Advanced Research in Electrical and Electronic Engineering
p-ISSN: 2349-5804; e-ISSN: 2349-5812 Volume 2, Issue 11 July-September (2015) pp. 68-72

http://www.krishisanskriti.org/Publication.html

Image Segmentation Method using K-means Clustering Algorithm for Color Image

Nameirakpam Dhanachandra[1] and Yambem Jina Chanu[2]

[1,2]NIT Manipur
E-mail: [1]dhana.namei@gmail.com, [2]jina@nitmanipur.ac.in

Abstract—*Many works has been done in area of image segmentation. But many are trying to improve for a better method. From image segmentation methods, segmentation using clustering technique is one of the commonly used image segmentation method. Many methods are there in the clustering technique, but k-means is most use and simplest method. But most of the images required pre-processing before applying for segmentation process. In this paper, a new image segmentation methods for color image is proposed where it uses local histogram equalization and k-means clustering. Local histogram equalization is used to enhance the color image by using the image information, by taking transformation of the image pixels. Then k-means clustering algorithm is apply for the segmentation of the color image. And lastly the proposed method is compared with classical methods like k-means, Fuzzy c-means, Subtractive clustering.*

1. INTRODUCTION

Image segmentation is a process of partitioning a digital image into different multiple segments, so it is a method to classify the pixels of an image correctly in a decision oriented application. It divides an image into a number of discrete regions such that the pixels have high similarity in each region and high contrast between regions. It is a valuable tool on many field including health care, medical image processing, traffic image, pattern recognition etc. There are different techniques for image segmentation like threshold based, edged based, clustering based, neural network based etc. From the different technique one of the most commonly used efficient methods is clustering method. Image segmentation using clustering [1] [2] can be divided into different types like K-means clustering, Fuzzy C-means clustering, Subtractive clustering methods etc.

One of most used clustering algorithm is k-means clustering. It is simple and computationally faster than the hierarchical clustering. And it can also work for large number of variable. But it produces different cluster result for different number of number of cluster. So it is required to initialize the proper number of number of cluster, k. Again, it is required to initialize the k number of centroid. Different value of initial centroid would result different cluster. So selection of proper initial centroid is also an important task. Many methods for segmentation of color images had been exited. But most of them are application based. So still there is no universal method for color image segmentation till now.

The objective of image segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyze. But most of the input images are taken from many areas based on the different applications. And some of the images are not clearly seen or some of them are present noise or some of them are of poor quality. So they need pre-processing before applying for segmentation. There are so many pre-processing techniques. These techniques can be divided into different types based on the type of processing like point processing, mask processing, noise removal etc. Depending on the type of input image, one of the those techniques are used. But point processing techniques are most commonly used because of many advantages.

2. RELATED WORKS

There have been many works done in the area of image segmentation by using different methods. And many are done based on different application of image segmentation. Many methods of image pre-processing techniques had been already existed. But with increase in the usage of image processing nowadays many are trying to produce better and new efficient method to meet the requirement different application. K-means algorithm is the one of the simplest clustering algorithm and there are many methods implemented so far with different method. And many researchers are also trying to produce new methods which are more efficient than the existing methods, and shows better segmented result. Some of the existing recent works are discussed here.

Khaled Hammouda[3] has done a survey on different techniques of data clustering in," A comparative study of data clustering technique". It implemented different data clustering technique- k-means clustering, Fuzzy c-means clustering, and Mountain clustering and subtractive clustering. It analyze these different technique, it is found that k-means and Fuzzy c-means are preferred when the number of cluster is known, and in such cases Mountain clustering is not usually used because of high number of dimension due to its exponential

proportionality to the dimension of the problem. But if the number of cluster is not known, Mountain clustering method is used. Again it is conclude that subtractive clustering is more advantages than Mountain clustering. Lastly it is stated that these different technique can be used in conjunction with other neural or fuzzy system so that it can improve the system performance.

H.D. Cheng, X.H.Jiang, Y.Sun and Jingli Wang [4] did a literature survey on various color image segmentation technique. They discussed about different approaches of segmentation of monochrome images. They had done comparision on many techniques and point out the advantages and disadvantages of them. They conclude HSI based segmentation as a better method, but still it had disadvantages like unstability of hue at low saturation. Lastly they conclude fuzzy based color image segmentation as promising means for color image segmentation.

Ibrahim A. Almerhag, Idris S Feghi and Ali A Dulla [5] proposed a new method of k-means clustering algorithm in," A modified k-means clustering algorithm for gray image segmentation". It proposed a new modified method of conventional k-means algorithm. In conventional k-means algorithm, usually cluster centers are randomly initialized. But here it intruded a new method to initialized the initial cluster centers. It uses minimum and maximum data points in the given data set to initialize the cluster centers. Then comparison is done between the standard and proposed k-means algorithm and it is found that the proposed algorithm have effective and more robust than the traditional k-means algorithm.

Nicholas Sia Pik Kong, Haidi Ibrahim and Seng Chun Hoo [6] had a survey on the histogram equalization techniques and implemented the methods like Histogram Equalization, Local Histogram Equalization, Mean Brightness prevention Histogram Equalization and Modified Histogram Equalization . They conclude that LHE are suitable where it required to reveal small and hidden image content. And MBPHE are useful when it aim to preserve the overall mean brightness of the image.

3. LOCAL HISTOGRAM EQUALIZATION

Local histogram equalization [7] is a method to improve the contrast of an area in an image by shifting intensity values so that there are an equal number of pixels in an image in each intensity. This algorithm tabulates the histogram for a collection of neighboring pixels, sometimes called a window and then assigns the pixel to the new histogram level. It uses the transformation derived from the image histogram to transform all the pixels. So it improves the image by transforming each pixel with a transformation function derived from a neighborhood region.

Neighborhood size greatly affects the reference image, since the intensity of each resulting pixel depends on the neighborhood. The more pixels in the neighborhood that are darker than the current pixel, the greater in intensity the current pixel becomes in the reference image. Likewise, the shape of the kernel determines the number of pixels in the neighborhood.

1. Load image.
2. Assign window size.
3. Padding of zeros.
4. Choose first window.
5. Find cumulative distributive function value of each pixel value in that window.
6. Replace the pixel value with Cumulative distributive function value.
7. Continue the process for every window.

4. K-MEANS CLUSTERING

Clustering is a method to divide a set of data into a specific number of groups. It's one of the popular method is k-means clustering. In k-means clustering, it partitions a collection of data into a k number group of data [8] . It classifies a given set of data into k number of disjoint cluster. K-means algorithm consists of two separate phases. In the first phase it calculates the k centroid and in the second phase it takes each point to the cluster which has nearest centroid from the respective data point. There are different methods to define the distance of the nearest centroid and one of the most used methods is Euclidean distance. Once the grouping is done it recalculate the new centroid of each cluster and based on that centroid, a new Euclidean distance is calculated between each center and each data point and assigns the points in the cluster which have minimum Euclidean distance. Each cluster in the partition is defined by its member objects and by its centroid. The centroid for each cluster is the point to which the sum of distances from all the objects in that cluster is minimized. So K-means is an iterative algorithm in which it minimizes the sum of distances from each object to its cluster centroid, over all clusters.

Let us consider an image with resolution of x×y and the image has to be cluster into k number of cluster. Let $p^i(x, y)$ be an input pixels to be cluster for each R,G,B channel where i=3 and $c_k{}^i$ be the cluster centers of i channel of color image. The algorithm for k-means [9] clustering is following as:

1. Load color image.
2. Separate the image into Red, Green, Blue channel.
3. Initialize number of cluster k and centre for each channel.
4. For each pixel of an image, calculate the Euclidean distance d, between the center and each pixel of an image using the relation given below. *i* represent each channel and *k* represent particular cluster.

$$d_k = \sum_{i=1}^{3} \|p^i(x,y) - {c_k}^i\|$$

5. Assign all the pixels to the nearest centre based on distance d.
6. After all pixels have been assigned, recalculate new position of the centre of each channel by averaging the pixels.

$${c_k}^i = \frac{1}{k} \sum_{y \in {c_k}^i} \sum_{x \in {c_k}^i} p^i(x,y)$$

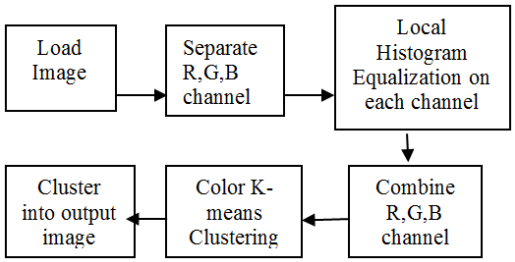
7. Repeat the process until it satisfies the tolerance or error value.
8. Reshape the cluster pixels into image.

Although k-means has the great advantage of being easy to implement, it has some drawbacks. The quality of the final clustering results is depends on the arbitrary selection of initial centroid. So if the initial centroid is randomly chosen, it will get different result for different initial centers. So the initial center will be carefully chosen so that we get our desire segmentation. And also computational complexity is another term which we need to consider while designing the K-means clustering. It relies on the number of data elements, number of clusters and number of iteration.

5. PROPOSED METHOD

The image to be segmented are mostly of poor quality like nosy image, low contrast etc. So they are generally pre-process before applying for the image segmentation process. Here we have used local histogram equalization to enhanced the contrast of the image i.e. to improve the quality of the image. Then improved image is applied for segmentation process using k-means clustering algorithm. For k-means clustering, we need to initialize the number of cluster and initial centroid, which are randomly initialize by user. For color image segmentation, we divide the image into three channel Red(R), Green(G) and Blue(B). The algorithm for proposed algorithm is given below.

1. Load the image.
2. Separate R,G,B channel.
3. Apply Local histogram equalization to each channel.
4. Apply K-means clustering algorithm.
5. Update the center and cluster the image until it satisfy the condition.
6. Reshape the cluster into output image.



6. RESULTS

Author Local histogram equalization technique required initialization of the window size. By varying the size of the window, we can check for better output image. Generally more is the size of the window, better will the output image. First we have taken the window size of 50x50. Then we check the output again by taking the window size of 75x75, 100x100 and 150x150. But with the increase in the size of the window, the computation time will increase. So to increase the quality of the output image by increasing the window size will cost computation time. So one have to decide between the computation time and quality of the image based on the type of the application. Matlab is used to implement the proposed method. The output result from the implementation at Matlab is shown in the fig.

Again we need to initialize number of cluster k and centre of each R,G,B channel. So these values are randomly initialize. We have taken k=3.

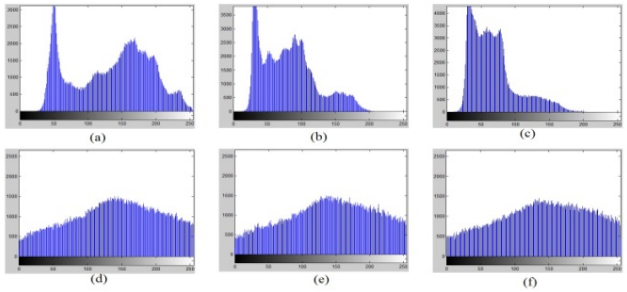


Fig. 1: Histogram diagram of Lena with window size 75x75.(a),(d) Red, (b),(e) Green, (c),(f) Blue

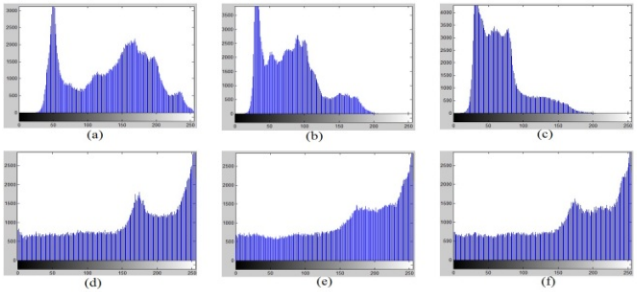


Fig. 2: Histogram diagram of Lena with window size 100x100.(a)(d) Red, (b)(e) Green, (c)(f) Blue.

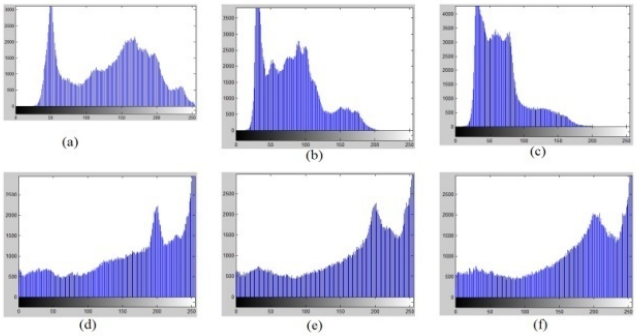


Fig. 3: Histogram diagram of Lena with window size 150x150.(a)(d) Red, (b)(e) Green, (c)(f) Blue.

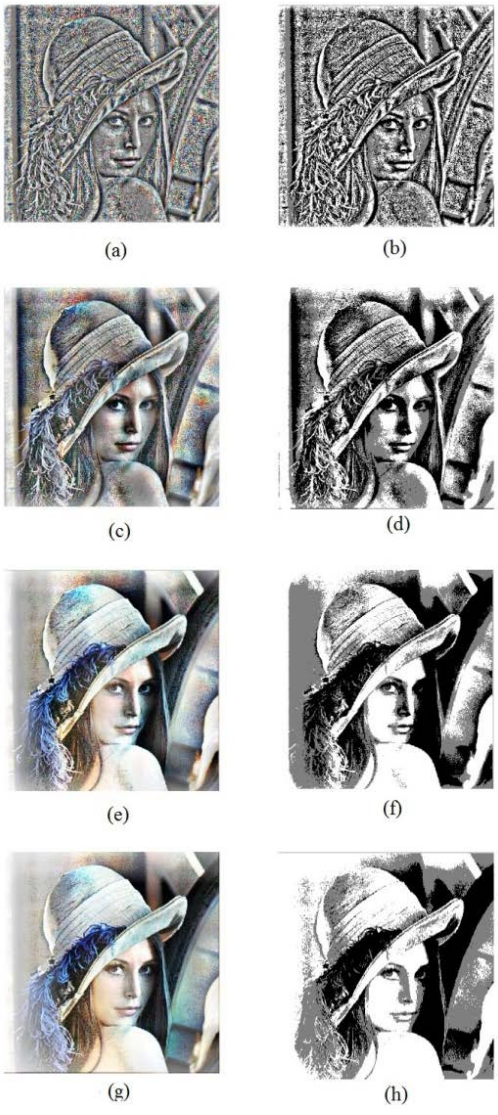


Fig. 4: First column represent Local Histogram Equalization output and Second column represent Segmentation output. (a)(b) window size 50x50, (c)(d) window size 75x75, (e)(f) 100x100, (g)(h) 150x150.

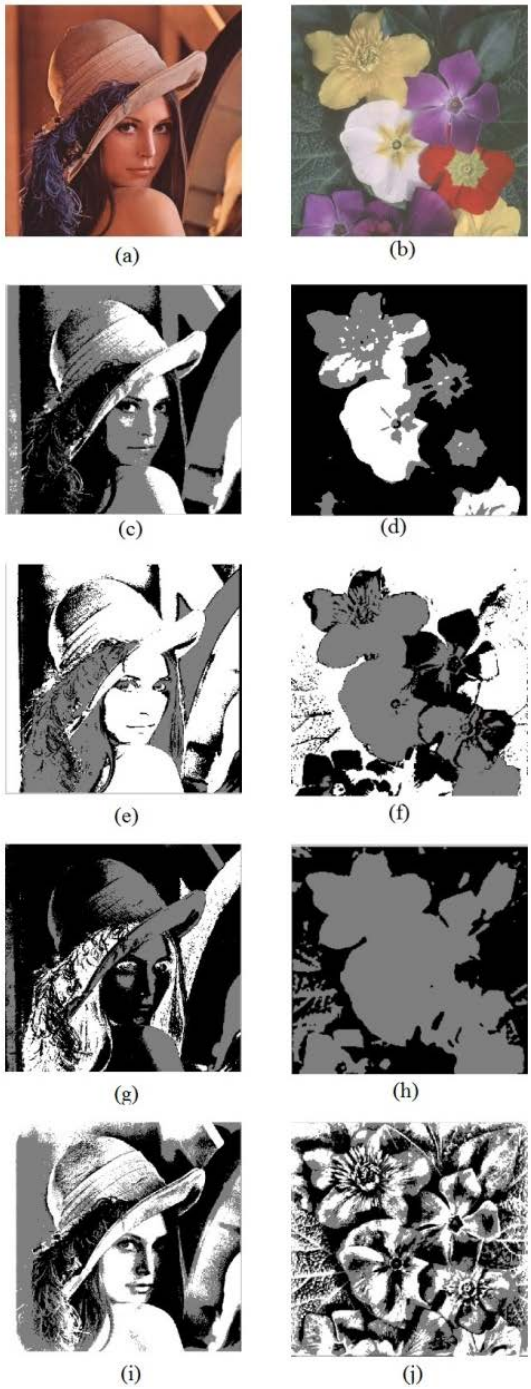


Fig. 5: (a)(b) Original image, (c)(d) K-means, (e)(f) Fuzzy c-means, (g)(h) Subtractive clustering, (i)(j) Proposed method

The proposed method is compared with k-means, fuzzy c-means and subtractive clustering method. Implementation is done using Matlab and output results are shown in the fig. 5. As shown in the fig., different output is getting for different method. But segmentation using proposed is segmented more efficiently as compared to others. But efficiency of the segmentation are depend mostly based on the type of

application. So we have to choose the segmentation method based on the type of the application we are using.

7. CONCLUSION

This paper proposed a method of color image segmentation using k-means clustering algorithm. We have used local histogram equalization method to adapt the contrast based on the input image. Pixel values of the images are transformed using cumulative distribution function. We have compare the result of the output image by varying the window size. Result shows that quality of the segmentation is better when the size of the image is large. The proposed method is compare with some of the classical methods like K-mean, Fuzzy C-means, Subtractive clustering.

In the future, we can introduce a new and better method of local histogram equalization to improve the computation time. Again more efficient method of histogram equalization base method can be approach in the future. The initial value of number of cluster and initial center plays an important role, so we can try in the future to find the number of cluster and initial center using the information from the histogram.

REFERENCES

[1] Jose Alfredo F. Costa and Jackson G. de Souza, "Image Segmentation through clustering based on natural computing techniques," Federal University of Rio Grande do Norte, Brazil

[2] D.L. Pham, C.Y.Xu, J.L. Prince," A survey of current methods in medical image segmentation", in Annual Review of Biomedical Engineer, 2000.

[3] Khali Hamuda, " A Comparative study of data clustering technique", Department of system Design Engineering, University of Waterloo, Canada.

[4] H,D, Cheng, X.H. Jiang, Y. Sun and J. Wang," Color image segmentation: advances and prospects," Pattern Recognition, 34: 2259-2281.

[5] Ibrahim A. Almerhag, Idris S. E, Feighi and Ali A Dulla, " Modified k-means clustering algorithm for gray image segmentation", Naseer International University, Tarhoona, Libya.

[6] Nicholas Sia Pik Kong, Haidi Ibrahim and Seng Chun Hoo," A literature Review on Histogram Equalization and its Variation for Digital Image Enhancement", in Internation Journal of Innovation, Management and Technology, vol 4, No. 4, August 2013.

[7] Huiz Zhu, Francis H.Y. Chan and F.K.Lam, " Image Contrast Enhancement by Constrained Local Histogram Equalizatio",in Computer Vision and Image Understanding, Vol. 73, No. 2, Febbruary,, pp. 281-290, 1999..

[8] Sorin Israil," An overview of clustering methods", in With Application to Bioinformatics.

[9] Aimi Salihai, Abdul Yusuff Masor, Zeehaida Mohamed, " Color image segmentation approach for detection of Malaria parasites using various color models and k-means clustering", in Wseas Transaction on Biology and Biomedicine, vol. 10, January 2013.