Image Segmentation and Analysis using Clustering Algorithms

Md. Shadab Hussain (16101033)
Vivek Yadav (16101057)

Computer Vision Course Project

Dr. Shyama Prasad Mukherjee

International Institute of Information Technology, Naya Raipur

(A Joint Initiative of Govt. of Chhattisgarh and NTPC)

Email: iiitnr@iiitnr.ac.in, Tel: (0771)2474040, Web: www.iiitnr.ac.in

CONTENTS

Table List

1. Problem Statement:

There are many image segmentation techniques to partition a digital image into segments available nowadays such as Thresholding, Clustering methods, Compression-based methods etc. But we found that some methods, especially K-Means clustering and are not working perfectly in some circumstances of image segmentation process.

There are two main problems that affect the performance of these two image segmentation methods. One of the problems is that a general K-Mean algorithm cannot correctly and automatically generate the value of K, which determines how many clusters will be segmented for the image. Randomly picking it will get us a value that's either too big or too small. In this case, some images, that have a relatively small RGB scale, may be over segmented.

2. Introduction:

Image segmentation has been considered as first step for image processing. An efficient segmentation output makes it easier in the next steps of the image processing. Therefore, many methods of image segmentation have been already proposed. Clustering is one of the commonly and widely used image segmentation approached because of its simplicity and efficiency. There are different clustering algorithm like K- means, fuzzy c-means, spectral clustering, expectation and maximization etc. [1] These different clustering approaches are proposed from different perspectives and are designed for different purposes. But existing clustering algorithm do require user-specified parameters as input. And the clustering performances depend highly on these user-specific parameters. One example of such parameter is the number of cluster, which is required by many clustering algorithms, e.g. K-means, fuzzy c-means etc.

In this paper, we have introduced new image segmentation techniques that are based on the predefined correctly predicted K values and inbuilt K-means clustering algorithm functions. The kernel function is applied in the subtractive algorithm to find the cluster centroid and these centroids are used in the K-means clustering algorithm.

K-means Algorithm:

K-means clustering algorithm is most commonly used clustering algorithms because of the simplicity and fast computation. It groups a collection of data into a k number group of data. The pixel and the centroids are the two important parameters which defined a cluster.

It is can be defined as an iterative algorithm in which it tries to minimizes the sum of distances from each object to its cluster centroid, over all clusters. Let us consider an image and number of cluster is defined as *k*. Let *P(x)* be the input pixels to be clustered and *c* be the cluster centroid. The algorithm for *k*-means clustering is given below:

1. Initialize the number of cluster k and the centroid for each cluster.

2. For each pixel of an image, calculate the Euclidean distance *d* can be calculated as in (1)

$$d = \| p(x) - c_k \| \quad (1)$$

where *p(x)* is x^{th} input pixel of the image, c_k is the center for k^{th} cluster.

3. Assign all the pixels to the nearest centroid based on the distance *d.*

4. After all the pixels have been assigned, a new position of the centroid are calculated using the relation given in (2)

$$c_k = \frac{1}{k} \sum_{y \in c_k} \sum_{x \in c_k} p(x) \quad (2)$$

5. Repeat the process until it satisfies the tolerance value or error value.

Although k-means has many advantages of being easy to implement and fast computation, it do has some drawbacks. The quality of the final clustering results of the k-means algorithm highly depends on the initial selection of initial centroid and number of cluster.

Performance Parameters

- **Mean Squared Error:**

  It is considered as a standard performance measurement of the output image. It defines how much the output image has deviated from the input image. A smaller value of RMSE defined that the image is of good quality.

- **Peak Signal-to-noise Ratio:**

  It can be defined as the ratio is the proportion between maximum attainable powers and the corrupting noise that influence likeness of the images. It measures the quality of the output image. A smaller value of PSNR means that the output image has poor quality.

3. Literature Review:

On reviewing the following papers we found that their implementation had some limitations. The papers were implemented in different platforms. The used methods are also mentioned in brief.

Table 1 : Limitation of existing Papers

S.no	Name Of The Paper	Implemented In	Papers Method	Limitation
1.	Adaptive K-means Image Segmentation Based On Meta Heuristic Algorithm [Somporn Tiachareon,IEEE 2018]	Python	Have used Meta Heuristic Algorithm and Genetic Algorithm	Have only used MTA and GA. Uses thresholding parameter only
2.	Improvement And Analysis Of Image Segmentation Techniques [Haoming Chen, Xue Han, CPMT 2018]	MATLAB	Have used spatial features	Complex calculation. High MSE values
3.	A new approach of Image Segmentation Method Using K-means and Kernel Based Subtractive Clustering Methods [Nameirakpam Dhanachandra, Yambem Jina Chanu, International Journal Of Applied Engineering Research 2017]	MATLAB	Uses kernel function in Subtractive Algorithm	Complex calculations for calculating kernel functions. High MSE values
4.	Image Segmentation Using K-means Algorithm and Subtractive Clustering Algorithm [Nameirakpam Dhanachandra, Yambem Jina Chanu, Kumanthem Manglem, International Journal Of Applied Engineering Research 2016]	MATLAB	Uses Subtractive Algorithm with K-means	Uses median filter and predefined value of K.

4. Proposed Work:

We proposed a method for finding the correct and predefined value of *K* using peak histogram values, of gray scale image and segment the image using MATLAB inbuilt functions.

The number of *K* values was generated randomly for any type of images segmentation even if not compatible. For example, an image with only colors red and green that was given an oversized K's value (e.g. $K = 10$) for segmenting will not work well with clusters.

Because a color image has three colors based histograms, which are red, green, and blue channel. It is hard to find/distinguish some valuable data in these histograms to help us to choose the K-value. Therefore, we transfer the image from a color image to a gray scale image to reduce the amount of data.

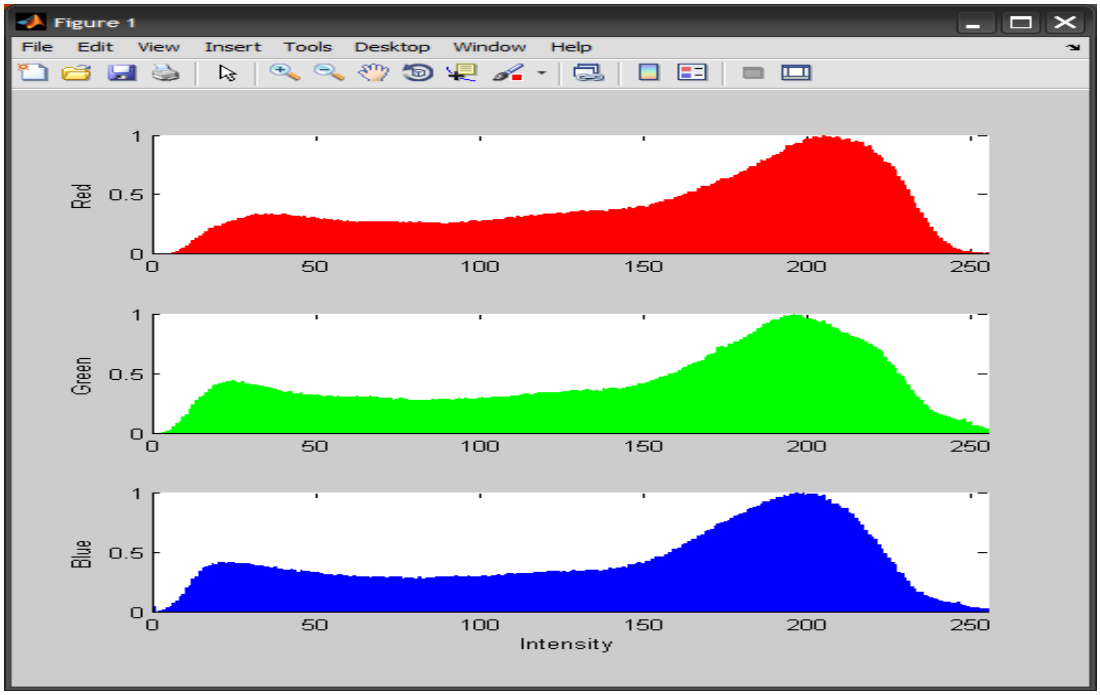

Figure 1. Histogram Of RGB Image

Then, we make a gray-level histogram for the image. After that, we can easily find that there are some "peaks" on the histogram, which are the local maxima of the gray-level intensity frequency. The amount of number of peaks will be used to decide how many K-values we will use in the image segmentation process. Because the peaks mean that the pixel count with those tone values were high which means there are many pixels that has same RGB values refer to the original color image.

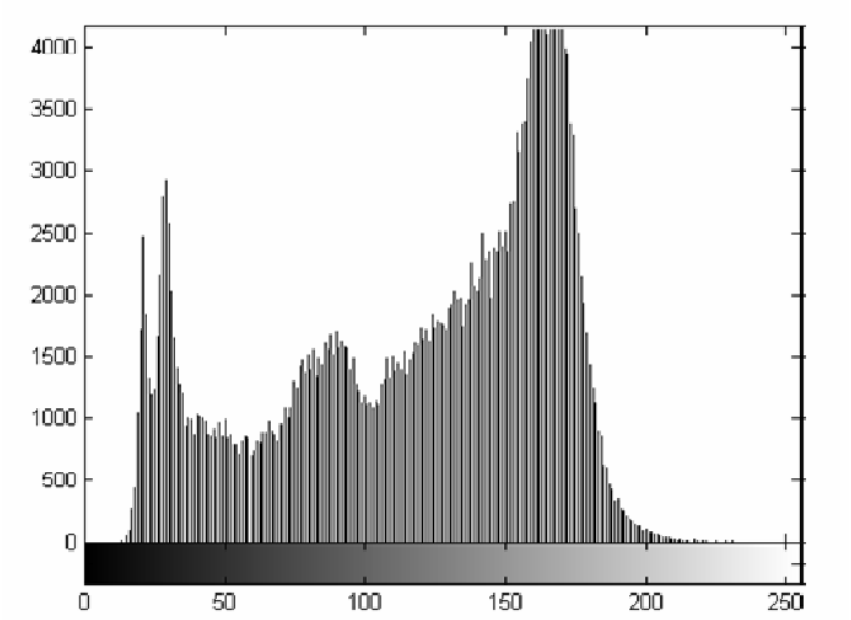

Figure 2. Histogram of Gray Scale Image

Proposed Algorithm:

1. Read the image from
2. Convert it to grayscale image from rgb image and find the number of peaks of its histogram.
3. Assign that number as *K* value.
4. Using the MATLAB functions segment the image and specifying the *K*-value as above.
5. Calculate the MSE and PSNR values of the segmented image, using MATLAB function. Develop the reference image from the input image only.
6. Use Gaussian filter with standard deviation 0.2 for generating the refrenced image.

5. Result Analysis:

The proposed method is implemented in Matlab and for the experimental purpose, we have used random images. In the implemented K-means clustering algorithm, the parameter K are need to give a prior value. Also, the paper mainly focuses on pixel distribution rather than clarity of segmented image.

Again, the number of K-value can be determined correctly but not randomly generalized. The number of clusters will be suitable for any images for the segmentation. How many pixels an image has, how much centroid will be generated. There will be no shortage or excess clusters a segmented image would have. From the Table 2, it can be observed that the proposed method have the small values of MSE and high values of PSNR.

Images	Method Of Implementation	MSE	PSNR
Station (Image 1)	Implemented [A New Approach of Image Segmentation Method Using K-Means and Kernel Based Subtractive Clustering Methods][1]	568.2683	20.5853
	Proposed	398.1031	22.1308
	Implemented [A New Approach of Image Segmentation	579.2950	20.5018

Hyderabad (Image 2)	Method Using K-Means and Kernel Based Subtractive Clustering Methods][1]		
	Proposed	423.8677	21.8585
Coins (Image 3)	Implemented [A New Approach of Image Segmentation Method Using K-Means and Kernel Based Subtractive Clustering Methods][1]	552.6088	20.7066
	Proposed	436.9038	21.7269
Lady (Image 4)	Implemented [A New Approach of Image Segmentation Method Using K-Means and Kernel Based Subtractive Clustering Methods][1]	556.7107	20.6745

	Proposed	418.7499	21.9113
Tree (Image 5)	Implemented [A New Approach of Image Segmentation Method Using K-Means and Kernel Based Subtractive Clustering Methods][1]	589.5097	20.4259
	Proposed	405.9077	22.0465

Table 2: Comparison of the proposed method with implemented method.

Figure *3* shows the graph plot of the RMSE values. Since RMSE measures the deviation between the images, the lowest value means less deviation. From the graph, it is clearly shown that the proposed method outperforms the implemented methods. Similarly, figure *4* shows the graph plot of PSNR values and it measures the likeliness between the images. As we observed, the proposed method have high PSNR value in all the images. Thus, these have proved the effectiveness of the proposed method.

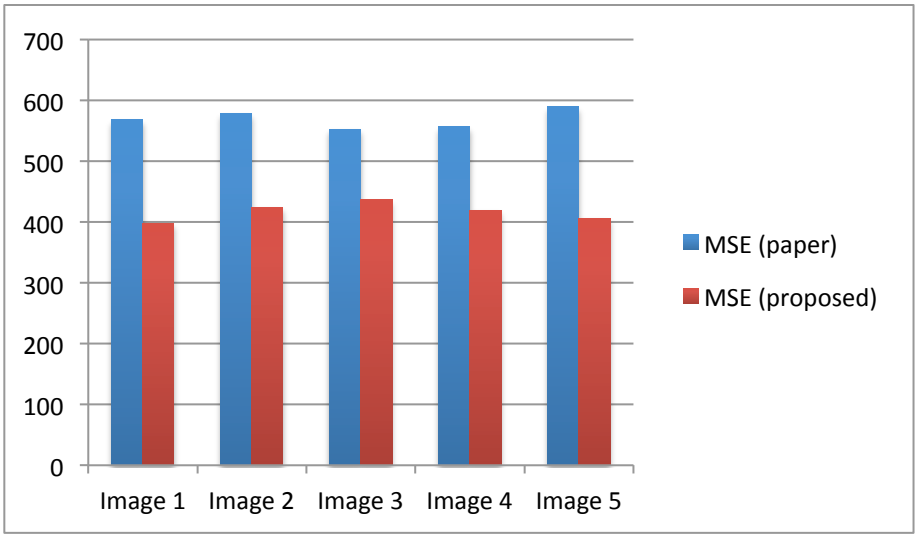

Figure 3 – Comparison Of MSE values

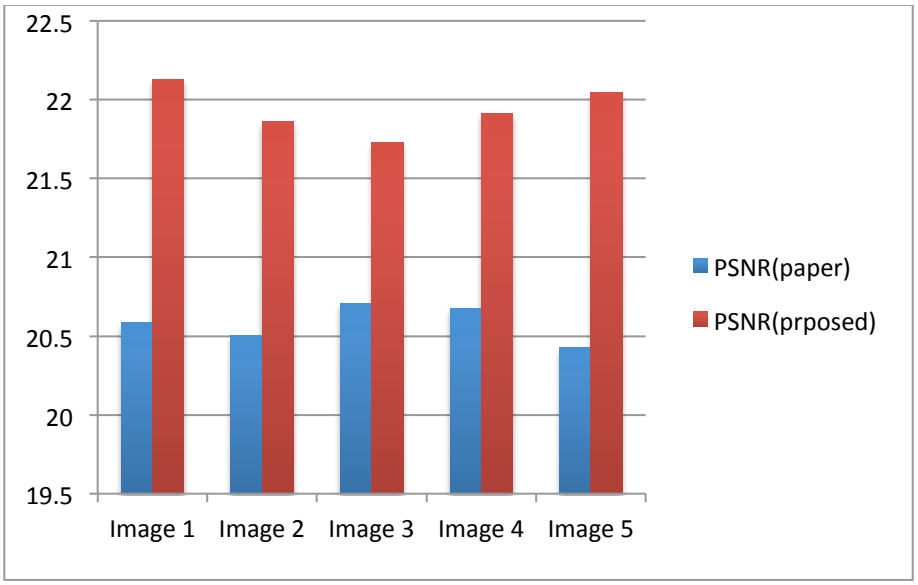

Figure 4 – Comparison Of PSNR values

6. Conclusion and Future Work:

A new image segmentation method is proposed based on the k-means and Matlab functions. The ambiguity caused by the color similarity can be mostly removed. In addition, because the number of K-value can be determined correctly but not randomly generalized. The number of clusters will be suitable for any images for the segmentation. The proposed method is compared with the other clustering based image segmentation method using the evaluation indices, RMSE and PSNR values and it validated the effectiveness of the proposed method.

In the future, an optimization method to define the optimal value of the cluster radius and the hyper cluster radius for calculating spatial features can be used.

7. References:

[1] Nameirakpam Dhanachandra, Yambem Jina Chanu, "A New Approach Of Image Segmentation Method Using K-Means and Kernel Based Subtractive Algorithm", in International Journal Of Applied Engineering Research, 2017.

[2] Haoming Chen, Xue Han, "Improvement And Analysis Of Image Segmentation Techiniques", in CPMT, 2018.

[3] Dragon M. Ristic, Milan Pavlovic, Irini Reljin, "Image Segmentaion method based on self organizing maps and K-Means algorithm", in IEEE 9th Symposium on Neural Network Application in Electrical Engineering, 2008.

[4] Nameirakpam Dhanachandra, Yambem Jina Chanu, Kumanthem Manglem, " Image Segmentation Using K-Means and Subtractive Clustering Algorithm", in International Journal Of Applied Engineering Research, 2016.

[5] Sugandhi Vij, Dr. Sandeep Sharma, Chetan Marwala, "Performance Evaluation of color Image segmentation using K means clustering and Watershed Technique", in 4th IEEE ICCCNT, 2013.