International Journal on Recent Researches In Science, Engineering & Technology
(Division of Computer Science and Engineering)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. It is an absolutely free (No processing charges, No publishing charges etc) Journal Indexed in JIR, DIIF and SJIF.
Research Paper
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

Volume 4, Issue 1, January 2016.

JIR IF : 2.54
DIIF IF : 1.46
SJIF IF : 1.329

AN IMPROVEMENT OF IMAGE SEGMENTATION USING K-MEANS CLUSTERING WITH KERNEL PCA FOR NATURE IMAGES

S.Saraswathi[1], Dr.M.Madheswaran[2]
[1]Assistant Professor, Department of Electronics and Communication Engineering
Mahendra Engineering College (Autonomous)
Mallasamudram, Tiruchengode - 637503 , Tamilnadu, India.
[2]Professor, Department of Electronics and Communication Engineering
Mahendra Engineering College (Autonomous)
Mallasamudram, Tiruchengode - 637503 , Tamilnadu, India.

[1]saraswathis@mahendra.info, [2]madheswaran.dr@gmail.com

ABSTRACT

In this paper, K-means clustering with kernel PCA and adaptive median filter is presented to segment and classify the noisy nature image that is taken from the satellite images. K-means clustering is utilized to segment the important area from the image background. It is mostly a partitioning technique applied to evaluate data and handles data observations as objects based on locations and distance between different input data points. In this process, kernel PCA is applied to extract the features from the image before applying the K-means clustering. Kernel PCA is the nonlinear form of PCA, which improved uses the complex spatial structure of high-dimensional features. The preprocessing method is used to remove the noise and unwanted background parts from the noisy nature image. Here, adaptive median filter is employed to remove the noise before image segmentation. Therefore the image quality is improved. Finally, nature image can be classified based on the RGB pixel values. By using MATLAB programming language this scheme is implemented.
Keywords: Image segmentation, K-means clustering, Kernel PCA, Adaptive Median filter.

I.INTRODUCTION

In nature, the natural images contain irresistible number of visual patterns produced by extremely varied processes of stochastic. Due to the surroundings, natural images are mostly noisy they were created [1][2]. A digital image processing has different algorithms on the image to develop the image quality by reducing noise and other unnecessary pixels and as well as to get further information on the image [3].

Image segmentation is very important step to evaluate the given image among the different image processing methods. Segmentation is utilized for object identification, estimation of occlusion boundary within movement or stereo systems, compression of image, editing of image [4-

6]. The process of dividing a natural image into regions or categories with standardized texture, normally referred to as image segmentation, is extensively established as a critical function for high-level image accepting, considerably minimizing the difficulty of content examination of images. A natural segmentation view is that we are attempting to establish which data set components obviously belong together [1][2] [4].

This is a problem called clustering. Usually, cluster is done in two ways. One way is partitioning and second way is grouping. In a data set to locate clusters is to discover relations between unlabeled data. The relation is a few data are in some way after to another that they can be clustered [2][3] . In today's scenario the natural image segmentation has become a very significant task. This natural image segmentation is one of the standard issues in computer vision. In images, image segmentation is characteristically utilized to find objects and boundaries and it is one of the mainly complicated tasks in image processing because it defines the final result quality of examination. Therefore, to overcome this problem an efficient K-means clustering algorithm with kernel PCA is applied to segment the nature image effectively. The goal of image segmentation in this process is to enhance the interpretability or awareness of information in images for human listeners, or to provide an enhanced input for other automatic image processing techniques.

K-means clustering is easy and computationally better than the hierarchical clustering. As well as it can also effort for huge number of variable. Hence it is necessary to initialize the suitable number of cluster k. Yet again, it is necessary to initialize the k number of centroid. Different initial centroid value is result various cluster. Therefore, variety of proper initial centroid is also a significant task. Classification of image decomposes the arithmetical properties of different image features and arranges data into categories [7][8]. K-Means clustering is also employed in brain tumor segmentation and its area calculation in brain MRI images is discussed in [9].

II. LITERATURE SURVEY

Bingquan Huo, et.al, presented a novel algorithm for segmentation of medical image. The multi-band active contour model based limit function is used to create the multilayer segmentation obtainable by recognizing the present state-of-the-art associated algorithms. Here a new penalty expression is employed to progress arithmetical stability and the step length is enhanced to progress efficiency. Cuadros Linares, et.al, proposed a new object-based empirical discrepancy error measure, known as Adjustable Object-based Measure (AOM). A penalty parameter is used which provides the way the capability to be more (or less) receptive in the over-segmentation presence.

Clustering is one method of utilized for land cover classification. Clustering is the task of objects into groups known as clusters so that objects from the identical cluster are more comparable to each other than objects from various clusters. Gurudatta V Nayak, et.al, presented an algorithm Utilizing K means clustering algorithm with color based thresholding for satellite image classification. The metrics of image quality are used that are overall accuracy, accuracy of users, accuracy of manufacturer, standard accuracy for manufacturer and user and Yu-guo Wang and Hua-peng Li presented an adaptive masking and a threshoding mechanism over every color channel to conquer over segmentation issue, before combining the segmentation from every channel into the final one. Thwe Zin Phyo et.al described a area classification for Landsat7 satellite image using K-means clustering.

This method is executed depending on color features with K-means clustering unsupervised algorithm. Aimi Salihai Abdul et. Al presented a colour image segmentation approach for discovery of malaria parasites that has been employed on malaria images of P. vivax species. In order to get the segmented red blood cells infected with malaria parasites, by utilizing partial contrast stretching, the images are first improved. Subsequently, an unsupervised segmentation method namely k-means clustering has been utilized to segment the infected cell from the background.

Pallavi Purohit and Ritesh Joshi applied K-means clustering algorithm for the assessment of student's academic performance for the reason of creating efficient decision by the councilors of student. Wei, B.C. et.al, proposed multi-objective nature-inspired clustering methods for segmentation of image. Three recognized significant stages contain design, intelligence, and option with respect to the issues of image segmentation clustering and design of multi-objective clustering algorithm. Saiful Islam and Dr. Majidul Ahmed used K-Means clustering, K-Medoids clustering and Hierarchical clustering to realize and comparisons between them for segmentation of Natural images. Naz, S et.al reviewed a Segmentation for noisy medical images with spatial probability, Fuzzy Local Information C-Means (FLICM) Clustering Algorithm, Novel Fuzzy C-Means Clustering (NFCM) and Improved Spatial Fuzzy C-Means Clustering (ISFCM) algorithm.

Median filters are used to decrease impulse noise level from bad images. Median filters are utilized to eliminate the salt-and-pepper noise. Vicky Ambule et.al proposed a method depending on median filter for the elimination of salt and pepper noise by its discovery followed by filtering in binary images has been presented. Yu-guo Wang, et.al, proposed a technique for purification model based on statistical examination theory which could cleanse training samples for enhanced classification of wetlands isolated sensing based on Artificial Neural Networks. With function of nonlinear mapping the BP ANN produces high-quality results of classification for versatile areas. Hui Zhang, et.al, used a hierarchical standard for the image sample evaluation of quality as a preliminary position. The model contains class, global and quality of image level, which regards both the quality of image and the class separability. On the image database experiments are conducted chosen from three synthesized databases and google earth.

III.PROPOSED WORK

Figure 1 shows the overall process of our proposed scheme. In this scheme, we take the one noisy nature image that is taken from the satellite digital images. The noisy image will be corrected by the image preprocessing methods. First, to smooth the image by removing the noise and unwanted parts from the background of given input nature image. Smoothing is frequently utilized to decrease noise within an image or to generate a less pixelated image. In the data an approximating function is created during the smooth the data set that tries to detain significant patterns, as removing noise or other fine-scale structures/rapid phenomenon., the signal data points are altered so individual points are decreased, and points that are lesser than the neighboring points are enhanced directing to a smoother signal in smoothing. In two important ways, Smoothing is utilized that can assist in analysis of data (1) From the data by being capable to extract extra information as long as the smoothing hypothesis is rational and (2) by being capable to give analyses that are both robust and flexible. Generally Smoothing is also based on a particular value describing the image, such as the mean value of the image or the median value. In this paper, we use the adaptive median filter to smooth the image by removing or reducing unwanted noise from background of image to improve the image quality. Due to different time the pixel values of the images cannot be the similar as the image is taken. In this state, the RGB pixel values of satellite image are dissimilar because of the time effects, light effects, and weather condition. Therefore, the adaptive median filter method is applied before the clustering to get clear image.

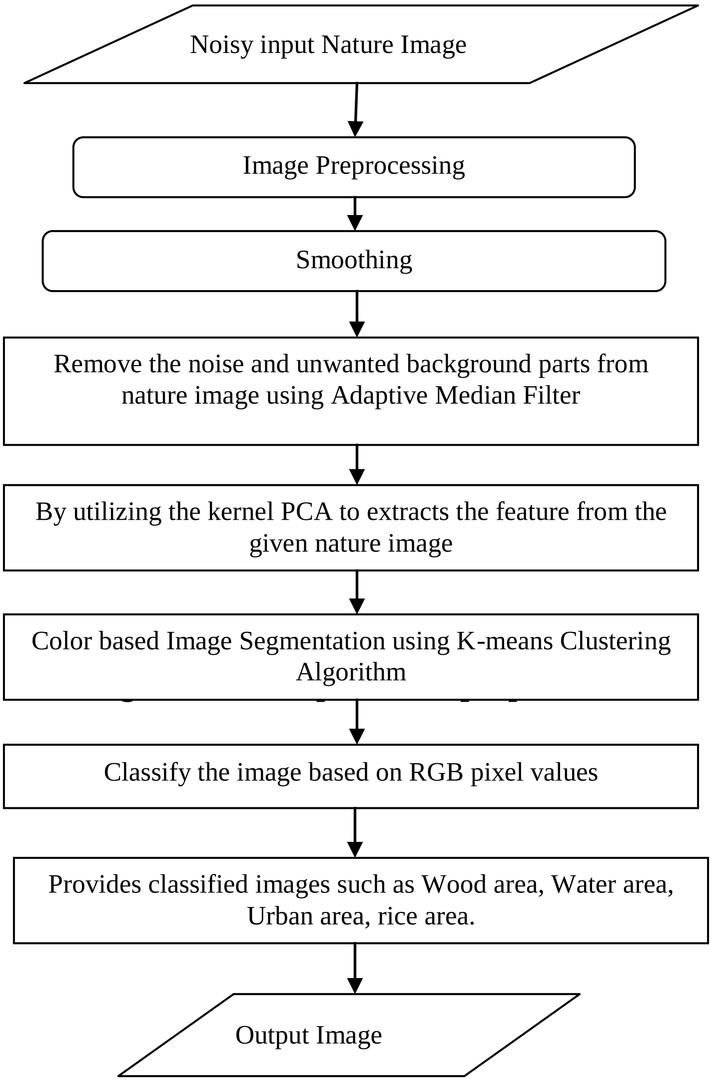

Figure 1 overall process of proposed scheme

The adaptive median filtering is employed extensively as an advanced technique compared with average median filtering. By impulse noise the Adaptive Median Filter executes spatial processing to establish which pixels in an image is pretended. In the image, by comparing every pixel to its surrounding neighbor pixels the Adaptive Median Filter categorizes pixels as noise. The size of neighborhood is adaptable and threshold is also adjustable for the comparison. From a majority of its neighbors a pixel is varied, and also not being structurally associated with those pixels, to which it is comparable and it is tagged as impulse noise. These pixels of noise are subsequently replaced in the neighborhood by the median pixel value of the pixels that have accepted the classification test of noise.The window size is employed to filter the image pixels is adjustable in nature that is the size of window is enhanced if the particular condition does not gather. The pixel is filtered utilizing the median of the window while the condition is meeting. Assume P_{xy} is the pixel of the ruined image, P_{max} is the maximum value of pixel and P_{min} is the minimum value of pixel in the window, W is current size of window employed, W_{max} is maximum size of window that is attained and window median allocated. Furthermore, this filtering algorithm executes in two steps as described

Step 1

- If $P_{min} < P_{med} < P_{max}$, after that the value of median is not an impulse, hence the algorithm going to step 2 to verify if the present pixel is an impulse.

- Else the window size is enhanced and then Step 1 is frequently occurred until the value of median is not an impulse thus the algorithm goes to Step 2; or else the maximum size of window is attained, here the median value is determined as the pixel value of filtered image.

Step 2

- If $P_{min} < P_{xy} < P_{max}$, after that the present value of pixel is not an impulse, therefore the filtered image pixel is unaffected.
- Else the image pixel is either equal to P_{min} or P_{max}(ruined), next the filtered imaged pixel is allocated the median value from Step 1.

This Adaptive Median filtering is establish to smooth the non disgusting noise from two-dimensional signals without dimming edges and protects details of image. This creates it mainly appropriate for increasing image quality.The kernel PCA algorithm can be applied in smoothed image to extract their features before clustering.

The newly proposed Kernel Principal Component Analysis (PCA) technique is utilized for feature extraction. To recognize the kernel PCA usefulness, mostly for clustering, examine that, as N points cannot in common be linearly divided in $d<N$ dimensions, they can approximately always be linearly divided in $d\geq N$ dimensions. That is, certain N points, X_i, if we plot them to an N-dimensional space with $\Phi(x_i)$. It is simple to make a hyperplane that seperates the points into random clusters. This Φ makes linearly independent points, therefore there is no covariance on which to execute eigen decomposition explicitly while we would in linear PCA.

A non-trivial, random Φ function is selected that is never evaluated clearly, permitting the option to utilize very-high-dimensional Φ's if there is no way to essentially estimate the data in that space in kernel PCA. As we usually attempt to avoid processing in the space Φ, which we call the feature space 'F', construct the N-by-N kernel

$$K(x, y) = (\Phi(x), \Phi(y))$$

Initially, let predictable novel features contains zero means, then the equation is

$$\frac{1}{n}\sum_{k=0}^{n}\Phi(x_i) = 0 \quad (1)$$

In high-dimensional feature space F the problem of PCA can be defined as the an n-sample estimation of the covariance matrix diagonalization

$$C=\frac{1}{n}\sum_{i=0}^{n}\Phi(x_i).\Phi(x_i)^T \quad (2)$$

Where $\Phi(x_i)$ is centralized nonlinear input variables mappings $x_i \in R^N$, $i = 1,\ldots,n$ here, the mapped data centralization in F

By orthogonal eigenvectors V, a diagonalization describes a unique data transformation to novel coordinates definite. Then locate eigenvalues ¸ eigenvectors of zero and non-zero $V\in F$ satisfying the equation of eigenvalue

$$CV = \lambda V \quad (3)$$

Substitute equation 2 in equation 3, then we will get

$$\frac{1}{n}\sum_{i=0}^{n}(\Phi(x_i).\Phi(x_i)^T).V= \lambda V \quad (4)$$

Resulting the problem of equivalent eigenvalue

$$n\lambda a = Ka \quad (5)$$

Where 'a' indicates the feature vector with ratios $a_1 \ldots a_n$ like that

$$V= \sum_{i=0}^{n} a_i\, \Phi(x_i) \quad (6)$$

$$K_{ij} = \Phi(x_i).\Phi(x_j) = K(x_i,x_j) \quad (7)$$

Where K is a (n x n) Gram matrix with the components
Standardizeing the solutions V^k analogous to the eigenvalues of non-zero ,λ_k of the matrix K, interprets into the state , $\lambda_k (a^k.a^k) = 1$. Finally, calculate the k-th nonlinear x principal component as the $\Phi(x)$ projection onto the eigenvector V^k

$$y_k(x) = V^k.\Phi(x) = \sum_{i=0}^{n} a_i^k \, K(x_i,x_j) \quad (8)$$

The normal steps of kernel PCA dimensionality decrease are followed as:

- Utilizing equation (7), to create the kernel matrix K from the training data set $\{x_i\}$.
- To employ equation (5) to resolve for the vectors a_i.
- Calculate the kernel principal components $y_k(x)$ utilizing Equation (8)

After that image feature extraction, contrast, the angular second moment, correlation and variance are estimated. After extracting the features from nature image, the k-means clustering can be applied for image segmentation.

K-means clustering method is used to image segmentation after feature extraction of given image. It is an iterative method that is utilized to divide an image into K clusters. K-means is easiest unsupervised learning algorithm, it solve the well recognized clustering difficulty. Figure 1 shows the working of K-means clustering algorithm in nature image.Through a definite number of clusters fixed a priori, the process follows a simple and easy way to categorize a data set. The major plan is to describe k centroids, one for every cluster. These centroids are placed in an archness method because of various location causes various result. Therefore, from each other the enhanced choice is to position them as much as probable far away. Next stage is to get every point pertaining to a given data set and relate it to the adjacent centroid. While there is no point is pending, first step is finished and previous group age can be done. Re-estimate k new centroids as clusters group's barycenters resulting from the previous stage. Then a novel binding is done between the similar data set points and the adjacent new centroid. Finally, the loop process is produced. As a result of loo, the k centroids are noticed to modify their position step by step until no more modifications are done. The centroids do not go anymore. K-Means clustering produces an exact number of displace, flat clusters. It is well suitable to creating global clusters. The K-Means technique is mathematical, non-deterministic, unsupervised and iterative.

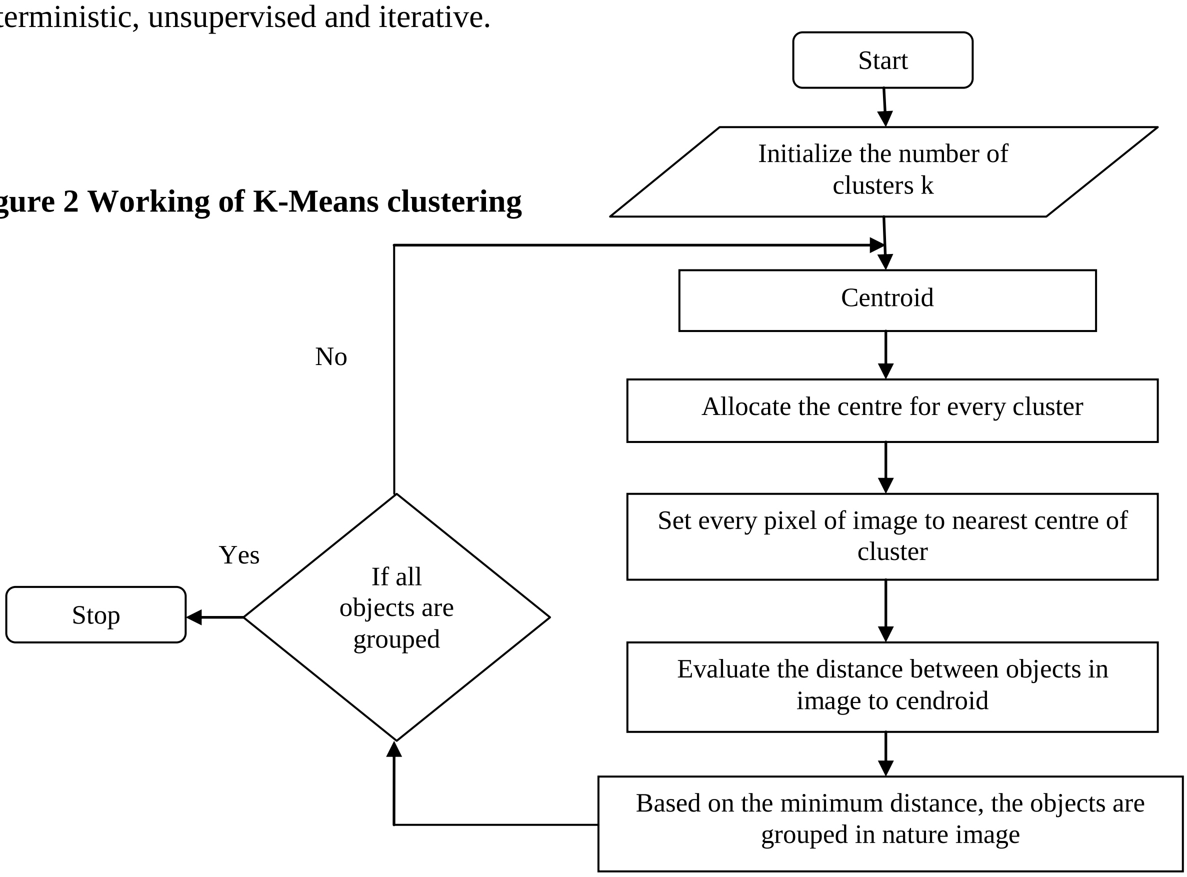

Figure 2 Working of K-Means clustering

A set of observations (m_1, m_2, …, m_n), where every observation is a d-dimensional actual vector, purpose of the k-means clustering to divides the n observations into $k \leq n$ sets $S = \{S_1, S_2, \ldots, S_k\}$ hence as to reduce the within-cluster sum of squares (WCSS). Its purpose is to discover

$$C_{k\text{-means}} = \sum_{i=1}^{n} \cdot \sum_{m \in S_i}^{n} [m - \mu_i]^2 \quad (9)$$

Figure 1 shows the working of K-means clustering. The K-Means algorithm contains following steps:

- Initialize the number of cluster k and allocate the center for every cluster at random
- Clusters k Centroids are selected from ***X*** arbitrarily.
- Distances between objects in image and cluster centroids are estimated.
- Every object is allocted to the cluster whose centroid is neighboring to it.
- By utilizing the formula Cluster centroids are updated in Equation (10)

$$V_i = \sum_{i=1}^{n} m_{ij} / n_i \; ; i \leq i \leq c \quad (10)$$

- From the updated cluster distances centroids are reevaluated.
- Based on the minimum distance, to group the objects in nature image.
- If all objects are grouped in image, then the run of algorithm is ended, or else the steps from 3 to 5 are repeated for possible objects movements between the clusters in nature image.

Even though *k*-means contains huge advantage of being simple to execute, it has a few disadvantages. A final clustering result quality is based on the random selection of primary centroid. Therefore if the initial centroid is arbitrarily selected, it can obtain various results for different initial centers. Hence the initial center is cautiously selected so that our desire segmentation can be obtained. And also complexity of computational is a further term which we require to consider as designing the *K*-means clustering. It depends on the number of data components, number of iteration and number of clusters. The image is segmented by the color using k-means clustering algorithm after the feature extraction of image. The input image is segmented into four clusters. The segmented clusters are classified based on the RGB pixel values.

In the color based image segmentation, the necessary function is to segment colors in an automated manner employing the L*a*b* color space and K-means clustering. The Figure 3 shows the color based image segmentation using the k-means clustering algorithm. In this process, the image is segmented based on the RGB color; the process is explained in the following steps.

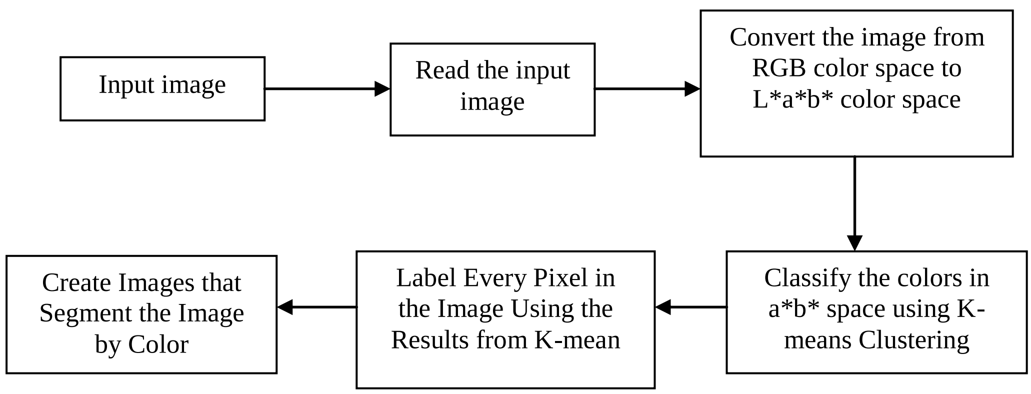

Figure 3 Color based image segmentation using K-means

Step 1: Read the input nature image.

In first step, read the nature image which is to be segmented that is taken from satellite. The image may be in PNG or JPG file format because the pixel values of PNG or JPG image file format is further precise.

Step 2: Convert the image from color space of RGB to color space of L*a*b*.

The color space of L*a*b* (CIELAB or CIE L*a*b*) permits to enumerate these image variances. The L*a*b* color space contains brightness layer or luminosity 'L*, chromaticity 'a*'denoting where color falls onwards 'b*'denoting where the color falls onwards the axis of blue-yellow. Every one of the information of color is the* and b* layers. It is precise the distinction between two colors utilizing squared Euclidean distance matric.

Step 3: Classify the colors using K-means Clustering.

Clustering is the one method to divide group of objects. K-means clustering handles every object as containing a position in space. It discovers separates such that objects within every cluster are as close to each other as probable, and as far away from objects in further clusters as probable. K-means clustering needs to specify the number of clusters to be divided. As the information of color subsists in a*b* space, with values of a* and b* the objects are pixels. K-means is employing the objects into four clusters utilizing the squared Euclidean distance metric.

Step 4: Label Every Pixel in the Image Using the Results from K-mean.

In the input nature image for each object, K-means restores an index parallel to a cluster.

Step 5: Create Images that Segment the Image by Color.

Employing pixel labels, in image the objects is divided by color, which can result in four clusters images.

Finally, we will get the output image that provides classified images such as Wood area, Water area, Urban area, rice area.

IV. RESULTS AND DISCUSSIONS

Figure 4 shows the original test image; we take the satellite nature image as input image to testing and the figure 5 shows that the smoothed image, in this figure the image is smoothed by reducing the noise from the original test image. For smoothing image, the adaptive median filter is used to reduce the noise efficiently. The adaptive median filter is one method of image preprocessing methods. This method is used to improve the quality of image. The satellite image may be not cleared because of light effects, time effects and weather condition. Therefore, this kind of image requires the filtering methods to filter the noise to give the clear image.

Figure 4 Original tested image

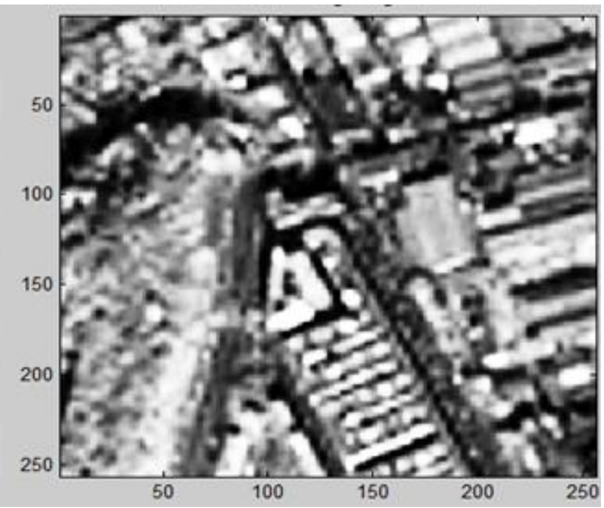

Figure 5: Smoothed image

Figure 6 Segmented image

In figure 6, the image is segmented using the k-means clustering. The input image is segmented into four clusters using K-means clustering algorithm using Matlab programming language. This K-means algorithm differentiates the given input image into various clusters of pixel in the feature space, every one of them described by its center. In the image every pixel is assigned to the adjacent cluster. After that the novel centers are estimated with the novel clusters.

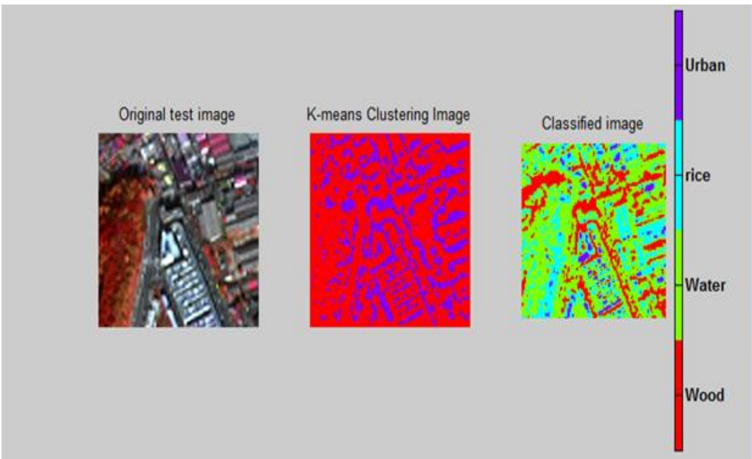

Figure 7 Final output

Figure 7 shows the final result of segmented images. The novel clusters are shown various color according color based segmentation technique. Four clustered images is assigned a exact range in order to obtain the classified images after clustering. The image is segment by the pixel range with color.

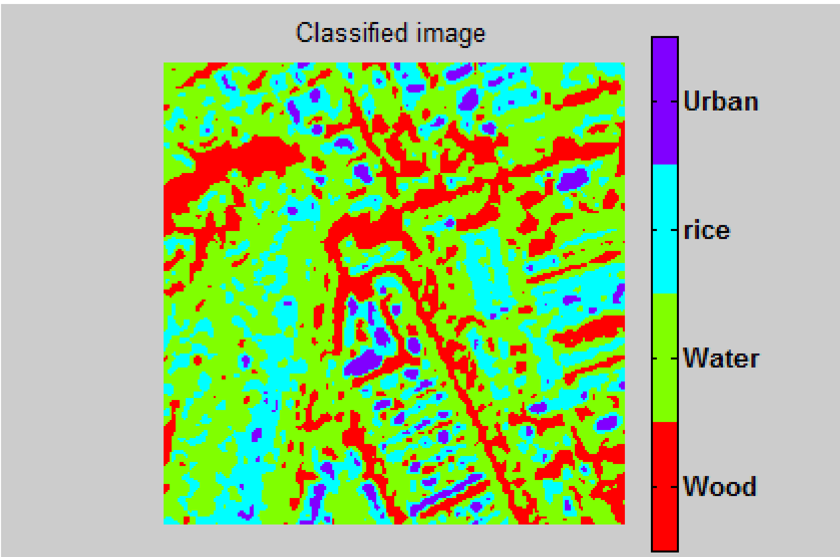

Figure 8 Classified image

In figure 8, given image is classified based on the RGB pixel values. Finally, the image is classified such as urban area, rice area, water area and wood area. The classified image results are reusable to classify the changes of Land cover in feature and it is used to recognize the amount of destroyed area of forest.

IV.CONCLUSION

The nature image segmentation has been described in this paper. In this process, adaptive median filter is applied to smooth the image by removing the noise effectively. Before the clustering, the feature is extracted using an effective kernel PCA algorithm. In the image data, using the k-means clustering algorithm the grouping of objects is performed. Here, the number of divided clusters based on the number of original image contents. For the clustering problem, the selection of K values is also very significant. The clusters are segmented by the color. In this paper, the segmented image has the four clusters because the user needs to segment this nature image into four clusters. Utilizing color based image segmentation; it is achievable to reduce computational cost avoiding for each pixel in the image. The clustered image results are simply to be used to classify the given image. Finally the results of clustered image must be assigned a precise range for every cluster according to their RGB pixel values. The image is classified into urban, water, wood and rice based on the RGB pixel values. The input image is segmented and classified efficiently. This scheme can provide the good quality of image.

REFERENCES

1. Saiful Islam and Dr. Majidul Ahmed 2013, ‘Implementation of Image Segmentation for Natural Images using Clustering Methods’, International Journal of Emerging Technology and Advanced Engineering, Vol.3, No.3, Pp.175-180.
2. Wei, B.C and Mandava, R 2010, ‘Multi-objective nature-inspired clustering techniques for image segmentation’, Cybernetics and Intelligent Systems (CIS), Pp.150 – 155.
3. Vicky Ambule, Minal Ghute, Kanchan Kamble and Shilpa Katre 2013, ‘Adaptive Median Filter for Image Enhancement’, International Journal of Engineering Science and Innovative Technology (IJESIT), Vol.2, No.1, Pp.318-323.
4. Bingquan Huo, Guoxin Li and Fengling Yin 2015, Medical and Natural Image Segmentation Algorithm using M-F based Optimization Model and Modified Fuzzy Clustering: A Novel Approach, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol.8, No.7, pp.223-234.
5. Cuadros Linares, O, Botelho, G, Rodrigues, F and Batista Neto, J 2015, An Adjustable Error Measure for Image Segmentation Evaluation, Graphics, Patterns and Images (SIBGRAPI), Pp. 80 – 86.
6. Naz, S, Majeed, H and Irshad, H 2010, ‘Image Segmentation using Fuzzy Clustering: A Survey’, International Conference on CET, Pp.181-186.
7. Gurudatta V Nayak, Anuja A Rao and NandanaPrabhu 2014, ‘K-Means Clustering Algorithm with Color-based Thresholding for Satellite Images’, International Journal of Computer Applications (0975 – 8887), Vol.105, No. 11, Pp.17-20.
8. Pallavi Purohit and Ritesh Joshi 2013, ‘A New Efficient Approach towards k-means Clustering Algorithm’, In International Journal of Computer Applications, (0975-8887), Vol. 65, No. 11. Pp.7-10.
9. Selvakumar, J, Lakshmi, A and Arivoli, T. 2012, ‘Brain tumor segmentation and its area calculation in brain MR images using K-mean clustering and Fuzzy C-mean algorithm’, IEEE International Conference on Advances in Engineering, Science and Management, Pp. 186-190.
10. Aimi Salihai Abdul, Mohd Yusuff Masor and Zeehaida Mohamed (2013), “Colour Image Segmentation Approach for Detection of Malaria Parasiter using Various Colour Models and

k-Means Clustering", In WSEAS Transaction on Biology and Biomedecine., vol. 10, Pp.41-55.

11. Yu-guo Wang and Hua-peng Li 2010, "Remote sensing image classification based on artificial neural network: A case study of Honghe Wetlands National Nature Reserve", Computer, Mechatronics, Control and Electronic Engineering (CMCE), Vol.5, No.1, Pp. 17 – 20.
12. Hui Zhang, Bangyu Li, Yan Wang, Jinfang Zhang and Fanjiang Xu 2013, 'Bag-of-Words model based image classification and evaluation of image sample quality in remote sensing', TENCON 2013 - 2013 IEEE Region 10 Conference (31194), Pp.1-4.
13. Md. Habibur Rahman and Md. Rafiqul Islam 2013, 'Segmentation of Color Image using Adaptive Thresholding and Masking with Watershed Algorithm', Informatics, Electronics & Vision (ICIEV), Pp. 1-6.
14. Yu-guo Wang and Hua-peng Li 2010, 'Remote sensing image classification based on artificial neural network: A case study of Honghe Wetlands National Nature Reserve', Computer, Mechatronics, Control and Electronic Engineering (CMCE), Vol.5, No.1, Pp. 17 – 20.
15. Wang, Q Wang 2010, 'Image Segmentation Based on Multi-scale Local Feature', on Image and Signal Processing (CISP),Vol.3, Pp.1406-1409.