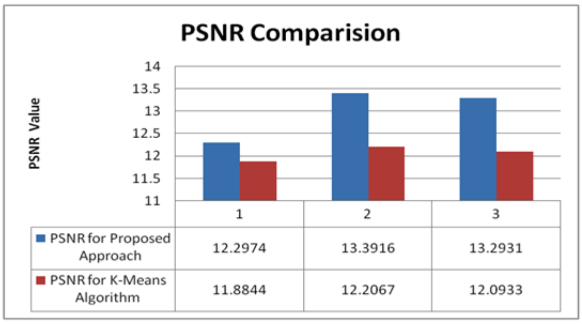
(b) For Image2



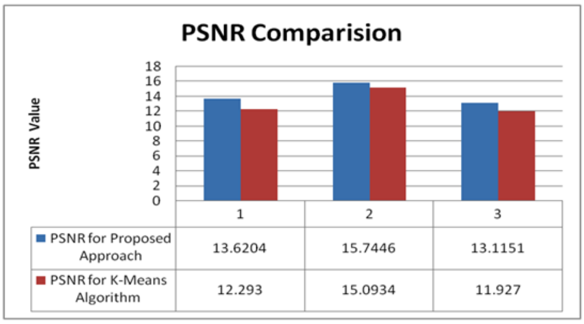
(c) For Image3



(d) For Image4



(e) For Image5



(f) For Image6

CHART 3: Comparison of AMBE values Between Proposed Approach And K-Means Algorithm

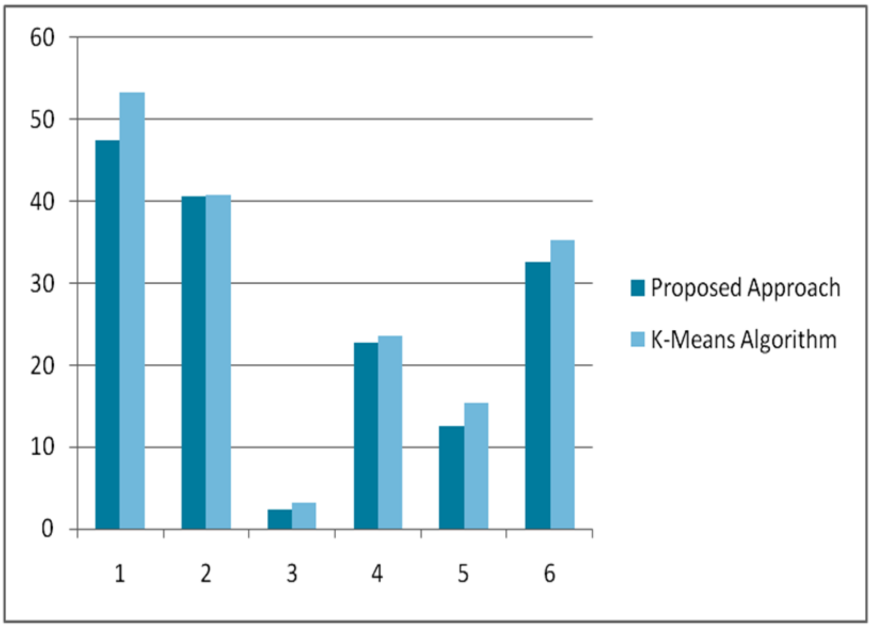


CHART 4: Comparison of NAE values Between Proposed Approach And K-Means Algorithm

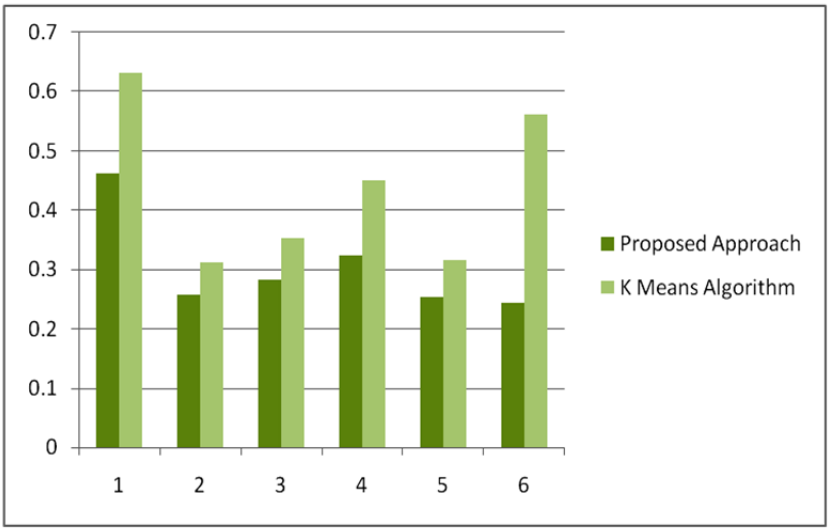
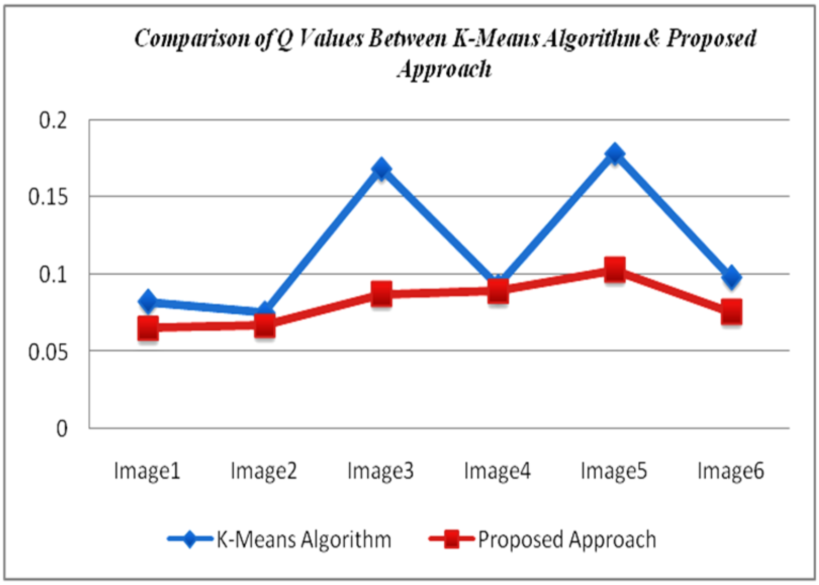


TABLE 3: Comparison of Values of Borsotti's Evaluation Function (Q-Values) between K-Means Algorithm & Proposed Approach

image no.	K-Means Algorithm	Proposed Approach
Image1	0.0821	0.0653
Image2	0.0751	0.0671
Image3	0.1682	0.0871
Image4	0.0928	0.0893
Image5	0.1785	0.1028
Image6	0.0978	0.0756



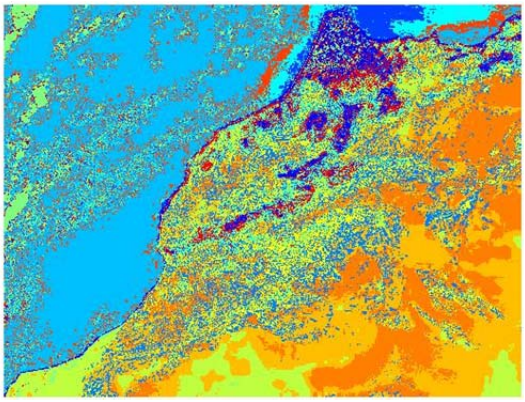
So, the experimental results show that if we compare the MSE values and PSNR values between the proposed approach and K-Means algorithm, then the MSE values for the Proposed approach are smaller than the same of K-Means algorithm while the PSNR values of the Proposed approach are higher than the same of the K-Means algorithm. Again,the NAE and AMBE values calculated for the Proposed approach are comparatively smaller than the same calculated for the K-Means algorithm. Most importantly, the Q-values calculated for the proposed approach are relatively smaller than those of K-Means algorithm. Note that smaller the Q-values, better is the result of the color image segmentation.All of these signify a better segmentation performance of the Proposed approach in comparison to the existent K-Means algorithm.

APPLICATION OF THE PROPOSED APPROACH FOR SATELLITE COLOR IMAGE SEGMENTATION:

Satellite images carry very important and sensitive information. But human eyes tend to fail in analyzing this vast amount of information. Also, small changes in characteristics of information such as intensity, color, texture etc are really difficult to get realized. Here arises the need of an efficient color image segmentation approach which will be able to analyze and segment the satellite color images accurately so that the hidden information can be brought out. From the above experimental results on the images collected from Berkeley Image Segmentation Dataset, it is proved that the proposed approach is an efficient one for color image segmentation. Now, we need to check whether the same efficiency is retained for satellite color image segmentation or not. Therefore, the proposed approach has been applied to different satellite images (images are collected from Earth Science World Image Bank [41]) and the results are analyzed. It is found that the boundaries of the ROIs in every image have been localized properly and also the segmented results are free from the noises and over segmentation. So, our proposed color image segmentation approach is appropriate for satellite color image segmentation. Following are the results of such 4 color satellite images which have been segmented with our proposed approach:

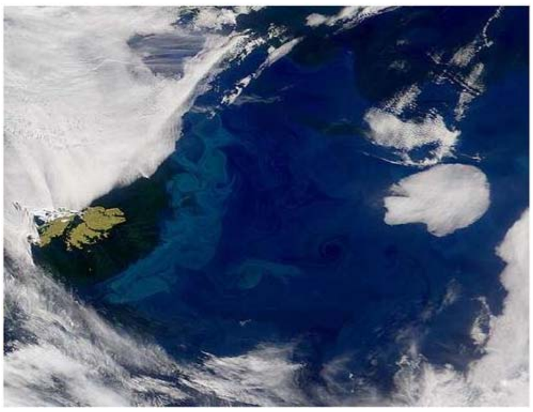
Image 1: This is the image of Morocco acquired in April 2000. The heavily-vegetated peninsula in the top center image is the southern shore of the Strait of Gibralter, which is the passage between the Atlantic Ocean and the Mediterranean Sea. On the tip of this peninsula is the coastal city of Tangier and towards the bottom right of the image is the edge of the Sahara Desert [41].

(a)Original Image

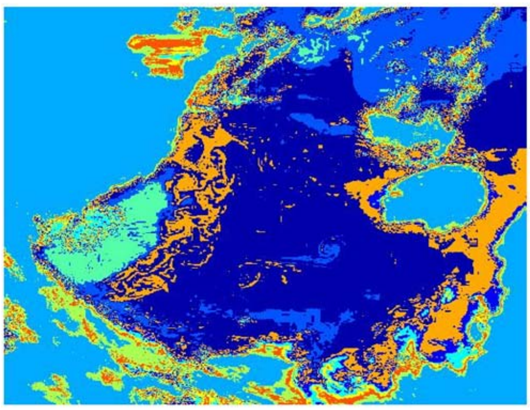


(b) Segmented Image

Image 2: This is the image of Falkland Islands, acquired in November 2001. Numerous eddies can be seen here in the South Atlantic Ocean east of the Falkland Islands. Eddies are highlighted by phytoplankton [41].

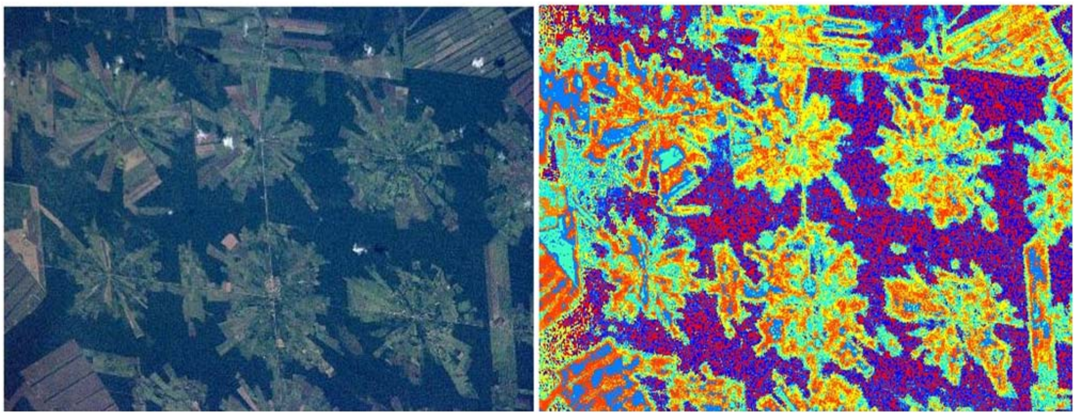


(a)Original Image



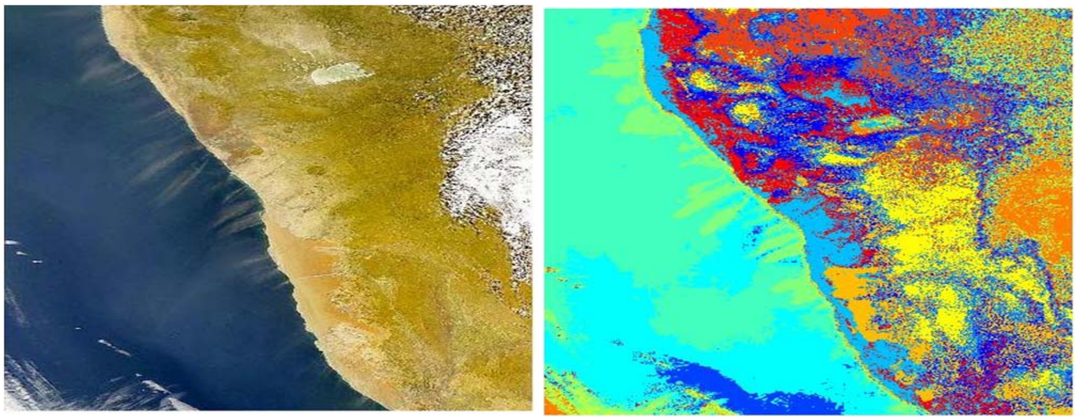
(b) Segmented Image

Image 3: This is the image of eastern Bolivia, showing the deforestation associated with the Tierras Bajas project. Each agricultural 'pinwheel' pattern is centered on a small community, spaced evenly across the landscape at 5-kilometer intervals and people have been resettled here from the Altiplano to cultivate soybeans [41].



(a) Original Image (b) Segmented Image

Image 4: *This is the image of the true color scene of dust streaming from the deserts along the southwestern coast of Africa, acquired in June 2000*[41].



(a) Original Image (b) Segmented Image

X. CONCLUSION

Color image segmentation is currently a very emerging research area in color image processing research due to its importance in the image analysis process. A number of algorithms exist for the same purpose among which the clustering based algorithms got frequent attention because of their simplicities. K-Means algorithm is one of the frequently chosen hard clustering algorithms for color image segmentation task. But, many times it results in segmentation which suffers from noises and also over segmentation. The reasons behind these are (1) an improper random selection of K value for K-Means algorithm, and (2) a wrong distance metric. By keeping these two things in mind, a novel color image segmentation approach has been developed where the value of K has been derived from the input image beforehand and "Cosine Distance Metric" is employed to deal with the distance calculations involved in K-Means algorithm. The experimental results show a better performance of our proposed approach. Also, the proposed approach through experiments on several satellite color images is found suitable for satellite color image segmentation. As the existence of noises and over segmentation both are the notable problems in satellite color image segmentation and as our approach is providing solutions for both of these two major problems, so, the proposed approach establishes a good framework for satellite color image segmentation.

REFERENCES:

[1] Image Segmentation. IntechOpen Book, Web link: http://www.intechopen.com/books/image-segmentation
[2] R. M. Haralick, L. G.Shapiro, "Image Segmentation Techniques",CVGIP, vol. 29, 1985,pp. 100-132.
[3] S. M. Aqil Burney, Humera Tariq, " K-Means Cluster Analysis for Image Segmentation", International Journal of Computer Applications (0975 – 8887) Volume 96– No.4, June 2014, pp. 1-8.
[4] Chris Solomon, Toby Breckon , "Fundamentals of Digital Image Processing", ISBN 978 0 470 84472 4.
[5] A.KaurBhogal, N. Singla, M.Kaur, "Color image segmentation using k-means clustering algorithm", International Journal on Emerging Technologies 1(2), 2010,pp. 18-20.
[6] A.K. Jain, M. N. Murty, P. J. Flynn, "Data Clustering: A Review",ACM Computing Surveys, vol. 31, pp. 264-323, Sep. 1999.
[7] D.J. Bora, A.K. Gupta, "A Comparative study Between Fuzzy Clustering Algorithm and Hard clustering Algorithm", International Journal of Computer Trends and Technology (IJCTT) V10(2):108-113, Apr 2014. ISSN: 2231-2803
[8] D.J. Bora, A.K. Gupta, "Clustering Approach Towards Image Segmentation: An Analytical Study", IJRCAR,Vol2,Issue 7,July 2014, pp. 115-124.
[9] D.J. Bora, A.K. Gupta, "A New Approach towards Clustering based Color Image Segmentation", International Journal of Computer Applications (0975 – 8887) Volume 107 – No 12, December 2014, pp. 23-30.
[10] G. Hamerly, C. Elkan, "Learning the k in k-means", In the proceedings of The Seventeenth Annual Conference Neural Information Processing Systems (NIPS), 2003.
[11] D.J. Bora, A.K. Gupta, "Effect of Different Distance Measures on the Performance of K-Means algorithm: An Experimental Study in Matlab" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, pp. 2501-2506.
[12] Tse-Wei Chen, Yi-Ling Chen, Shao-Yi Chien, "Fast image segmentation based on K-Means clustering with histograms in HSV color space," Multimedia Signal Processing, 2008 IEEE 10th Workshop on, Cairns, Qld, 2008, pp. 322-325.
[13] N. Dhanachandra, K. Manglem, Y.Jina Chanu, "Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm", IMCIP-2015, Procedia Computer Science 54 (2015), pp.764 – 771.
[14] Sadia Basar, Awais Adnan, Naila Habib Khan, Shahab Haider, "Color Image Segmentation Using K-Means Classification On RGB Histogram", Recent Advances In Telecommunications, Informatics And Educational Technologies,pp.257-262.
[15] R.Verma, S.Singh Rathore, A. Verma, "MRI Segmentation using K-Means Clustering in HSV Transform", International Journal of Advanced Research in Computer Engineering & Technology , Volume 4 Issue 10, October 2015, pp.3925-3929.
[16] D.J. Bora, A.K. Gupta, "A Novel Approach Towards Clustering Based Image Segmentation", International Journal of Emerging Science and Engineering, Volume-2 Issue-11, September 2014, pp. 6-10.
[17] M.Luo, Y.F.Ma ,H.J. Zhang, "A Special Constrained K-Means approach to Image Segmentation", proc. The 2003 Joint Conference of Fourth International Conference on Information Communications and Signal Processing and the Fourth Pacific Rim Conference on Multimedia,Vol.2,pp.738-742.
[18] G.Stockman and L.Shapiro, "Computer Vision", Prentice Hall, 2001.
[19] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the HSV color space for image retrieval," presented at IEEE International Conference on Image Processing, Rochester, New York, 2002.
[20] The HSV Colorspace, http://www.dig.cs.gc.cuny.edu/manuals/Gimp2/Grokking-the-GIMP-v1.0/node51.html
[21] HSV color solid cylinder, Web link: http://commons.wikimedia.org/wiki/File%3AHSV_color_solid_cylinder_alpha_lowgamma.png.
[22] Irwin Sobel, 2014, History and Definition of the Sobel Operator.
[23] Lihua Sun,Enliang Zhao, Long Ma,Li Zheng, "An Edge Detection Method Based on Improved Sobel Operator", Advanced Materials Research Vols. 971-973 (2014), pp. 1529-1532.
[24] A. Nosrat, Y. S. Kavian, "Hardware Description of Multi-Directional Fast Sobel Edge Detection Processor by VHDL for Implementing on FPGA", International Journal of Computer Applications (0975 – 8887) Volume 47– No.25, June 2012, pp. 1-7.
[25] A.Singhal, "Modern Information Retrieval: A Brief Overview",Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 24 (4), 2001, pp. 35–43.
[26] Pairwise distance between pairs of objects, Web link: http://www.mathworks.in/help/stats/pdist.html.
[27] Lloyd, Stuart P., "Least squares quantization in PCM", IEEE Transactions on Information Theory 28 (2),1982, pp. 129–137.
[28] A.K. Jain, M. N. Murty, P. J. Flynn, "Data Clustering: A Review",ACM Computing Surveys, vol. 31, Sep. 1999, pp. 264-323.
[29] A.K. Jain, "Data clustering: 50 years beyond K-means", Pattern Recognition Letters, Elsevier, 2010.
[30] Meyer, Fernand, "Topographic distance and watershed lines," Signal Processing, Vol. 38, July 1994, pp. 113-125.
[31] D.J. Bora, A.K. Gupta, "A New Efficient Color Image Segmentation Approach Based on Combination of Histogram Equalization with Watershed Algorithm", International Journal of Computer Sciences and Engineering Vol.-4(6), Jun 2016, E-ISSN: 2347-2693,pp.156-167.
[32] R. L. Solso,"Cognition and the Visual Arts ", Cambridge, MA: The MIT Press, 3rd ed, 1999.
[33] S.D. Chen and A. R. Ramli, "Preserving brightness in histogram equalization based contrast enhancement techniques", Digital Signal Processing, vol. 14, (2004), pp. 413-428.
[34] S.D. Chen, "A new image quality measure for assessment of histogram eualization-based contrast enhancement technique", Digital Signal Processing, vol. 22,2012,pp. 640-647.
[35] Raju. A , Dwarakish. G. S and D. Venkat Reddy, "A Comparative Analysis of Histogram Equalization based Techniques for Contrast Enhancement and Brightness Preserving", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.6, No.5 (2013), pp.353-366
[36] D.J. Bora, A.K. Gupta, F.A. Khan, "Color Image Segmentation Using An Efficient Fuzzy Based Watershed Approach",*SIPIJ,Vol.6,No.5,*pp. 15-34.
[37] R. Sakuldee, S. Udomhunsakul, "Objective Performance of Compressed Image Quality Assessments", Proceedings Of World Academy Of Science, Engineering And Technology Volume 26 December 2007 Issn 1307-6884, pp. 434-443.
[38] Q. Huynh-Thu, M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," Electronics Letters, vol. 44, no. 13,June 2008, pp. 800-801.
[39] M. Neubert, H. Herold, and G. Meinel, "Evaluation of remote sensing image segmentation quality—Further results and concepts," in Proc. 1st Int. Conf. OBIA, Salzburg, Austria, Jul. 2006.
[40] Berkeley Image Segmentation Dataset
Web link: http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/BSDS300/html/dataset/images.html
[41] Earth Science World Image Bank
Web link: http://www.earthscienceworld.org/images/search/browseresults.html?begin=0&num=1&numBegin=1&Category=37

AUTHOR'S PROFILE:

Dr. Anil Kumar Gupta is actively involved in Data Mining and Pattern Recognition research. His research work brings many new ideas to Classification and Clustering techniques. He has PhD in Computer Science. He is currently serving as HOD of the Department of Computer Science and Applications, Barkatullah University. He is also the chairman of Board of Studies (Computer Science) of the same university. Currently he is guiding 7 PhD research scholars. He has over twenty years of teaching experience.

Mr Dibya Jyoti Bora is a researcher in the field of Image Processing and Machine Learning, particularly Cluster Analysis and Image Segmentation. His research works solve many problems related to Color Image Segmentation process. He has contributed more than 12 research papers as his 1st authorship in international journals and national and international conferences including IEEE . Most of them are highly cited and applied in different technical fields including agriculture and medical imaging areas. As per academic details, he got distinction in his Graduation with honors in Mathematics, University 1st rank in PG(Information Technology) and currently pursuing PhD in Computer Science. He has qualified several lectureship ability tests including GATE CS/IT two times, UGC SET in Computer Science and Applications. He has 4 years of teaching experience in university PG level. He is currently teaching in the Department of Computer Science And Applications, Barkatullah University, Bhopal for PG students of CS and IT. Previously, he taught in the Computer Application Department, NEHU. He has two certifications in Image Processing Specialization, one from Duke University and the other from Northwestern University.