A NEW APPROACH OF IMAGE SEGMENTATION USING UNSUPERVISED CLUSTERING TECHNIC

Ganesh Hosamani[1], MR. S. Sreedhar Kumar[2]
Department of Computer Network and Engineering
Don Bosco Institute of Technology, Bangalore, India

Abstract: Image segmentation is basic operation that is to be performed on the image before it is processed application specifically. Processing the entire image is not the efficient way to get the interested result. By segmenting, the image will have more meaning and only segment of the image can be processed so that the work load will also decrees and the deep investigation can be done on the image. There are several approaches available for the image segmentation but there is no universal method is available. For scientific and other purposes k-mean technique is widely accepted. In k-mean technique we have to choose number of k manually. More the number of k samples the quality of the segmentation is high. In the segmentation the quality also depends on the sample values chosen. Choosing sample value is simple if the number of k is less. For the large number of k choosing sample values from image is difficult if it is done manually. So this paper proposes a new updating to the existing k-mean technique so that for the larger number of k, samples are chosen automatically. By using this technique the iteration in k-mean can be reduced to get the quality results.

Keywords: K-Means clustering; image processing; Segmentation;

I. INTRODUCTION

Image segmentation [1] is very important process in all image processing. It is very impractical to process entire image for any application. In image only some part of the image will be containing interested data. So it is important and efficient to extract only part of the image for the further processing. There are two types of segmentation supervised and unsupervised methods. This paper mainly concentrating on unsupervised methods. There are lots of unsupervised methods are proposed but all methods will provides quality of output to particular types of images only. So there is no universal methods available for this. But for scientific and other purposes k-mean technique is widely accepted. In k-mean[2] technique we have to choose number of k manually. More the number of k samples the quality of the segmentation is high. In the segmentation the quality also depends on the sample values chosen. Choosing sample value is simple if the number of k is less. For the large number of k choosing sample values from image is difficult if it is done manually. For choosing the samples in gray scale image it has a limit. It can have maximum of 255 samples. Each pixel in grayscale have the limits of 0 to 255 that represents black to white. Those are representing the intensity[3] level. But in the color image the pixel has more values then the gray scale values that is represented with red green and blue color combination each contains values from 0 to 255 so for each pixels 16581375 possible values are there. So choosing sample is difficult. Choosing the initial vectors. Improper vectors produces inappropriate segmented results. Because of randomness of the image for the segmentation cluster initial values choosing static value is impractical. The initial values must be chosen with respect to the image only. For the unsupervised methods it is better to minimize the human involvement as possible as much. Even though the choosing of initial vector from image is simple if the value of k is less. The color is represented in digital form that comprises of color and intensity of light. Hear it is difficult to identify the unique pixel manually. So for choosing the automatic values from the given image will make the segmented results more acceptable and good quality of result.

II. K-MEAN ALGORITHEM

k-mean is unsupervised clustering algorithm that is not dedicated to particular field like image processing, this can be used in any domain in which the data has to be clustered according to some similarity. Originally k-mean is proposed for signal processing, widely it is used for the cluster analysis in data mining. For the image segmentation, unsupervised clustering k-man is one of the well suited method and for scientific purposes widely accepted. K-mean is simple to understand and to implement only have to change the distance formula depending on the data it is operating. For the image operation k-mean uses pixel distance to find the group of which pixel belongs to. Then the normal iteration is carried on to perform the segmentation process. Mathematical expression of k-mean is given below that will take x1 to xn observations as a data and k initial samples S1 to Sk.

$$\arg min \sum_{i=1}^{k} \sum_{x \in Si} ||x - \mu i||^2 \text{ ---------- } (1)$$

μi is mean of data which are similar to the previous clustered data.

Steps involved in k-mean algorithm.

Step 1: Analyze the input data and decide the no of k samples might need.

Step 2: Initialize the k samples as initial vector. Make sure that the samples are not same or almost equal.

Step3: Use the distance formula appropriate for the application. Find the group for each values. And compute new centroid for each group.

$$Ci = \arg min \, || xi - \mu i ||2 \text{---------- } (2)$$

Step 4: Compute the new centroid for each of the clusters.

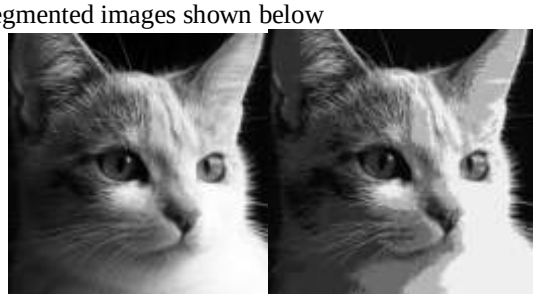
$$\mu_j = \frac{\sum_{i=1}^{m} 1\{c_{(i)}=j\}x^{(i)}}{\sum_{i=1}^{m} 1\{c_{(i)}=j\}} \text{---------- (3)}$$

Step 4: Repeat the steps until the centroid of the each group doesn't change.

Step 5: Cluster the values depending on the final centroids.

III. GRAY SCALE IMAGE SEGMENTATION USING KMEAN

Gray scale image segmentation is simple operation.While performing gray scale image segmentation, to find the difference between two pixels the simple subtraction can be used. The pixel values in the grayscale values are limited to 0 to 255 representing black to white varying the intensity of the color. This kind of segmentation will not allow to investigate as deeply as color images. The following image is segmented using size of k is 5 and initial vectors {0, 55, 110, 200, 255} the segmented images shown below



(a) Original image (b) segmented image

Fig 1 original and final segmented image.

The 5 segmented parts of the images are

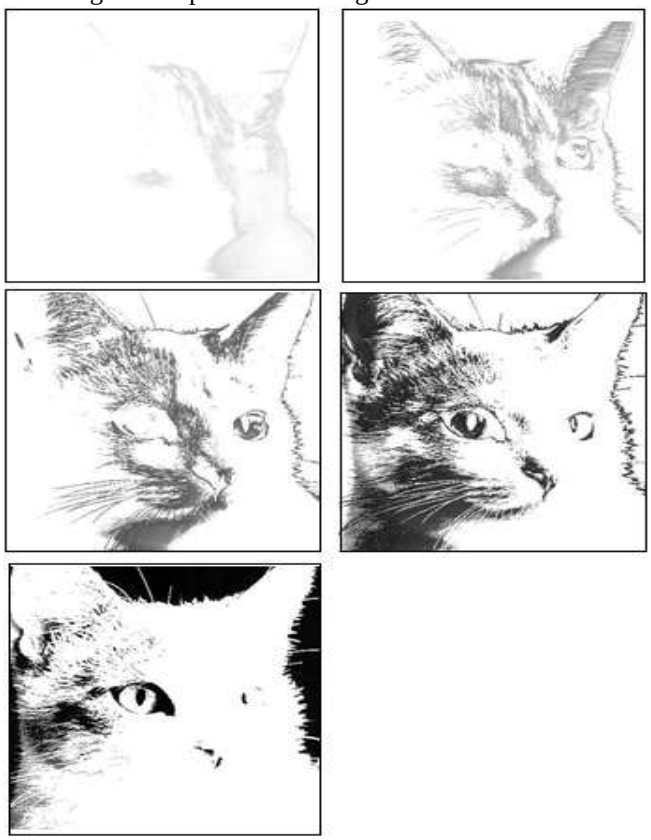
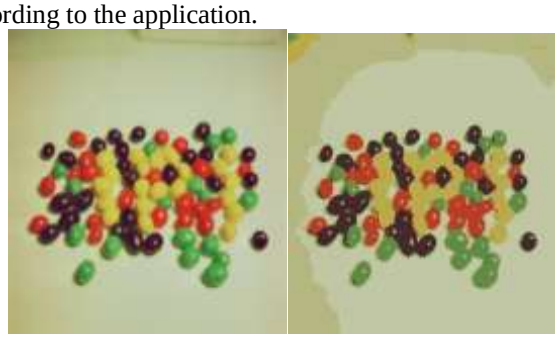


Fig 2 segmented images k=5.

IV. COLOR IMAGE SEGMENTATION USING KMEAN

The color image provides more flexibility while segmenting image. In gray scale intensity is only pixel property but for the color pixel it is mixed value of red green and blue with its level of intensity. By transforming colors in to different color spaces pixel difference can be found in different ways according to the application.

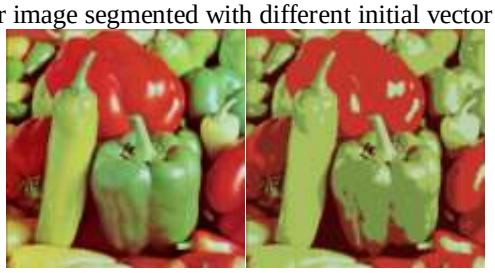


(a) Original image (b) segmented image

Fig 3 original and final segmented image using color scheme.

The above image is segmented using size of k is 6 and initial vectors {{0, 3, 11}, {183, 196, 205}, {182, 199, 206}, {176, 193, 201}, {182, 199, 206}, {180, 197, 207}}.

Another image segmented with different initial vector k=6



(a) Original image (b) segmented image

Fig 4 original and final segmented image.

Fig 5 segmented images k=5.

The above image is segmented using size of k is 6 and initial vectors {{0, 3, 11}, {183, 196, 205}, {182, 199, 206}, {176, 193, 201}, {182, 199, 206}, {180, 197, 207}}. Vectors are chosen manually by analyzing the color on image. For the small no of k it is simple to select but as the k is grater that is complicated. Human error will occur most of the times.

V. INITIAL PIXEL VECTOR SELECTION

For the k-mean technique choosing the proper initial sample values is important. The quality of the segmented images is depends on it. For most of the initial vectors if the distance between the pixels are reasonably high the k-mean will calculate new cluster mean values. Disadvantages might be the iteration taken to compute new cluster head might be more. For the deeper investigation on the image the no of k is increased. In that case the choosing of large no of sample manually is not efficient. So this proposed method will choose the values automatically from the given image. The chosen values will different for each and every image so that the quality of the image will be produced and for some low end application the iteration of the k-mean will be omitted. The proposed method will provide an updating to the k mean so that it produce quailed segmentation and more acceptable for the image processing. By adding this update the k-mean only input need is the no of k. this makes easier to use the k-mean algorithm.

Steps involved in updated k-initialization method:
Step 1: Load the image and find the unique pixels in the image.
Step 2: For each pixels find the occurrences count in the image.
Step 3: Sort them according to the occurrences count of the pixels.
Step 4: Take top 50% of the values and sort according to the pixel values.
Step 5: Choose the k values from that list make sure that the distance between the each pixels are high.
Step 6: Give those values to the k-mean algorithm.
As described in the above steps first we have to locate the unique pixels in the image, Then count the occurrences of them in the image consider "U" as unique set of pixel in image initially it is null. Consider "X" is set of pixel in image check each pixel is in "U". If the pixel present in "U" then skip the pixel. If not present in "U" add it to "U". The below equation shows in mathematical form.

$$for\ each\ xi \in X \begin{cases} if\ xi \in U & count(xi) += 1; \\ else & count(xi) = 1; \end{cases}$$

The above equation will find the unique pixels and its respective count. Now the unique pixels with its count is available. Next step is to sort the pixel array with respect to the count values. The sorted values indicates that that the top pixel in the array is occurred most frequently in the image that will be one of the most well suited initial value for the k-mean. But for choosing other values we can't directly go for the next highest occurred pixel. Some times that pixel might be very near to the above pixel. In that case choosing both pixel is not efficient way. Even though both are good samples any one can be chosen. To solve the above problem, from the unique pixels take to 50% of the pixels that will be containing the highest occurrence count. Now sort them according to the pixel values. Divide the sorted array in to |(Length_Of_Sorted_Array/k)| groups choose the pixel value from that group which has the highest count. Now the pixels are chosen with the highest count and with the acceptable distance between the pixels.

VI. EXPERIMENTAL RESULTS

For the implementation of the proposed method the system is with i5 4th gen processor 6GB ram and 2GB graphic card is used. For the programming development jdk 1.5 is used with the windows environment. And for the image jpeg format is used. The below images are segmented using the proposed algorithm for the large number of k.

Table 1: Samples chosen from proposed method.

no	k	Pixels samples
1	10	{192,199,166},{193,202,171},{192,199,165}, {192,201,170},{193,202,173},{196,201,171}, {193,200,167},{196,205,176},{192,200,163}, {197,200,169}
2	20	{192,199,166},{193,202,171},{192,199,165}, {192,201,170},{193,202,173},{196,201,171}, {193,200,167},{196,205,176},{192,200,163}, {197,200,169},{192,203,173},{194,203,174}, {195,204,175},{191,199,162},{191,197,161}, {191,200,169},{195,204,173},{195,202,168}, {194,203,172},{196,199,168},{189,195,159}
3	25	{192,199,166},{193,202,171},{192,199,165}, {192,201,170},{193,202,173},{196,201,171}, {193,200,167},{196,205,176},{192,200,163}, {197,200,169},{192,203,173},{194,203,174}, {195,204,175},{191,199,162},{191,197,161}, {191,200,169},{195,204,173},{195,202,168}, {194,203,172},{196,199,168},{189,195,159}, {192,198,162},{193,201,164},{194,202,165}, {191,198,164}
4	30	{192,199,166},{193,202,171},{192,199,165}, {192,201,170},{193,202,173},{196,201,171}, {193,200,167},{196,205,176},{192,200,163}, {197,200,169},{192,203,173},{194,203,174}, {195,204,175},{191,199,162},{191,197,161}, {191,200,169},{195,204,173},{195,202,168}, {194,203,172},{196,199,168},{189,195,159}, {192,198,162},{193,201,164},{194,202,165}, {191,198,164},{195,202,171},{193,200,159}, {193,200,166},{194,201,168},{192,198,160}

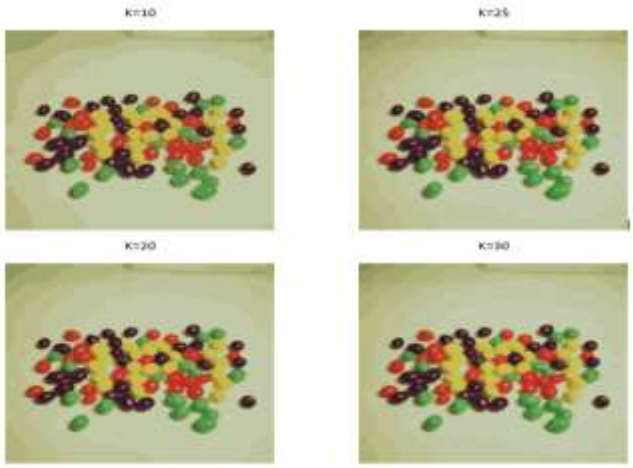


Fig 6 segmented image with different k values.

The results of the segmentation while the k is 30.

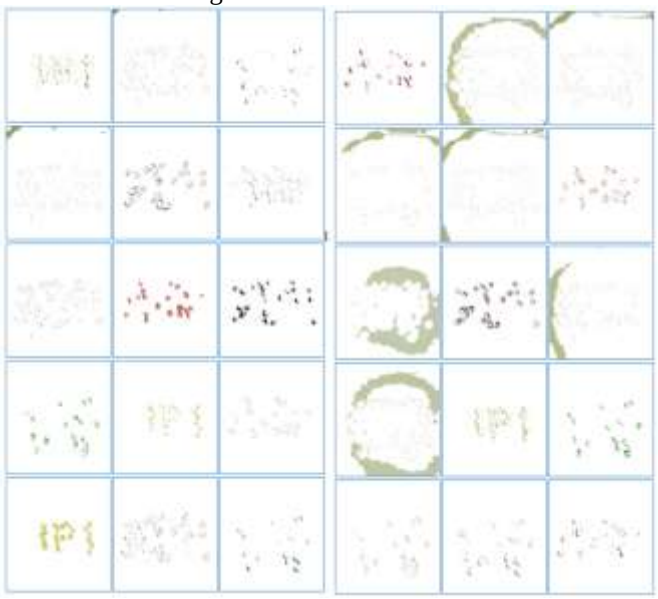


Fig 7: 30 parts of the segmented image

The segmented images enables for the deep investigation by seeing unseen patterns. And this will make k-mean more acceptable.

VII. CONCLUSION

For the selection of the k samples are much simpler and the appropriate initial values are chosen automatically. User involvement is fully minimized only choosing of k and image is dependent on the user. The initial vectors are calculated depending on the image so the quality of the initial vectors are much more acceptable for k mean. The output produced by the k mean algorithm is enabling the user to investigate deeply for his specification application. By segmenting the color image segmentation can be done with different manners like color separation, intensity so on. This method makes the k-mean method more acceptable for several applications. The quality of the output is increased and less amount of iteration are done in k-mean so the time for segmentation also less.

REFERENCES

[1] S. P. Lloyd, ―"Least squares quantization in PCM," IEEE Trans. Inf.Theory, vol. IT-28, no. 2, pp. 129–136, Mar.1982.

[2] J. Shi and J. Malik, ―Normalized cuts and image segmentation,‖ IEEE Trans. Pattern Anal.Mach. Intell., vol. 22, no. 8, pp. 888–905, Aug.2000.

[3] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy,"Sonar image segmentation using a hierarchical MRFmodel,"IEEE Trans. Image Process., vol. 9, no. 7, pp.1216–1231, Jul. 2000.

[4] Muthukannan.K, Merlin Moses.M, Color Image Segmentation using K-meansClustering and Optimal Fuzzy C-Means Clustering ,Kongu Engineering College, Perundurai, Erode, T.N.,India.27

[5] Kaiping Wei1, Tao Zhang et al., "An Improved Threshold Selection Algorithm Based on Particle Swarm Optimization for Image Segmentation" proceeding of IEEE, vol:07, 2007.

[6] Rafael C. Gonzalez; Richard E. Woods (2008). Digital Image Processing. Prentice Hall. pp. 1–3. ISBN 978-0-13-168728-8.

[7] Berlin, B. and Kay, P., Basic Color Terms: Their Universality and Evolution, Berkeley: University of California Press, 1969

[8] M.D. Fairchild, Color Appearance Models, 2nd Ed., Wiley, Chichester (2005).

