Color based segmentation using clustering techniques

[1] Deepali Jain, [2] Shivangi Chaudhary
[1]Communication Engineering,
[1]Galgotias University, Greater Noida, India

***Abstract* - Segmentation of an image defines as a process of partitioning the image into its constituent parts or similar objects. There are two approaches for image segmentation i.e. discontinuity based and similarity based. In this paper, color based segmentation on noisy image using K-means and Fuzzy c-means is shown. Images are the best means of conveying information. Image is in RGB color space, transforming it in LAB color space which is more compatible to human vision. Clustering is one of the best methods for segmentation even if the data is large. The main aim of this paper is to compare the performance of K-means and Fuzzy c-means techniques over the noisy image. The best results can be seen after finding the accuracy of the segmented images using clustering techniques.**

***Index Terms* - Image segmentation, k-means clustering, Fuzzy c-means clustering**

I. INTRODUCTION

Image segmentation is a salient part of image processing. Segmentation refers to a process in which a digital image is separated into uniform and non-overlapping regions. Each region should be homogeneous means based on some similarity measures [1]. Segmentation has wide applications in the field of biomedical, pattern recognition, image coding etc. Image segmentation is a basic and key step for better analyzing an image. Noisy background or adhere noise in the image made the image segmentation process a big and challenging research issue. The main objective is to clarify the image which is easy to analyze. The main focus is to segregate the digital image into set of pixels where pixels should be similar in color, intensity, texture etc. There are many approaches for segmentation like region based segmentation, edge based segmentation, thresholding and clustering. Clustering is one of the best and simplest techniques for differentiating an image into similar attributes.

In literature review we come to know about the image segmentation via soft computing method, this approach helps in upgrading the clarity and the recognition capabilities of grey level image [2]. Some other important clustering techniques were proposed earlier like shape based clustering where HMM is used for finding the number of clusters and then data is clustered according to their shape similarity [3]. Large data set is a challenging task for segmentation but via clustering the execution time can be minimized [4].Fuzzy ISODATA algorithm is known for the class of convergence of class of clustering. In this paper arbitrary sequence created by Picard iterations constantly terminate at local minima, this is proved by the help of Zangwill theory [5].

II. CLUSTERING

Clustering is a process of collecting objects into groups of similar attributes and separate it with the dissimilar ones. Cluster analysis is an essential part of clustering which help us in discriminating the similar object in one cluster and dissimilar ones in another cluster [6]. Clustering methods tells us about the interconnection between the data. In recent scenario, clustering is kind of boom, it has significance in many real world applications such as marketing, land use, urban planning, bank and internet security, spam pattern discovery and in image processing etc. Clustering techniques assist in dealing with the noisy data, dealing with different types of attributes, clusters of arbitrary shapes and to deal with high-dimensional data [7]. It has various future applications like in biomedical field; if a patient is suffering from a disease then doctors can predict what will be the impact of this disease after few years with the help of clustering.

Clustering Methods

Many clustering algorithms have been proposed till now in the literature. There are two types of clustering algorithms one is hard (or crisp) and other is fuzzy. Hard means each data point of the data set belongs to one and only one cluster and fuzzy means one data point of the data set can belong to more than one cluster [6]. Major clustering methods can be classified as partitioning methods, Hierarchical methods and Density based clustering [7]. The main objective is to choose correct algorithm for particular data type and this depends on the data type available, the nature and motive of the application.

III. PROPOSED ALGORITHM

In the present work, the main focus is on the two clustering techniques i.e. K-means and Fuzzy c-means clustering. Starting with the K-means clustering, it is widely used and simple algorithm. For efficient high dimensional data set, K-means clustering is used. K-means comes under the category of center based clustering algorithm. K-means is a partitioned and non-hierarchal clustering method [8]. K-means is based on iteration method which helps in partitioning the n datasets into k clusters where the initialization of clusters is fixed from the starting. The main disadvantage of K-means is, it works only on numerical data which restricts its domain of working. Numerical data means each cluster has its center known as mean.

K-means comprise of two phases one is initialization phase means randomly the data is allocated into k clusters by the algorithm and second is iteration phase in which the distance between the data points and cluster centre is calculated by the algorithm and assigns the nearest cluster to the data point on this basis of distance. Only on homogenous non-textured color image, K-means gives some realizable results. Advantages of K-means are easy implementation, converge quickly and give good results but conditions applied [9][10]

Algorithm:

1. Randomly assign the cluster centers.
2. Compute the Euclidean distance from the data points and the cluster centre.
3. Assign each object to its nearest cluster temporarily.
4. Again compute the mean of the clusters and reassign the object into cluster.
5. This process repeats until there is no movement or no changes in the mean.
6. Otherwise go to step 2.

Now, coming on to Fuzzy c-means clustering, it is a soft computing method. It is the supplement of K-means because fuzzy allot membership function to the data points. Membership function means each data point is allowed to be in more than one cluster. Fuzzy c-means is an extension of the fuzzy set theory. Membership can be assigned by calculating the distance between the cluster center and the data points, the data point having less distance with the cluster centre comes in that particular cluster [11]. Number of clusters, the fuzziness exponent and the termination tolerance are three main parameters of Fuzzy c-means clustering.

Algorithm:

1. Randomly assign the cluster centers.
2. Compute the Euclidean distance from the data point to each cluster center.
3. Check the distance from data points to the centers, the data point having minimum distance with the
4. Cluster center has high membership and assigns it with that particular cluster.
5. Evaluate new cluster centers.
6. This process repeat until there is no movement in the centers.
7. Otherwise go to step 2.

IV. METHODOLOGY

Considering a standard image in .jpeg format, first we convert it into .png format because it compresses image properly without any loss and supports millions of colors. Further this RGB image is converted in LAB color space where we define the luminosity and chromaticity layer. After this we make the cluster index so that whole image will be indexed between black, grey and white colors. Now we calculate the pixels of the indexed image then we apply K-means clustering technique. This technique segments the image on the basis of their colors and frame different clusters according to our initialization of clusters. Now, we add salt and pepper noise in the image and then find the cluster index of the noisy image and calculate the pixels of it. Now, for accuracy we calculate the difference of the noisy cluster index pixels and noiseless cluster index pixels and divide it from the total pixels of the original image. The two main components of fuzzy logic system are fuzzy set and membership functions. Fuzzy set includes the infinite range of true values between 0 and 1. It defines only partial degree of membership. Now, membership function assigns the real values to the data points of the object as shown in figure1 and figure2. The whole procedure is carried out in Fuzzy Inference System. In fuzzy inference system we initialize the value of three parameters that are input, output and rule base. The advantage of using Sugeno rule base is we can represent output in linguistic pattern as shown in figure4 where the output is either linear or constant as shown in figure3. The same procedure is applied with Fuzzy c-means technique [12] and then we are able to compares the accuracy of both the techniques.

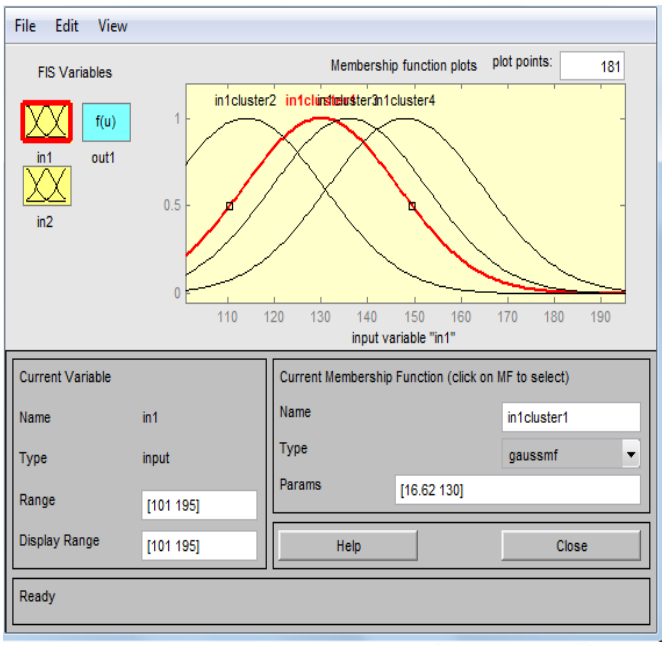

Figure1: Membership function of input variable 1

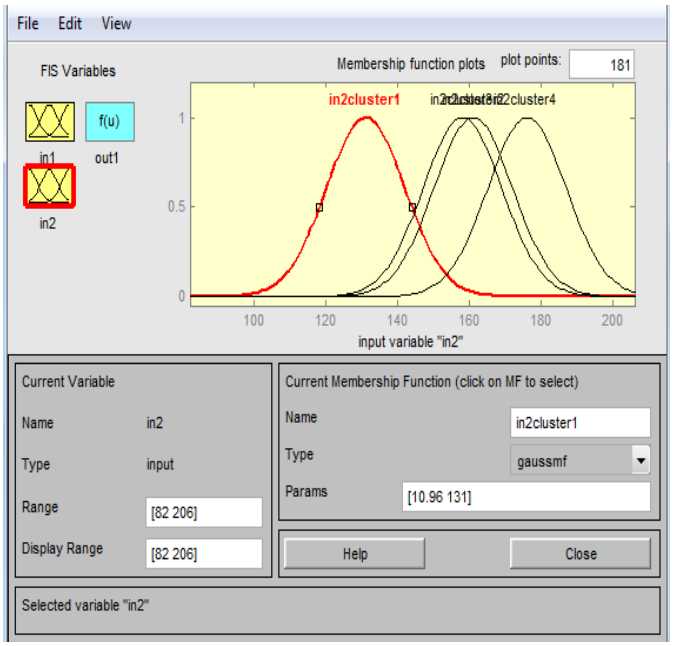

Figure2: Membership function of input variable 2

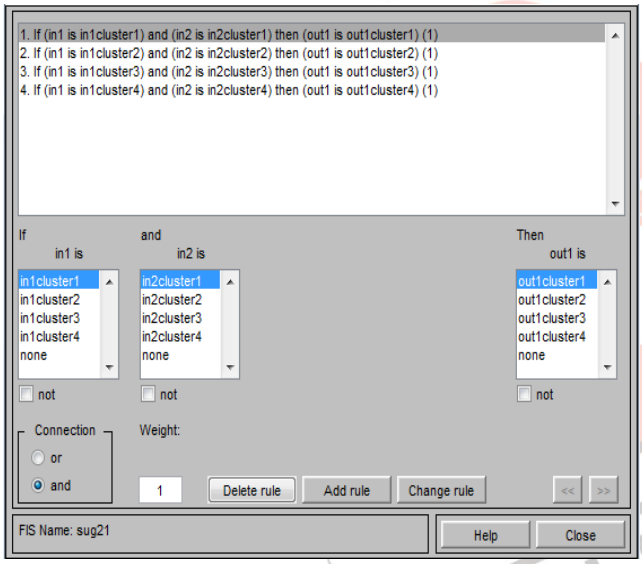

Figure3: Sugeno rule base

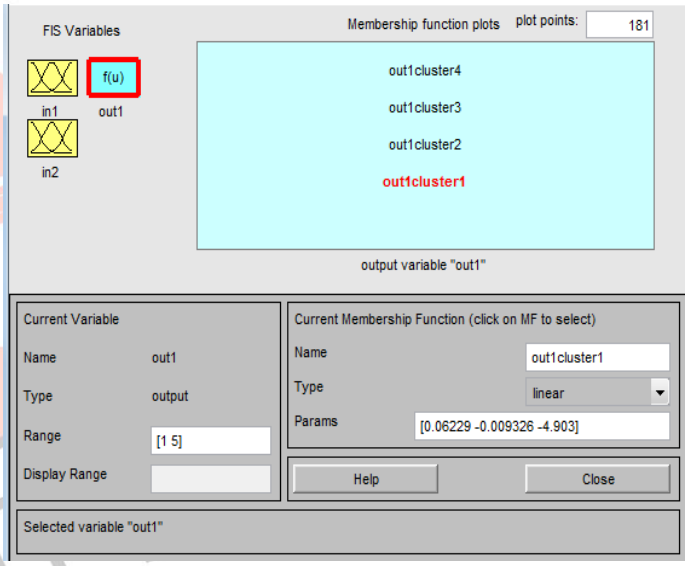

Figure4: Output

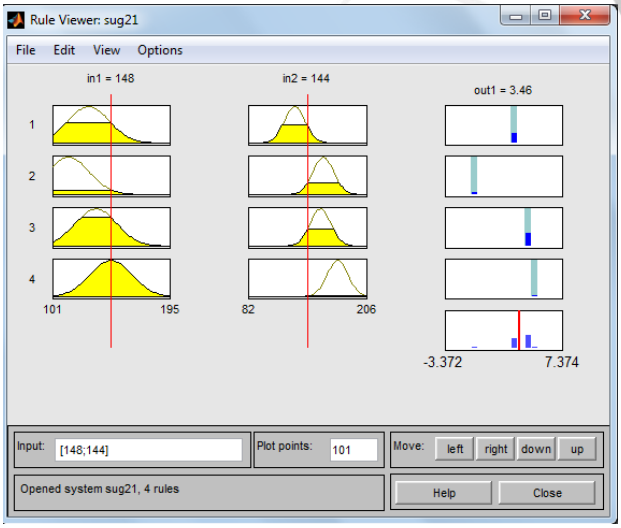

Figure5: Rule viewer

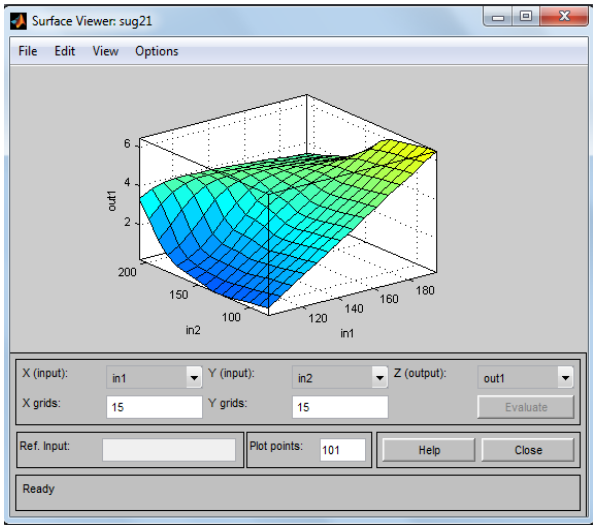

Figure6: Surface viewer

V. EXPERIMENTAL RESULTS

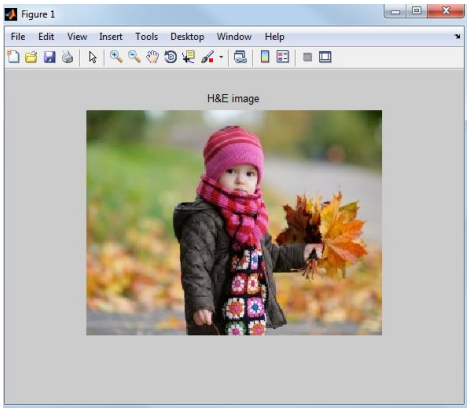

Figure7: Original image

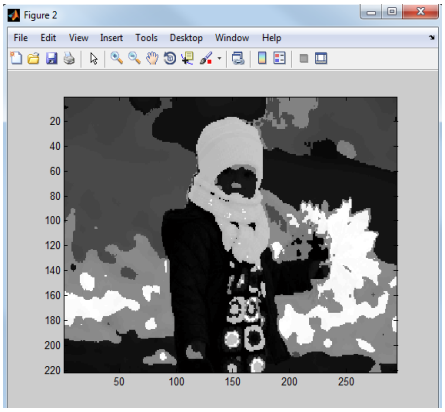

Figure8: Cluster index of Original image

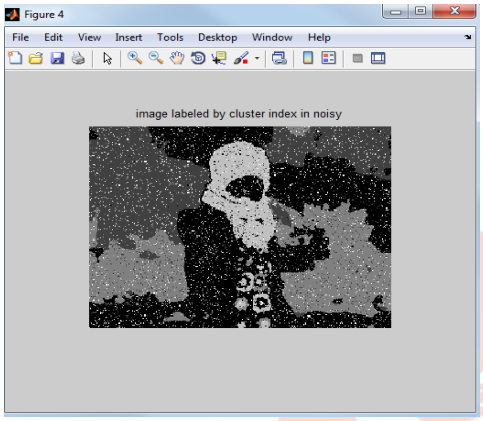

Figure 9: Cluster index of noisy image

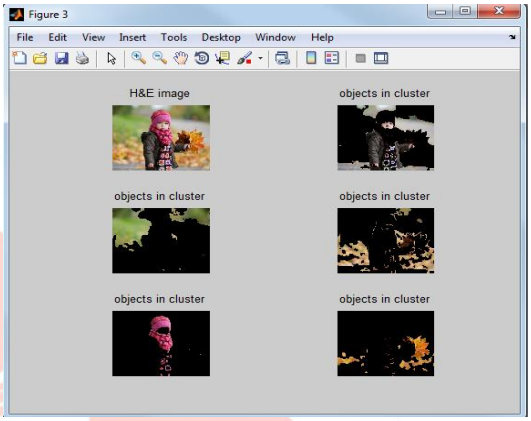

Figure 10: Color based segmentation using k-means Clustering

Cluster index means it segregate the whole image in the pixel labels where all the pixels labeled in the form of numbers means which pixels fall in which cluster. Hence, the mapping is done between the noisy and noiseless image, the technique which gives the more corrected segments considered as the best technique.

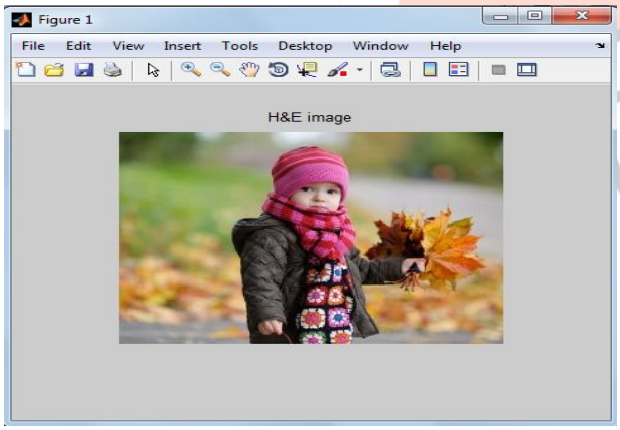

Figure11: Original image

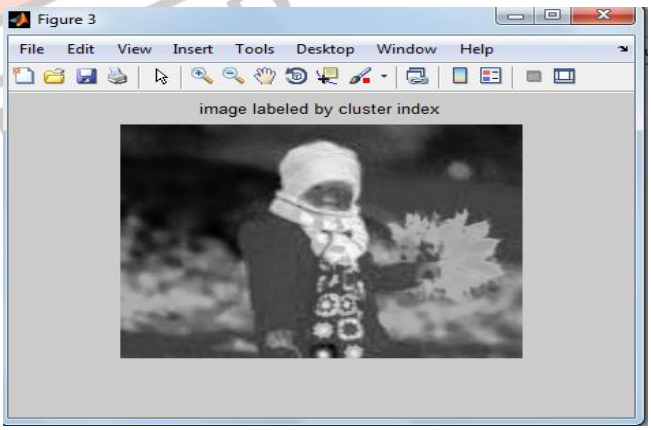

Figure12: Cluster index of original image

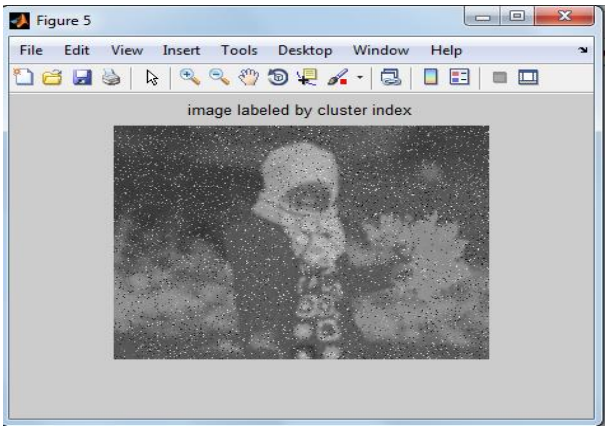

Figure13: Cluster index of noisy image

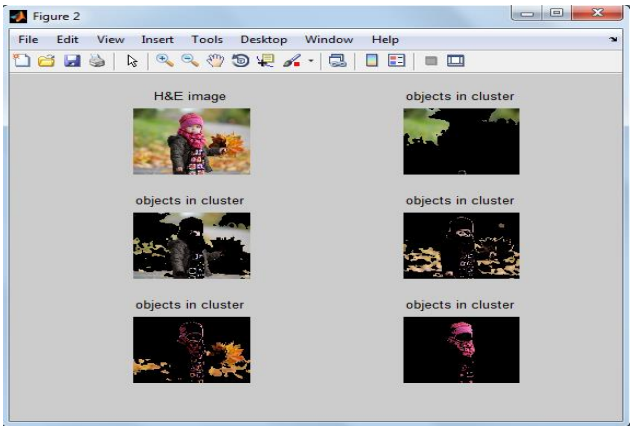

Figure14: Color based segmentation using Fuzzy c--means Clustering

After doing the whole clustering procedure, the accuracy we get from the Fuzzy c-means is 88.92% and that we get from K-means is 75.82%. This shows fuzzy c-means gives better results and accuracy than K-means in case of color based segmentation as shown in TableI.

Table I Clustering techniques results

Clustering Techniques	Accuracy
K-means	75.82%
Fuzzy C-means	88.92%

VI. Conclusion

Fuzzy c-means is superior to K-means because it gives good results on whether the data is incomplete, overlapped or uncertain and has wider applicability. K-means-means belongs to hard clustering and Fuzzy c-means is a soft computing method. K-means is not efficient because the mean of the cluster changes on every iteration. Fuzzy c-means is better because it can properly work on high dimensional data than K-means. No doubt, Fuzzy c-means is slower and takes times to process but gives better and improved efficiency than K-means. In this paper, the effect of noise on the image using different clustering techniques is measured. Fuzzy c-means technique gives more segmented image and more appropriate accuracy than K-means.

References

[1] F.Meyer. (1995). Color image segmentation. IEEE International Conference, (pp. 303-306).
[2] Senthilkumaran, N., & Rajesh, R. (2009). Image segmentation -A Survey of Soft computing Approaches. International Conference, (pp. 844-846).
[3] Bhardwaj, S., Srivastava, S., & Gupta, J. (2013). A Novel Clustering Approach Using Shape Based Similarity. Springer, (pp. 17-27).
[4] Saraswathi, S., &Allirani, A. (2013). Survey on image segmentation via clustering. International Conference, (pp. 331-335).
[5] J.Bezdek (1980) A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithm. IEEE Journal, (pp. 1-8).
[6] Xu, R., &Wunsch, D.(2005). Survey of Clustering Algorithms," Neural Networks IEEE Transaction, (pp. 645-678
[7] Lukic, I., Kohler, M., &Slavek, N. (2012). The segmentation of Data set Area method in the clustering of uncertain data International Convention, (pp. 405-410).
[8] Dharmarajan, A., &Velmurugan, T. (2013). Applications of partition based clustering algorithms. IEEE International conference, (pp. 1-5, 26-28).
[9] K, M., & M.M., M. (2010). Color image segmentation using k-means clustering and Optimal Fuzzy C-Means clustering. International Conference (pp. 229-234).
[10] Kalam, R., &Manikandan, K. (2011). Enhancing K-means Algorithm for image segmentation. International Conference, (pp. 1-4, 20-22).
[11] Ningbo, G. (2004). Fuzzy cluster analysis and applications. China: House of Indian University).
[12] DING Zhen, Zhong-shan, H., Jing-yu, Y., Zhen-min, T., & Yong-ge, W. (1997). A Fuzzy clustering based approach to image. IEEE, (pp. 536-541).
[13] Bourbakis, A. a. (1997). A Fuzzy Region Growing Approach for Segmentation of Color Images. IEEE, (pp. 867-881).