Performance Analysis of Color Image Segmentation Techniques (K-means Clustering and Probabilistic Fuzzy C-Means Clustering and Density based Clustering)

Farah Jamal Ansari
Section of Computer Engineering, University Polytechnic, Faculty of Engineering and Technology, Jamia Millia Islamia (A central University) New Delhi-25(India)

Vaseem Durrani
Technical Analyst, Aedifico Tech Pvt Ltd., New Delhi, India

Abstract—A comparative analysis of color image segmentation of three different techniques is presented in this paper with different noise levels. The robustness of each algorithm is checked on the basis of its accuracy of the clusters mapped. We have taken K-Means, FCM and Density based image segmentation approach for the analysis. In these methods of segmentation, the objects are distinguished clearly from the background. Basically Image is that type of information which has to be processed effectively and correctly. Segmentation of an image requires the separation or division of the image into regions of similar and multiple attribute. Image segmentation assigns a tag to each and every pixel in an image such that each pixel with the same tag share certain visual characteristic. Basic attribute for segmentation of an image is its luminance amplitude for an image and color components for a color image and its intensity level. Clustering is one of the methods which are used for segmentation. The objective of this paper is to compare the performance of various segmentation techniques for color images at different noise levels and attacks.

Keywords*:* K-Means clustering ,Optimal Fuzzy C-Means clustering ,Segmentation PSNR and MSE.

I. INTRODUCTION

Image segmentation is the process of dividing an image into multiple parts and that parts collectively cover the entire image. Basically, partitions are different objects in which image have the same texture or color. Because of its complicated background noise, the image segmentation is the difficult and very hot research issues on the image processing. This is typically used to identify objects or other relevant information in digital images. There are many different ways to perform image segmentation; one is Color-based Segmentation such as K-means clustering, and another is Optimal Fuzzy C- Means clustering. All of the pixels of a image in a region are similar with respect to some computed property, like color or intensity. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. Adjacent regions of a image are significantly different with respect to the same characteristics. Some of practical applications of image segmentation are: image processing, computer vision, face recognition, medical imaging, digital libraries, image and video retrieval [1]. Image segmentation methods have five categories: Region based segmentation [6], Pixel based segmentation [7], Edge based segmentation, and Clustering based segmentation [1], Edge and region Hybrid segmentation. The K-Means clustering technique is one of the most popular method that has been applied to solve low-level image segmentation tasks. In k-means clustering, it partitions a collection of data into a k number group of data [11, 12]. 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 [15]. The images are restricted by two main problems. The first problem is generated by the fact that no spatial (regional) cohesion is applied during the space partitioning process starting condition (the initialization of the initial cluster centers), while the second is generated by the Fuzzy clustering techniques have been effectively used in image processing, pattern recognition and fuzzy modeling .The OFCM was proposed by Gath and Geva and Xuejian Xiong, Kap Luk Chan. An effective method for unsupervised optimal fuzzy clustering based on a generalized objective function is implemented in this paper [12].

II. COLOR IMAGE SEGMENTATION USING K- MEANS CLUSTERING

The K-Means clustering technique is one of the most popular methods that have been applied to solve low-level image segmentation tasks. In k-means clustering, it partitions a collection of data into a k number group of data [11, 12]. 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 [15]. Clustering algorithm is convergent and its basic aim is to shape up the partitioning decisions based on a user-defined pre or initial set of clusters. Applications of the clustering algorithms of segmentation of image for complex color images are restricted because of two problems. The first problem is with generating the cluster centers and second is generating by the fact that no regional cohesion is applied during the space partitioning process. The k-means clustering was proposed by Bo Zhao, Zhongxiang Zhu, Enrong Mao and Zhenghe Song [12].

To produce erroneous decisions the selection of initial centers is very important because it prevents the algorithms from converge to local minima. The basic initialization procedure randomly selects the centers of the cluster from the input data. This procedure doesn't eliminate the problem of converging and also the segmentation results different any time when the algorithm is applied to the input data.

In this paper, by extracting the most power colors from the color histogram by using a different approaches to select the cluster centers. When this procedure is applied on the large images then this procedure proved very efficient.

During the processing, the algorithm doesn't consider the connections between the color components of each pixels nd its neighbors this is the limitation with the K-Means. Because of this reason this algorithm is restricted with some applications of the clustering to complex color-textured images.

By considering to the sample the local color smoothness, The input image is filtered by the gradient operator for the local complex texture image while for filtering the image an adaptive diffusion scheme is used. So during the space partitioning, algorithm tries to optimize the fit diffusion gradient distribution in a local neighborhood (around the pixels) under the analysis with the color for each cluster. Until the process is reached to converges, the method is continuously applied. This is applied to the clustering algorithm on large images of complex texture and on to test the data that have been corrupted due to noise.

2.1 The Algorithm

The procedure of the algorithm is easy and simple that fallows some steps to classify a set of data though some number of clusters (k clusters). For each cluster define k centroid then these centroid placed in a different way at different locations so that it can produce different results. Make these centroids far from each other to take a better result. Associate each point to the nearest centroid which belongs to a given data set, now each object assigned a nearest cluster group. Now we have to re-calculate k for new centroid from the previous steps so that binding has to be done between nearest points of centroid and same data set then a loop has been generated now we notice that k centroid changes their location and then no more changes have been seen, means now centroid do not change their location or doesn't move anymore.

$$\sum_{i=1}^{n} \zeta(ti;Cj) \qquad (1)$$

Cj is the nearest cluster centroid of ti

$$q©=1/n \quad \sum_{i=1}^{n} \zeta(ti; Cj) \qquad (2)$$

This can be computed from R as,

$$q(R, W) = \sum_{j=1}^{k} Wj \quad \sum_{i=1}^{d} Rij \qquad (3)$$

2.1.1. *Steps of the classical K-Means clustering algorithm*

Steps in the algorithm are as follows [11]:

Step1. Define the number of cluster k and randomly selects the centroid of the clusters.

$$Center_i^0 = GL_{min} + [(i-i/2)(GL_{max} - GL_{min})/k)]$$
$$i = 1, 2 \ldots k \quad (4)$$

Step2. Generate k clusters and determine the clusters Centre.

$$Distance_{i,j} = abs(GL_j - Center_i)$$

$$i = 1,2,\ldots, K;$$
$$j = 1,2,\ldots, N. \quad (5)$$

Step3. Assign each pixel in the image to the clusters that minimize the distance between the pixel and the cluster Centre (Distance is the squared or absolute difference between a pixel and a cluster Centre).

$$Center_i^m = 1/Ni \sum_{i=1}^{Ni} GLj$$

$$j=1, 2, \ldots K. \quad (6)$$

Step4. Re-compute cluster Centre by averaging all of the pixels in the cluster.

Step5. Repeat steps 2 and 3 until convergence is attained (for example cluster Centre remains unchanged).

$$Threshold = 1/2(center_k + center_{k-1}). \quad (7)$$

III. FUZZY CLUSTERING

In Fuzzy clustering means methods, a piece of data can be from one or more number of clusters. In patterns recognition this technique is used. Clustering means dividing data points of cluster into homogenous classes, so that similar data items can be taken into same class and dissimilar data can be taken into different class as possible. Data compression can be though as a clustering where large data converted into a small number of clusters. The formation of cluster depends on the data and the application, to identify the classes' different similarity-dissimilarity measures can be used. Areas of application of fuzzy cluster analysis include data analysis, pattern recognition, and image segmentation [4].

3.1The Unsupervised Optimal Fuzzy C-Means Clustering

There are two problems in Fuzzy C-Means Clustering algorithm, one is clusters needs to be specified in advance which is used, and other is limitation of FCM(restriction to spherical clusters). Also there are several derivatives of FCM one is the UFP-ONC algorithm [3], it is a two-step approach to finding a good results of a clustering. The core part of the algorithms is FCM. UFP-ONC algorithm is the improvement in the FCM.

3.2The Optimal Fuzzy Clustering

Consider a collection of patterns, x1,x2,……xnЄRp. Form a input data set X.

$$J(U,V;X) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m \{(1-g)*(\|x_k - V_i\|_l)^l + g*(\sum \{(x_k - V_i)\bullet s_{ij}\}^2)\}, l \geq 1 \quad (8)$$

c is the number of the clusters and V_i is the prototype of i^{th} cluster.

$$E_i = \sum_{K=1}^{N} (\mu_{ik}) m (x_k - V_i)(x_k - V_i)^T \quad (9)$$

These r eigenvectors,

n n

$$V_i^t = \sum_{k=1} (\mu_{ik}^{\ t-1})^m \, x_k \Big/ \sum_{k=1} (\mu_{ik}^{\ t-1})^m \qquad (10)$$

$$\mu_{ik}^t = 1/\{\sum_{j=1}^{r} \delta_{ik}^t / \delta_{jk}^t\}^{(1/(m-1))} \qquad (11)$$

Where t is the iterative step number, and

$$\delta_{ik}^t = (1-g)*(\| x_k - V_i^t \|_l)^l + g*(\sum\{(x_k - V_i^t)\bullet s_{ij}^t\}^2 \qquad (12)$$

Explanation of this equation is in [15].Obliviously, these two values are the necessary conditions for J to have a local minimum [15]. However, minimization of J with Eq. (8) forms a class of constrained nonlinear optimization problems whose solution is unknown [12,15]. This problem arises by placing the l-norm in the objective Function [15]. For the present research, only l=2 norm is considered for testing the UOFC algorithm while other norm measures will be studied in future work [15]. It can be seen that the two main advantages of the UOFC algorithm are [15]

(1) Its additional linear in the generalized objective function;

(2)The l-norm distance measure used.

3.3. The Algorithm

The algorithm consists of following parts:

(1) Choose initial cluster centroids Vi .

(2) calculate the centers vectors

$$u\_\{ij\} = \frac{(1/d^{\wedge}2(Xj, Vi))^{\wedge}(1/(q-1))}{\text{Sum}(k=1, K)\,(1/d^{\wedge}2(Xj, Vk))^{\wedge}(1/(q-1))} \qquad (13)$$

(3) Compute new centroids Vi_new:

$$Vi\_new = \frac{\text{Sum}(j=1, N)\,(u\_\{ij\})^{\wedge}q\, Xj}{\text{Sum}(j=1, N)\,(u\_\{ij\})^{\wedge}q} \qquad (14)$$

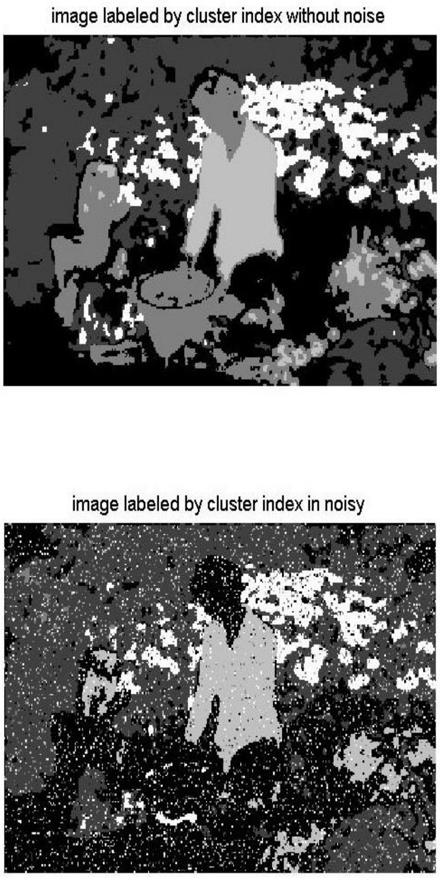
(4) When the movement of centroids (relative changes) is less than a predetermined threshold MOVETHRS, stop the iteration. Otherwise go to step 2.

IV. EXPERIMENTAL RESULTS

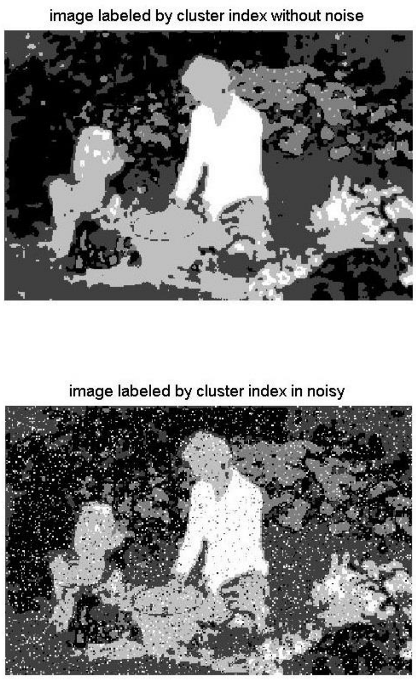
We have considered three clustering algorithms for the complete analysis. these algorithms are kmeans clustering, fcm clustering and density based clustering. Clustering of color in one standard image is done for all the three algorithms on MATLAB software tool. Fig 1 displays the standard image.

Fig 1 original image

Figure 2(a) displays the clustering image after kmeans algorithm and figure 2(b) displays the clustering image after fcm.



(a)



(b)

Fig2 (a) kmeans result (b) fcm result

In figure 3 (a), (b) and (c) color based separate clusters are shown for kmeans, fcm and density based respectively

Fig 3(a)

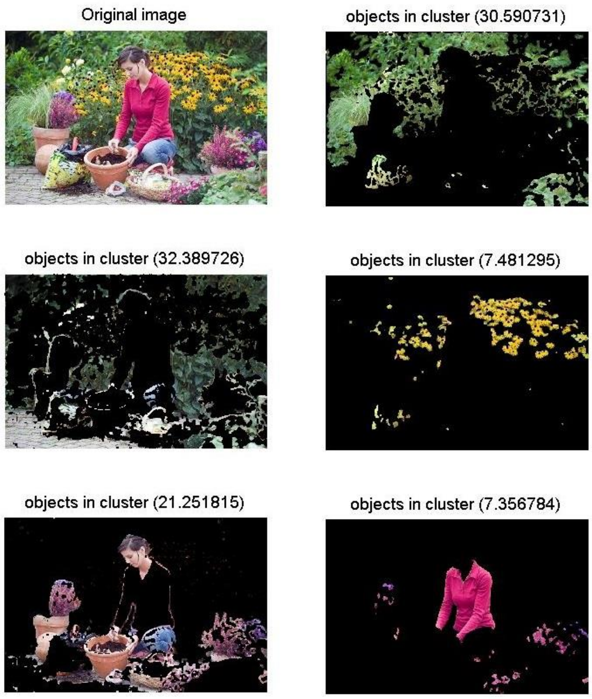


Fig 3(b)

Fig 3(c)

Fig 3 (a) kmeans color clusters (b) fcm color clusters (c) db color clusters

Noise Salt & Pepper (0.05 level)	
Algorithm	**Accuracy (%)**
Kmeans	80.24
Fcm	90.15
Density based	78.45

Table 1

The accuracy of fcm is better than the density based and kmeans clustering algorithms as shown in the table 1.

Noise Salt & Pepper (0.15 level)	
Algorithm	**Accuracy (%)**
Kmeans	71.02
Fcm	88.45
Density based	71.21

Table 2

Noise Salt & Pepper (0..25 level)	
Algorithm	**Accuracy (%)**
Kmeans	59.12
Fcm	78.23
Density based	61.12

Table 3

Table 1 2 and 3 are displaying the algorithms performance analysis at different salt and pepper noise. At 0.05 level where noise is almost negligible all the three are above 75 % accuracy level but as the noise level increases the performance of the algorithm degrades. And at 0.25 level of noise the kmeans accuracy goes down to 59% whereas FCM still shows 78.23%. hence fcm performs better in context with the noisy environment it founds the cluster above 75% correct.

V. CONCLUSION

After performing the experiments on all the three algorithm's for image segmentation and clustering. we have concluded that the fcm clustering algorithm is quite accurate in-terms of detecting the clusters as compared to density based and kmeans algorithm at different level of salt and pepper noise.

REFERENCES

[1] J.Bezdek, A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithm, IEEE Trans.Pattern. Anal. Mach. Intel, 2:1-8, 1980.
[2] GAO Xin-bo.Fuzzy cluster analysis and its applications [M].Xi'an, China: House ofXidian University, 2004. (in Chinese)
[3] Ahmed M N, Yamany S M, Mohamed N, etal. A modified fuzzy Cmeans algorithm for bias field estimation and segmentation of MRIdata [J].IEEE Trans on MedicalImaging, 2002, 21(3):193-199.
[4] DINGZhen, HU Zhong-shan, YANG Jing-yu, TANG Zhen-min, WU Yong-ge.Afuzzy-clustering-based approach to image 1997, 34(7):536-541. (in Chinese)
[5] F.Meyer, Color image segmentation, In IEEE International Conference on Image Processing and its Applications, pages 303–306, May 1995.Maastricht, The Netherlands.
[6] A.Moghaddamzadeh and N. Bourbakis, A Fuzzy Region Growing Approach for Segmentation of Color Images,Pattern Recognition, 30(6):867- 881, 1997.
[7] G.A. Ruz, P.A.Estevez, and C.A. Perez, A Neurofuzzy Color Image Segmentation Method for Wood Surface Defect Detection, Forest Prod. J. 55 (4), 52-58, 2005.
[8] A.A. Younes, I. Truck, and H. Akdaj, Color Image Profiling Using Fuzzy Sets, Turk J Elec Engin, vol.13, no.3, 2005.
[9] Refael C.Gonzalez and Richard E.Woods, Digital Image processing (second edition) Pearson Education Asia Limited and Publishing House of Electronics Industyt, 2007.
[10] Pham DL and Prince JL, An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities, Pattern Recognit Lett, 1999, 20(1):57—68. [11] B. Zhang,"Generalized K-harmonic Means–boosting in Unsupervised Learning, Technical report (HPL-2000–137), Hewlett-Packard Labs, 2000.
[11] ZhongxiangZhu ,Bo Zhao ,Enrong Mao and Zhenghe Song, Image segmentation based on Ant colony optimization and K-means clustering, Proceedings of the IEEE international conference on Automation and logistics, August 18-21,2007,jinan,china.
[12] Gath,I. and Geva A. B. (1989), Unsupervised Optimal fuzzyclustering,IEEE Transaction on Pattern Analysis Machine Intelligence, 11(7): 773-781
[13] Monika Xess and S.Akila Agnes", Survey On Clustering Based Color Image Segmentation And Novel Approaches To Fcm Algorithm", IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308
[14] Muthukannan.K, Merlin Moses.M ", Color Image Segmentation using K-means Clustering and Optimal Fuzzy C-Means Clustering", International Conference on Communication and Computational Intelligence – 2010,