Comparison of Clustering Methods for Segmenting Color Images

S. Arumugadevi[1*] and V. Seenivasagam[2]

[1]Department of Information Technology, Sri Krishna Engineering College, Chennai, India; sa_devi@yahoo.co.in
[2]Department of Computer Science and Engg, National Engineering College (Autonomous), Kovilpatti, India;
yespee1094@yahoo.com

Abstract

Background/Objectives: Image segmentation is the first step for any image processing based applications. The Conventional methods are unable to produce good segmentation results for color images. **Methods/Statistical analysis:** We present two soft computing approaches namely Fuzzy C-Means (FCM) clustering and Self Organizing Map (SOM) network are used to segment the color images. The segmentation results of FCM and SOM compared to the results of K-Means clustering. **Results/Findings**: Our experimental results shown that the Fuzzy C-Means and SOM produced the better results than K-means for segmenting complex color images. The time required for the training of SOM is higher. **Conclusion/Application**: The trained SOM network reduced the execution time for segmenting color images. The performance of FCM and SOM is higher than the K-means for segmenting color images. Applications of color image segmentation are video surveillance, face recognition, fingerprint recognition, object detection, medical image analysis, and Automatic target detection.

Keywords: Clustering, FCM, Image Segmentation, K-Means, SOM, Subtractive Clustering

1. Introduction

Color image segmentation is the challenging task in image processing and contains two critical issues, firstly which color model to be used and secondly, which segmentation technique should be applied. A color space is a method by which we can specify, create and visualize colors. Several color representations, such as RGB, HSI, CMY, CMYK, YIQ, CIE L*a*b*, etc., are employed for color segmentation, but none of them can dominate the others for all kinds of colors images[1]. Each color representation has its advantages and disadvantages. RGB model is device dependent. A CIE L*a*b* color space is a color opponent space with dimension L for lightness and 'a' and 'b' for the color-opponent dimensions, based on nonlinearly compressed CIE XYZ color space coordinates[2].

The advantages of CIE L*a*b* color space are: It is designed to approximate human vision; The CIE L*a*b* color space includes all perceivable colors; It is device independent; It is used in many industries apart from printing and photography; It provides exact color specifications for paint.

The segmentation of color image has been proved to be difficult because it involves a vast amount of data processing. In computer vision literature, various methods dealing with segmentation have been discussed. Although great efforts have been devoted to it, some issues are still not fully addressed. Shirakawa and Tomoharu[3] proposed evolutionary image segmentation based on multiobjective clustering. Two objectives, overall deviation and edge value, are optimized simultaneously using a multiobjective evolutionary algorithm. Yang and Huang[4], have modified the objective function of the standard FCM algorithm with a penalty term that takes into account the influence of the neighboring pixels on the centre pixels for image segmentation. In [5] fuzzy clustering algorithms and competitive neural network was used for color image segmentation. The self estimation algorithm was suggested for automatically finding the number of clusters using Euclidean distance. The features extracted using

Discrete Wavelet Transform (DWT) fed to fuzzy C-means algorithm and the membership function created by FCM was used as a target to be fed to the back-propagation neural network[6].

Deshmukh and Shinde proposed the neuro-fuzzy system ACISFMC. It uses a Multilayer Perceptron (MLP) like network which performs color image segmentation using multilevel thresholding. Threshold values used for finding clusters and their labels are found automatically using FMMN clustering technique[7]. FCM algorithm was used to define the target of the supervised feed-forward neural network and a fuzzy entropy method was deployed to set a threshold value for improving the segmented image[8]. Dong et al.[9] proposed a hybrid system which comprises unsupervised segmentation and supervised segmentation. The unsupervised segmentation is achieved by a two-level approach, i.e., color reduction and color clustering. The supervised segmentation involves color learning and pixel classification. Simulated Annealing (SA) has been used for finding the optimal clusters form SOM prototypes.

The color image segmentation results based on soft computing techniques are proven to be better than conventional hard clustering techniques[10]. FCM[11,12] used for color image segmentation. A.Borji et al.[13] proposed new method for color image segmentation using fuzzy logic. Wen-Xiong Kang et al. is compared the various Image Segmentation Algorithms[14]. Some researchers modified and expanded the typical SOM[15,16]. Most of the researchers modified and expanded the FCM[17–19]. N. Senthilkumaran and R. Rajesh, explained the various edge detection techniques using soft computing approaches[20].

Dongxiang Chi suggested that SOM with K-Means. SOM-K, a new unsupervised natural image segmentation method based on SOM and k-means and also stated that clustering method like k-means and their variants are not acceptable for considering the computational cost and a priori cluster number k needed[21]. The various clustering methods are applied on the time series databases. When compare the clustering methods, the SOM performs well in forming clusters[22]. Rajiv Kumar and A. M. Arthanariee[23] proposed hybrid method that combines region-based and cluster-based. The time consumption of FCM is less when compare with the K-Means clustering method[24]. P. Ganesan et al. proposed fuzzy based segmentation method in YCbCr Color space. RGB color space is not efficient for object specification and recognition of colors[25]. A hybrid method[26], which is based on

split and merge approach, is proposed for detecting the fruit defects. The k-means algorithm is used to split the original image into regions and merging by minimum spanning tree procedure.

2. Fuzzy C-Means Clustering

FCM Clustering algorithm is frequently used in pattern recognition. FCM assigns membership to each data point on the basis of distance between the cluster and the data point. The algorithm is based on minimization of the objective function (1).

$$\sum_{i=1}^{N} \sum_{j=1}^{C} U_{ij}^{m} \left\| x_i - c_j \right\|^2 \qquad (1)$$

Where N is the number of data, c is the number of clusters, uij is the degree of membership of xi in the cluster j.

There are four main steps in this approach for segmenting the color images:

Step1: Image acquisition
Step2: Preprocessing: To convert an image in RGB into CIE L*a*b*
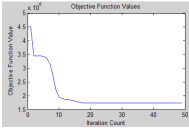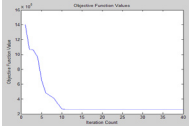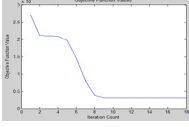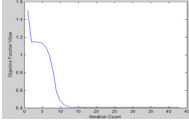Step3: Clustering by using Fuzzy clustering method.
Step4: Segmentation process: Assign label to highest membership value in each column of the partition matrix U that indicates a data point belongs to which cluster center. Find the segmented images from the labeled matrix.
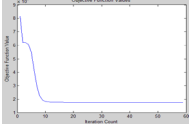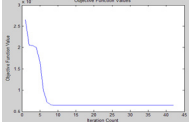
The FCM clustering process has been carried out with variety of images. The segmentation results are shown in Figure 3 and Table 1.

3. Self-organizing Map (SOM)

A Self-organizing Map (SOM) network is a type of artificial neural network that is trained using unsupervised learning. The SOM[21] map consists of a competitive layer which can classify a dataset of vectors with any number of dimensions into as many classes as the layer has neurons. The neurons are arranged in a 2D topology, which allows the layer to form a representation of the distribution and a two-dimensional approximation of the topology of the dataset. The SOM learn to cluster data based on similarity, topology. The goal of learning in the self-organizing map is to cause different parts of the network to respond similarly to certain input pattern.

Table 1. Experimental results of FCM on L*a*b* color space

Image	Size	No. of cluster	Running Time in seconds	No. of iterations	Objective function minimization status
	225x225	4	2.23272	49	
	148x111	4	0.68907	40	
	150x120	4	0.79848	48	
	225x225	3	1.542816	42	
	259x194	3	0.64425	18	
	225x225	4	2.41344	43	

SOM Training algorithm

Step1: *Create a Self Organizing map in which neurons are arranged by hexagonal.*

Step2: *Distances between neurons are calculated from their positions.*

Step3: *Preprocessing: Perform Histogram equalization for the intensities can be better distributed on the histogram.*

Step4: *Create a input vector from an pre-processed image.*

Step5: *Pick each input vector in the input data set and do the following*
  a. *Find the similarity between the input vector and the map's node's weight vector using distance.*
  b. *Track the node that produces the smallest distance is called best matching unit, BMU.*
  c. *Update the nodes in the neighborhood of the BMU.*

4. Determining the Number of Clusters and Cluster Validity

If the number of clusters is manually specified, the segmentation may not be effective[5]. Two main issues in clustering are determining the number of clusters and measure to evaluate the quality of the clusters.

4.1 Finding Number of Cluster

In this paper we have used an algorithm based on co-occurrence matrix for finding the number of clusters.

Algorithm: Co-occurrence matrix based method

Step1 Read an image I

Step2 Transform image I to HSV color space.

Step3 Calculate the co-occurrence matrix T for H values.

Step4 Pick the diagonal values of co-occurrence matrix and store them in d.

Step5 Find the local maximum from the diagonal values by i) finding the mean of the diagonal values of co-occurrence matrix, ii) defining K as the number of values that are greater than or equal to the mean value.

Step6 K is the final number of clusters.

Subtractive clustering is a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a set of data. When Compared to K-means and FCM, this result is a little bit behind the accuracy achieved in those other techniques. But we can use this method for initializing the cluster center.

4.2 Method of Cluster Validation

The hardest problem in comparing different clustering algorithms is to find an independent measure to evaluate the quality of the clusters. In our paper we have used the silhouette index value for evaluating the cluster compactness. The silhouette value for each pixel is a measure of how similar that pixel is to pixels in its own cluster vs. pixels in other clusters, and it ranges from –1 to +1. For pixel i, it is defined as

$$s(i) = \frac{bi - ai}{\max(ai, bi)} \qquad (2)$$

Where 'ai' is the average distance from the ith pixel to the other pixels in the same cluster, and 'bi' is the minimum average distance from the ith pixel to the pixels in a different cluster, minimized over all different clusters. It is clear

from the above definition $-1 <= s(i) <= 1$. The silhouette value is calculated using (1) for each pixel of the clustered image. For fixing the correct number of clusters the following method is used in our paper.
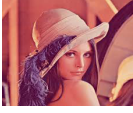
Algorithm for Finding Cluster Validity

Step 1 *Read an image I.*
Step 2 Preprocessing.
Step 3 *Initialize the number of clusters K = 2 and set the previous silhouette value = 0.*
Step 4 *While k>=2 do the following*
Step 5 *Cluster the image using any algorithm.*
Step 6 *Find the silhouette index value for the image by silhouette = mean (silhouette values of all pixels in image I).*
Step 7 *If silhouette>previous silhouette value*
   k = k + 1 / increase the number of clusters*
    Previous silhouette value = silhouette
    goto step4.
   Else return k–1 / K -number of clusters.*
   Step 8 Stop.

5. Result and Discussion
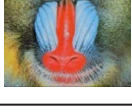
Table 2. Cluster count found by co-occurrence matrix based method and Subtractive clustering

SNo.	Image	No. of clusters found by Co-occurrence matrix based method	No. of clusters found by Subtractive *clustering*
1	Test Image 1	4	3
2	Test Image 2	4	3
3	Test Image 3	5	2
4	Test Image 4	2	2
5	Test Image 5	3	3
6	Test Image 6	3	4
7	Test Image 7	5	4
8	Test Image 8	4	3
9	Test Image 9	3	5
10	Test Image 10	4	4
11	Test Image 11	3	3
12	Test Image 12	2	3
13	Test Image 13	3	2
14	Test Image 14	3	3

(*Contiued*)

Table 2. *Continued*

SNo.	Image	No. of clusters found by Co-occurrence matrix based method	No. of clusters found by Subtractive *clustering*
15	Test Image 15	3	4
16	Test Image 16	4	3
17	Test Image 17	3	3
18	Test Image 18	3	3
19	Test Image 19	3	3
20	Test Image 20	3	3
21	Test Image 21	3	3
22	Test Image 22	4	4
23	Test Image 23	3	4

All the algorithms are executed and tested using MatlabR2013a with a variety of images. For a sample, the results of few images are discussed. FCM is applied to the L*a*b* color space transformed image for improving the segmentation results which are shown in Figure 3. The advantage of FCM is that unlike K-means clustering a pixel may belong to more than one cluster. The main drawbacks of FCM are computational time and more sensitive to initialization conditions of cluster number and cluster center. Different number of clusters is also chosen for segmenting the various test images. In many clustering algorithms such as K-means and FCM need number of clusters as input. Two Methods co-occurrence based and subtractive clustering for finding number of clusters of test images and the results are shown in Table 2 and the comparison between the methods have been shown in Figure 1. Both methods are produced same results for 10 test images out of 23 images. From the silhouette index for the clustering images after applying K-means, FCM and SOM with various number of cluster counts, the results shown that co-occurrence based method is better than subtractive clustering method to determine the number of clusters needed for images. Most of the images have higher silhouette index for k = 2. But k = 2 is not sufficient for many complex color images. In such cases start from k = 3. So these methods act as prior step for clustering algorithms. The execution time of subtractive clustering is higher than the co-occurrence based method. The execution time is lowest for K-means and highest for FCM and SOM. The execution time is very less for the trained SOM used for segmentation.
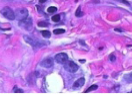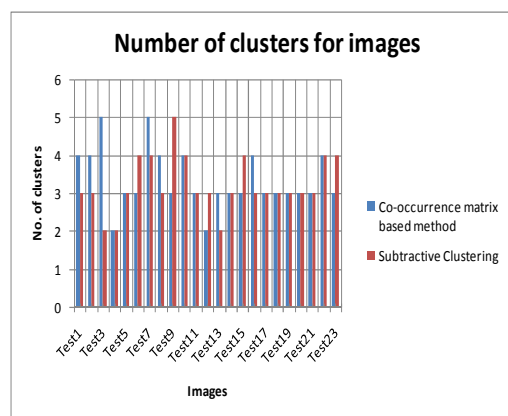

Figure 1. Comparison of cluster count found by co-occurrence matrix based method and Subtractive clustering.

As compared to K-means algorithm FCM is better for complex color images. The Table 4 providing the parameter silhouette values for the evaluation of cluster compactness of K-Means, FCM and SOM. As it is clearly evident from the Table 4 the performance of K-Means method are less effective compare to FCM and SOM for the complex color images. Advantage of FCM is each pixel is assigned membership to each cluster unlike K-means clustering a pixel may belong to more than one cluster. The main drawback of FCM is Computational time, more Sensitive to initialization condition of cluster number and cluster center.

We have chosen random initialization for SOM because with random initialization our cluster (segment) centers have good chance of searching the maximum color segmentation space available. SOM produces result close to FCM but results of SOM can be more optimized by combining another method. The experimental results of SOM are shown in Figure 4. In Table 4 it can be seen that silhouette values of 13 test images segmented by FCM and SOM are higher than K-Means. Table 3 consists of this details and comparison chart shown in Figure 2. When the k value is small then K-Means performance is good.

Table 3. Silhouette values of test images with FCM and SOM performance higher than the K-Means

| Image	K-Means	FCM	SOM
Test Image 6	0.7213	0.7939	0.7426
Test Image 7	0.769	0.7939	0.7387
Test Image 8	0.696	0.7517	0.7389
Test Image 9	0.6417	0.7479	0.7395
Test Image 11	0.7195	0.7495	0.7383
Test Image 12	0.7433	0.7931	0.7401
Test Image 13	0.9564	0.9679	0 .9041
Test Image 15	0.7345	0.7513	0.7367
Test Image 16	0.7348	0.7494	0.7362
Test Image 18	0.525	0.7492	0.7338
Test Image 19	0.6965	0.7469	0.7288
Test Image 20	0.6556	0.7926	0.7334
Test Image 22	0.7281	0.7942	0.7392
Test Image 23	0.6768	0.7497	0.7396
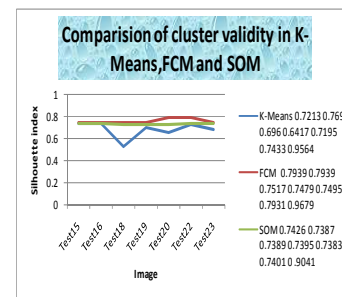
Figure 2. Comparison chart for silhouette values of the the images clustered by FCM and SOM with high Silhouette values than K-Means clustering.

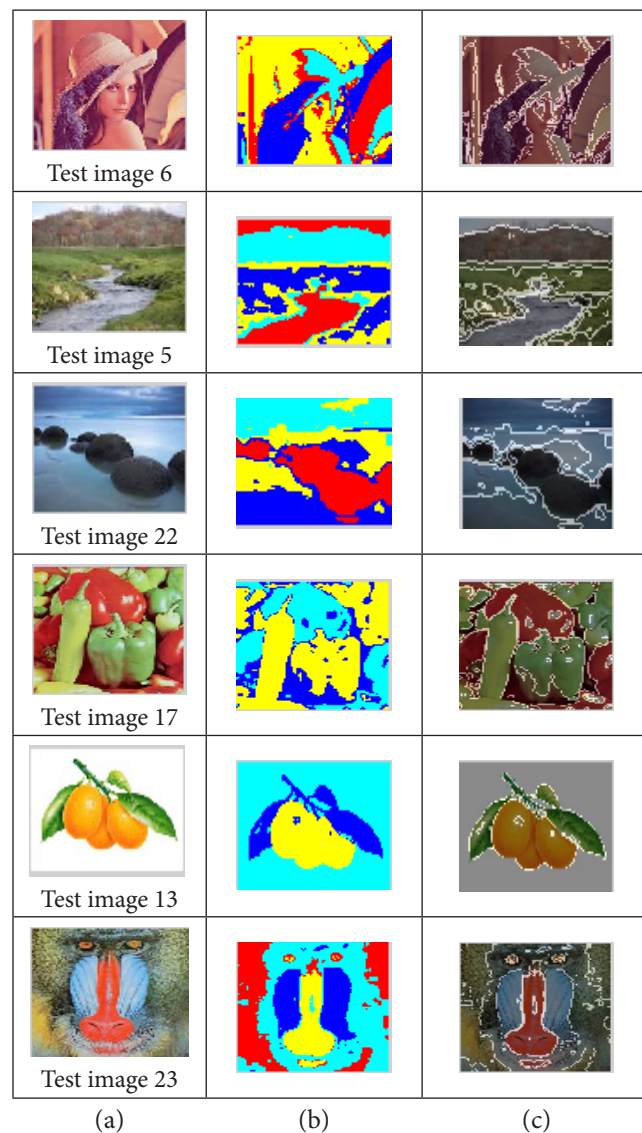|		
Test image 6		
Test image 5		
Test image 22		
Test image 17		
Test image 13		
Test image 23		
(a)	(b)	(c)

Figure 3. The experimental results of FCM on L*a*b color space test images. a) Original image. b) Segmented image. c) Marker image.
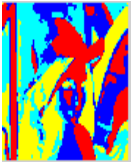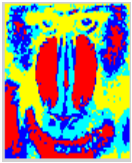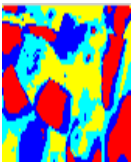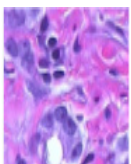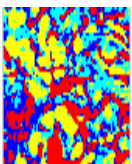
			Cluster1 – 0.7030 Cluster2 – 0.7778 Cluster3 – 0.7901 Cluster4 – 0.6997 Silhouette index – 0.7426
			Cluster1 – 0.6955 Cluster2 – 0.7817 Cluster3 – 0.7874 Cluster4 -0.6933 Silhouette index – 0.7396
			Cluster1 – 0.6872 Cluster2 – 0.7826 Cluster3 – 0.7907 Cluster4 –0.6958 Silhouette index – 0.7387
			Cluster1 –0.8047 Cluster2 –1 Cluster3 – 0.7784 Silhouette index – 0.9041
			Cluster1 – 0.6680 Cluster2 – 0.7391 Cluster3 – 0.7853 Cluster4 –.7511 Silhouette index– 0.7352
			Cluster1 –0.7325 Cluster2 – 0.7764 Cluster3 – 0.7360 Cluster4 –0.6873 Silhouetteindex – 0.7334
			Cluster1 – 0.6799 Cluster2 – 0.7901 Cluster3 – 0.7961 Cluster4 –0.6862 Silhouette index– 0.7377
			Cluster1 – 0.6760 Cluster2 – 0.9996 Cluster3 – 0.7829 Cluster4 –0.6991 Silhouette index– 0.7928
(a)	(b)	(c)	(d)

Figure 4. The experimental results of SOM arranged in 4 columns namely a,b,c and d. a) Original image. b) Segmented RGB image. c) Edge image. d) Silhouette index.

Table 4. Silhouette index value for K-Means, FCM and SOM for the test images

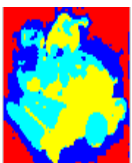| SNo.	Image	K-Means		FCM		SOM
		No. of clusters	Silhouette index value	No of clusters	Silhouette index value	
1	Test Image 1	3	0.7534	3	0.7480	0.7377
		4	0.7977	4	0.7341	
2	Test Image 2	3	0.7519	3	0.7492	0.7372
		4	0.7328	4	0.7344	
3	Test Image 3	3	0.7629	3	0.7508	0.7240
				4	0.7297	
4	Test Image 4	6	0.6795	6	0.7236	
		5	0.7110	5	0.7380	
		4	0.7740	4	0.7486	0.7421
		3	0.8408	3	0.7565	
				2	0.7953	
5	Test Image 5	2	0.8268	2	0.7937	
		3	0.7352	3	0.7586	0.7358
				4	0.7358	
6	Test Image 6	2	0.7213	2	0.7939	
		3	0.7276	3	0.7475	0.7426
		4	0.6988	4	0.7375	
7	Test Image 7	2	0.7690	2	0.7939	
		3	0.7116	3	0.7486	0.7387
		4	0.6541	4	0.7356	
		5	0.6254	5	0.7252	
8	Test Image 8	3	0.6960	3	0.7517	0.7389
		4	0.6488	4	0.7348	
9	Test Image 9	3	0.6417	3	0.7479	0.7395
		4	0.6513	4	0.7342	
10	Test Image 10	3	0.8100	3	0.7734	0.7928
		4	0.8014	4	0.7849	
11	Test Image 11	3	0.7195	3	0.7495	0.7383
		4	0.6593	4	0.7338	
12	Test Image 12	2	0.7433	2	0.7931	
		3	0.6887	3	0.7476	0.7401
		4	0.6497	4	0.7377	
13	Test Image 13	2	0.9564	2	0.9679	0 .9041
		3	0.9097	3	0.9038	
14	Test Image 14	2	0.8621	2	0.7945	
		3	0.8099	3	0.7485	0.7375
		4	0.7316	4	0.7330	
15	Test Image 15	3	0.7345	3	0.7513	0.7367
		4	0.6699	4	0.7333	
16	Test Image 16	3	0.7348	3	0.7494	0.7362
		4	0.6052	4	0.7344	
17	Test Image 17	2	0.8343	2	0.7934	
		3	0.7463	3	0.7488	0.7352
		4	0.7245	4	0.7341	
18	Test Image 18	3	0.5250	3	0.7492	0.7338
		4	0.5290	4	0.7288	
19	Test Image 19	3	0.6965	3	0.7469	0.7288
		4	0.6349	4	0.7332	
20	Test Image 20	2	0.6556	2	0.7926	
		3	0.6464	3	0.7480	0.7334
		4	0.6505	4	0.7293	
21	Test Image 21	3	0.7864	3	0.7471	0.7337
		4	0.7334	4	0.7361	
23	Test Image 23	3	0.6768	3	0.7497	0.7396
		4	0.6275	4	0.7348	

6. Conclusion

In this paper we have segmented the color images using K-means, FCM and SOM. FCM clustering algorithm applied on the CIE L*a*b* color reduced image. The trained SOM network reduced the execution time for segmenting color images. The performance of FCM and SOM is higher than the K-means for segmenting color images.

7. References

1. Gonzalez RC, Woods RE. Digital image processing. Pearson Education; 2002.
2. Patel JB, Anand RS. Color image segmentation for medical images using L*a*b* color space. IOSR Journal of Electronics and Communication Engineering. 2012; 1(2):24–45.
3. Shirakawa S, Tomoharu N. Evolutionary image segmentation based on multiobjective clustering. Evol Comput. 2009; 2466–73. doi: 10.1109/CEC.2009.4983250,
4. Yang Y, Huang S. Image segmentation by fuzzy C-means clustering algorithm with a novel penalty term. Comput Informat. 2007; 26(1):17–31.
5. Sowmya B, Sheelarani B. Colour image segmentation using fuzzy clustering techniques and competitive neural network. Applied Soft Computing. 2011; 11(3):3170–8.
6. Ebrahim A, Gharpure D. Incorporating FCM and back propagation neural network for image segmentation. International Journal of Computer and Communication Technology. 2011; 2(8):121–6.
7. Deshmukh K, Shinde GN. An adaptive neuro-fuzzy system for color image segmentation. J Indian Inst Sci. 2006; 86(5):493–506.
8. Krishnan N, Babu CNK, Rajapandian VVJ, Devaraj NR. A fuzzy image segmentation using feedforward neural networks with supervised learning. Proceedings of the International Conference on Cognition and Recognition; 2009.
9. Dong G. Xie M. Color clustering and learning for image segmentation based on neural networks. IEEE Trans NeuralNetwork. 2005 Jul; 16(4):925–36.
10. Sowmya B, Sheelarani B. Colour image segmentation using soft computing techniques. International Journal of Soft Computing Applications. 2009; 4:69–80.
11. Mújica-Vargas D, Gallegos-Funes FJ, Rosales-Silva AJ. A fuzzy clustering algorithm with spatial robust estimation constraint for noisy color image segmentation. Advances in Pattern Recognition Methodology and Applications. 2013; 34(4):400–13.
12. Yambal M, Gupta H. Image segmentation using fuzzy C means clustering: a survey. International Journal of Advanced Research in Computer and Communication Engineering. 2013 Jul; 2(7).
13. Borji A, Hamidi M, moghadam AME. CLPSO-based fuzzy color image segmentation. Fuzzy Information Processing Society, 2007. NAFIPS '07. Annual Meeting of the North American; 2007. p. 508–13.
14. Kang W-X, Yang Q-Q, Liang R-P. The comparative research on image segmentation algorithms. First International Workshop on Education Technology and Computer Science; 2009.
15. Jaffar MA, Ishtiaq M, Hussain A, Mirza AM. Wavelet-based color image segmentation using self-organizing map neural network. IPCSIT; Singapore: IACSIT Press; 2011.
16. Ortiz A, Górriz JM, Ramírez J, Salas-González D, Llamas-Elvira JM. Two fully-unsupervised methods for MR brain image segmentation using SOM-based strategies. Applied Soft Computing. 2013; 13:2668–82.
17. Harrabi R, Ben Braiek E. Color image segmentation based on a modified fuzzy c-means technique and statistical features. International Journal of Computational Engineering Research. 2(1):120–35. ISSN: 2250–3005.
18. Le Capitaine H, Frélicot C. A fast fuzzy c-means algorithm for color image segmentation. EUSFLAT-LFA 2011; 2011 Jul; Aix-les-Bains, France.
19. Paúl R, Aguirre E, García-Silvente M, Muñoz-Salinas R. A new fuzzy based algorithm for solving stereo vagueness in detecting and tracking people. International Journal of Approximate Reasoning. 2012; 53:693–708.
20. Senthilkumaran N, Rajesh R. Edge detection techniques for image segmentation – a survey of soft computing approaches. Int J Recent Trends Eng. 2009 May; 1(2):250-4.
21. Chi D. Self-organizing map-based color image segmentation with k-means clustering and saliency map. International Scholarly Research Network ISRN Signal Processing. 2011; 2011:Article ID 393891.
22. Devi DMR, Thambidurai P. Similarity measurement in recent biased time series databases using different clustering methods. Indian Journal of Science and Technology. 2014 Feb; 7(2):189–98.
23. Kumar R, Arthanariee AM. Performance evaluation and comparative analysis of proposed image segmentation algorithm. Indian Journal of Science and Technology. 2014 Feb; 7(1):39–47.
24. Sheshasayee A, Sharmila P. Comparative study of fuzzy c means and k means algorithm for requirements clustering. Indian Journal of Science and Technology. 2014 Jun; 7(6):853–7.
25. Ganesan P, Rajini V, Sathish BS, Kalist V, Basha SKK. Satellite image segmentation based on YCbCr color space. Indian Journal of Science and Technology. 2015 Jan; 8(1):35–41.
26. Pham VH, Lee BR. An image segmentation approach for fruit defect detection using k-means clustering and graph-based algorithm. Vietnam J Comput Sci. 2015; 2:25–33 DOI: 10.1007/s40595-014-0028-3.