_____

Evaluation of Clustering Image Using Steady State Genetic and Hybrid K-Harmonic Clustering Algorithms

Assist Prof. **Dr. Baheja K.Shukur** [1]**, Lecturer Sura Zaki Naji Alrashid**[2]
[1]Network Department, Information Technology, Babylon University, Hilla
[2]Software Department, Information Technology, Babylon University, Hilla
Email: baheja2003@yahoo.com, sura_os@itnet.uobabylon.edu.iq

Abstract- The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data depending on some similarity measure (e.g. Euclidean distance).In this paper a steady state genetic algorithm (SSGA) approach is used to cluster true color images. After splitting the original images into red, green and blue components and displaying the image of each part, Steady State Genetic Algorithm (SSGA) is used to cluster the image to determine the number of clusters for the image by generating an initial population randomly and then applying the different operations of GA such as fitness function computation, selection, crossover, mutation and stopping condition. In the Crossover stage 1X, PMX and UX methods used for crossover between two parents to produce a new child. In addition to that another clustering method which combines k-mean algorithms and k-harmonic mean algorithms are used. The last clustering algorithm uses two functions to find the cluster centers for each image. Finally root mean square error is used to find the difference between the clustering and original image.

Keywords - The steady state genetic algorithm (SSGA), UX method, k-mean.

1.   Introduction

Clustering can be considered as the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way" [1]. A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters [2]. The similarity criterion is distance: two or more objects belong to the same cluster if they are "close" according to a given distance. Clustering algorithms may be classified as listed below [3]:

- Exclusive Clustering
- Overlapping Clustering
- Hierarchical Clustering
- Probabilistic Clustering

In the first case data are grouped in an exclusive way, so that if a certain data belongs to a definite cluster then it could not be included in another cluster [4]. On the contrary the second type, the overlapping clustering, uses fuzzy sets to cluster data, so that each point may belong to two or more clusters with different degrees of membership. In this case, data will be associated to an appropriate membership value. Instead, a hierarchical clustering algorithm is based on the union between the two nearest clusters. The beginning condition is realized by setting every data as a cluster. After a few iterations it reaches the final clusters wanted. Finally, the last kind of clustering uses a completely probabilistic approach [5]. An important component of a clustering algorithm is the distance measure between data points. If the components of the data instance vectors are all in the same physical units then it is possible that the simple Euclidean distance metric is sufficient to successfully group similar data instances [6].

2.   Related Works

Clustering Image segmentation is a process of dividing an image into different regions based on certain attributes such as intensity, texture, color, etc. This process is fundamental in computer vision and many applications, such as object recognition, image compression, image retrieval, and visual summary, which can benefit from it.

Color images can increase the quality of segmentation, but increase the complexity of problem. Genetic algorithms are well suited to optimizing this complex problem. Pachunde et al., [7] gave a state- of - the–art Genetic Algorithm based clustering methods. Gui et al., [8] applied K-harmonic means instead of K-means to improved spectral clustering procedure in order to raise its stability and performance. Experimental results show that the proposed approach is effective on SAR image segmentation and has the advantage of calculating quickly. Sankari and Chandrasekar [9] used genetic algorithm to find optimal color classes so that the colors in the natural image are clustered sharply. . Al-Guwaizani [10] Suggested a Variable Neighborhood Search (VNS) to improve a recent clustering local search called K-Harmonic Means (KHM). Hauschild et al., [11] proposed a hybrid genetic algorithm to perform image segmentation based on applying the q-state Potts spin glass model to a gray scale image. First, the image is converted to a set of weights for a q-state spin glass and then a steady-state genetic algorithm is used to evolve candidate segmented images until a suitable candidate solution is found. Hierarchical local search is used on each evaluated solution in order to speed up the convergence to an adequate solution.

_____

3. The genetic algorithms

Most of the classical search methods used to solve the optimization problems suffered from a number of problems and difficulties especially when used to solve a complex problem. Below are some of the problems which the methods suffered from [12].

1- The approximate to the optimal solution depended on the selected initial solution.

2- The classical methods are not efficient to treat with the discreet variables.

All the genetic algorithms in random population, represented as the set of the solution, allocated to each solution specific "fitness" connected directly with the object function for the specific problem and then another population is generated from applying a set of genetic operators like selection, crossover, mutation, continuously and sequential to the generation of that population until the condition is satisfied [13]. Moreover, genetic algorithm characteristically solves most difficulties and problems mentioned above which the classical methods suffered from [14]. Below are the following properties:-

1- The genetic algorithms worked on coding the set of parameters which represent the variables of the decision to the problem and not on the parameters itself.

2- The genetic searches from a set of point in the search space and not on the one point like the classical methods.

3- The genetic algorithm uses the object function information and not derivative or no other information.

4- The genetic algorithms used the probabilistic rules and not deterministic rules [15].

There are two basic forms of the genetic algorithms:-

1- The simple genetic algorithm (SGA).

2- The steady state genetic algorithm (SSGA)

The second of the genetic algorithms have been used in our study, in which the population changes gradually to generate the new generation one after one.

4. The Proposed System

Figure 1 shows the block diagram of the proposed system. In the following, the function of each block will be described.

4.1 Read Image

In this stage of the proposed system the images acquired consist of four parts (red, green, blue, additional part). After separating these parts, we need to contract the resulted image and then display images. The read processing of images file depends on image file format.

Algorithm:-read image
```
Pos = 54
 i = 1
 Do
 Get #1, pos, r1 (i): pos = pos + 1
 Get #1, pos, b1 (i): pos = pos + 1
 Get #1, pos, g1 (i): pos = pos + 1
 i = i + 1
 Loop Until i >= height * width
 x = 1
 y = h
 i = 1
 Do
 R(x, y) = r1 (i): r2(x, y) = r1 (i)
 G(x, y) = g1 (i): g2(x, y) = g1 (i)
 B(x, y) = b1 (i): b2(x, y) = b1 (i)
  x = x + 1
 If x > width Then
  x = 1
  y = y - 1
 End If
  i = i + 1
 Loop Until i > w * h
 i = 1
 Do
 Get #1, i, a
 Put #2, i, a
 i = i + 1
 Loop Until i = 55
 Close #1
  Close #2
```
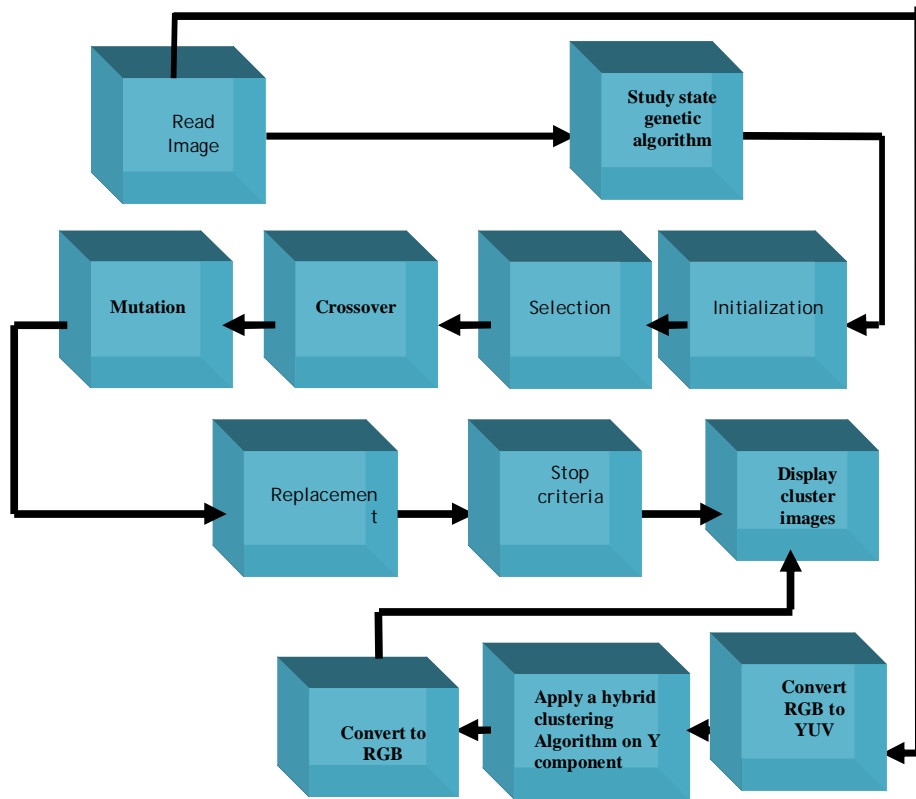
Figure 1: A block diagram of proposed system

4.2. Study state genetic algorithm
4.2.1. Initialization step

In this stage each chromosome is constructed from

X: taken from the width of the image randomly

Y: taken from the height of the image randomly.

The population consists of a number of chromosomes. Each chromosome contains a number of genes. Each gene consists of (x, y, red, green, blue) where x, y represented the location of the pixels in the image selected randomly.

<u>**Algorithm :-initialization**</u>
For i1 = 1 to popsiz
For j1 = 1 to the length of the chromosome
  *popu.popu1 (i1).chromosome mosome (j1).x = Int ((w * Rnd) + 1)*
  *popu.popu1 (i1).chromosome mosome (j1).y = Int ((h * Rnd) + 1)*

  X1 = popu.popu1 (i1).chromosome mosome (j1).x
  Y1 = popu.popu1 (i1).chromosome mosome (j1).y
  popu.popu1 (i1).chromosome mosome (j1).red = r(X1, Y1)
  popu.popu1 (i1).chromosome mosome (j1).grn = g(X1, Y1)
  popu.popu1 (i1).chromosome mosome (j1).blu = b(X1, Y1)

4.2.2. Evaluation step

The fitness of a chromosome is computed using the Euclidean distance. If $p_1 (x_1, y_1)$ and $p_2 (x_2, y_2)$, it is $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

<u>**Algorithm Evaluation**</u>

For i=1 to pop size
Sum=0
For i11 = 1 to pop size
  Sume = 0
  dist = 0
  fit = 0
For i = 1 to image width

_____

For j = 1 to image height
pix.xx = i
pix.yy = j
pix.rr = r (i, j)
pix.gg = g (i, j)
pix.bb = b (i, j)

For j11 = 1 to chromosome length
Dx = ((pix.xx - popu.popu1 (i11).chromosome (x) ^ 2) / width
Dy = ((pix.yy - popu.popu1 (i11).chromosome (j11).y) ^ 2) / height
Dr = ((pix.rr - popu.popu1 (i11).chromosome (j11).red) ^ 2) / 256
Dg = ((pix.gg - popu.popu1 (i11).chromosome (j11).grn) ^ 2) / 256
Db = ((pix.bb - popu.popu1 (i11).chromosome (j11).blu) ^ 2) / 256
Dst = dx + dy + dr + dg + db
ds1 = (ds) ^ 0.5
Minn (j11) = ds1
Mine = minn (1)
For j1 = 1 to chlen
If minn (j1) <= mine Then
Mine = minn (j1)
k = j1
End If
Varr (k) = varr (k) + mine
For k1 = 1 to chlen
Sume = Sume + varr(k1)
Next k1
*Dist = Sume / (h * w * 5)*
Fit = 1 / (1 + dist)
fitt(i11) = fit
For i11 = 1 to pop size
popu.popu1 (i11).fit = fitt(i11)

4.2.3. Selection step

The first crossover operator is selected randomly from the population size which represented the first parent and then the second crossover operator is randomly chosen representing the second parent with a condition that any two parents are not to have the same value at each iteration.

Algorithm Selection:-
Step1:-Generate first parent (r1)
Step2:- generate second parent (r2)
Repeat step (2) until r1 become different from r2
Step3:- if fitness (r1)>fitness (r2) then
Per=r1
Else
Per=r2
End if

Step4:- selection r1 or r2 is depending on the fitness function.
Step5:- repeat these steps from (step1) to (step4) to generate the second parent.

4.2.4. Crossover step

Perform crossover step by using (UX). This operator is considered to be very different from others, which has been randomly exchanged with probability (0.5). From the genetics of parents two children are produced. This crossover method has large effective to destroy the chromosome but almost gives a good performance. The power of this method is its ability to enhancement the variety of the population. This method depends on the SSGA.

Algorithm Crossover:-
For i = 1 to the length of the chromosome
Generate w = Rnd (2)
If w = 1 Then
Chld (1).chromosome (i).x = popu1 (per1).chromosome (i).x
Chld (1).chromosome (i).y = popu1 (per1).chromosome (i).y
Chld (1).chromosome (i).red = popu1 (per1).chromosome (i).red
Chld (1).chromosome (i).grn = popu1 (per1).chromosome (i).grn
Chld (1).chromosome (i).blu = popu1 (per1).chromosome (i).blu
End If
If w = 0 Then
Chld (1).chromosome (i).x =popu1 (per2).chromosome (i).x
Chld (1). Chromosome (i).y =popu1 (per2).chromosome (i).y
Chld (1). Chromosome (i).red = popu1 (per2).chromosome (i).red
Chld (1). Chromosome (i).grn = popu1 (per2).chromosome (i).grn
Chld (1). Chromosome (i).blu = popu1 (per2).chromosome (i).blu
End If
For i = 1 to the length of the chromosome
Generate w1 = Rnd (2)
If w1 = 0 Then
Chld (2). Chromosome (i).x =popu1 (per1).chromosome (i).x
Chld (2).Chromosome (i).y =popu1 (per1).chromosome (i).y
Chld (2).Chromosome (i).red =popu1 (per1).chromosome (i).red

_____

Chld (2).Chromosome (i).grn =popu1 (per1).chromosome (i).grn
Chld (2).Chromosome (i).blu = popu1 (per1).chromosome (i).blu
End If
If w1 = 1 Then
Chld (2).Chromosome (i).x = popu1 (per2).chromosome (i).x
Chld (2).Chromosome (i).y = popu1 (per2).chromosome (i).y
Chld (2).Chromosome (i).red = popu1 (per2).chromosome (i).red
Chld (2).Chromosome (i).grn = popu1 (per2).chromosome (i).grn
Chld (2).Chromosome (i).blu = popu1 (per2).chromosome (i).blu
End If

4.2.5. Mutation

From the two children generated from the crossover stage a number of genes is randomly chosen not exceeding the length of the chromosome. Then x, y are randomly generated which represent a pixel of the image and exchange the x, y contain at this selected genetic with new generated x,y. then evaluated the new population generated at the same fitness function.

Mutation algorithm

Generate the first mutation operator ra1randomly from the chromosome (ran1)
Generate the first mutation operator ra2 randomly from the chromosome (ran2)
Generate m1genx randomly from the width
Generate m1genyrandomly from the height
Generate m2genx randomly from the width
Generate m2geny randomly from the height
Chld (1).chromosome (ran1).x = m1genx
Chld (1).chromosome (ran1).y = m1geny
Chld (1).chromosome (ran1).red = r (m1genx, m1geny)
Chld (1).chromosome (ran1).grn = g (m1genx, m1geny)
Chld (1).chromosome (ran1).blu = b (m1genx, m1geny)
Chld (2).chromosome (ran2).x = m2genx
Chld (2).chromosome (ran2).y = m2geny
Chld (2).chromosome (ran2).red = r (m2genx, m2geny)
Chld (2).chromosome (ran2).grn = g (m2genx, m2geny)
Chld (1).chromosome (ran2).blu = b (m2genx, m2geny)

4.2.6. The replacement

At this study we use Binary Tournament Replacement to replace new people instead of the worst of the two persons chosen randomly. The parameter of choosing the best person depends on the height fitness function.

Replacement algorithm

If Chld (benu).fit > popu.popu1 (popsize).fit Then
*   popu.popu1 (popsize) = Chld (benu)*
End If
Where benue is the best person result.

Also in our work the error function has been used to find the amount error produced through the image clustering.

4.2.7. Stop criteria

To stop all above steps when we found the more fitness function have the similar or approximately value consequently.

5. Convert RGB to YUV color space

There are many slightly different formats to convert between YUV and RGB. These formulas assume U and V are unsigned bytes [16] [17].

$$Y=R*0.299000+G*0.587000+B*0.114000$$

(1)

$$U=R*-0.168736+G*-0.331264+B*0.500000+128$$

(2) $$V=R*0.500000+G*-0.4188+B*-0.081312+128$$ (3)

6. The algorithm of the hybrid clustering Algorithm

1. Begin
2. Initialize Dataset DN, K, C_1, C_2, . . . , C_K, Current Pass=1; where D is dataset, N is size of data set,
K is number of clusters to be formed,
C_1, C_2, . . . , C_K are cluster centers.
Current Pass is the total no. of scans over the dataset.

_____

3. Do assign the n data points to the closest C_i;

If Current Pass%2==0

Recomputed $C_1, C_2, \ldots, C_K$ using Harmonic Mean function;

Else

Recomputed $C_1, C_2, \ldots, C_K$ using Arithmetic Mean function;

Increase Current Pass by one.

Until no change in $C_1, C_2, \ldots, C_K$;

4. Return $C_1, C_2, \ldots, C_K$;

5. End

7. Case study

This system is tested for different types of images. Some of these cases can be shown below in Tables 1 and 2 and Tables 3 and 4.

Case 1: Test the system on the flower image in Table (1), multi objective image based on SSGA clustering algorithm in Table 2 as shown in Figures 1 and 3.

Table (1): The tested system on the flower image using SSGA

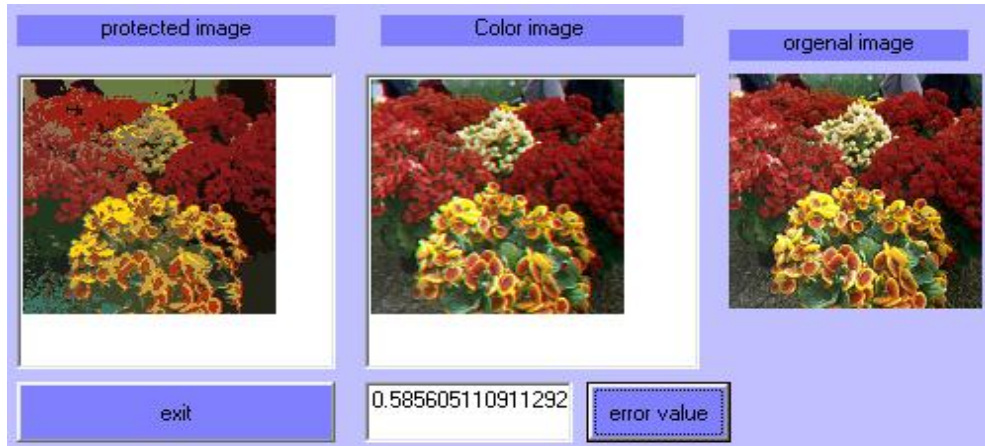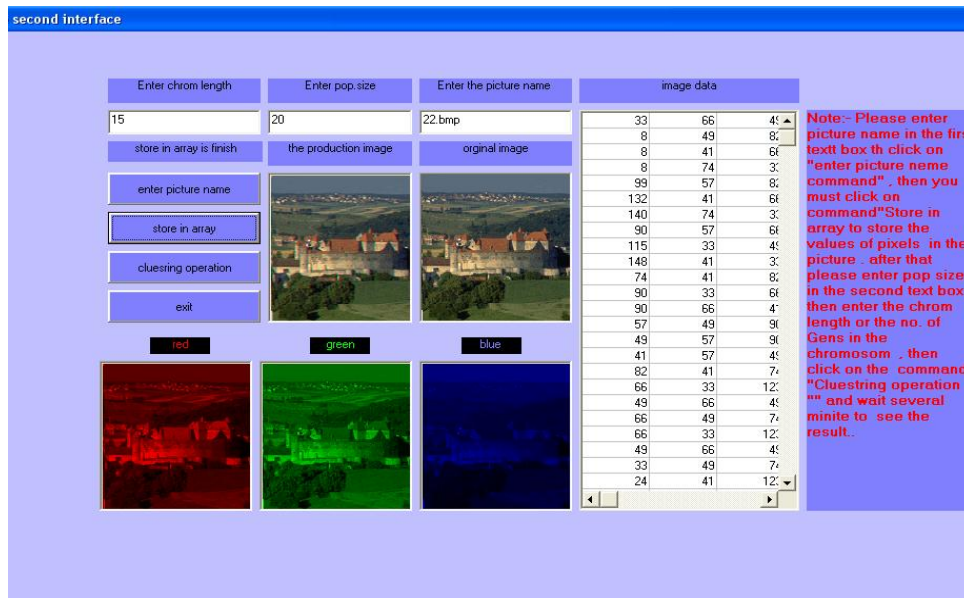| Image size	128*128
No. of population	20
Chrom.length	10
Max fitness	0.6892686
Error value	0.868438837140105

16

_____

Figure (2): the implementation of the SSGA on flower image

Table (2): The tested system on the multi objective image using SSGA

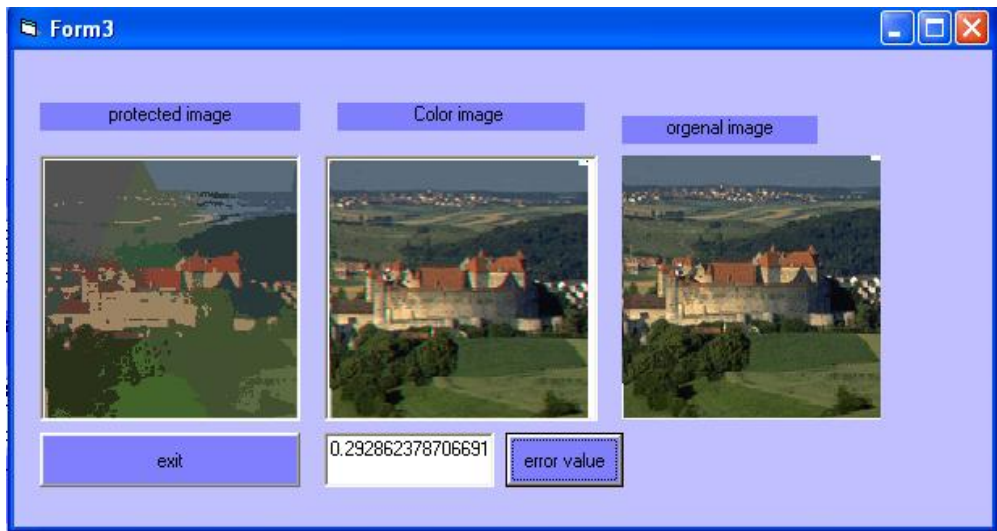|	
Image size	**160*160**
No.of population	**20**
Chrom.length	**15**
Max fitness	**0.6295041**
Error value	**0.292862378706691**

_____

Figure (3): the implementation of the SSGA on multi object image

Case 2: flower image and multi object image tested on the hybrid clustering method

Table (3): The tested system on the flower image using hybrid k-harmonic algorithm

Image size	128*128
No. of iteration	30
Number of cluster	10
Error value	1.41576730294866E-05

**Table (4): The tested system on the multi objective image using hybrid k-harmonic algorithm**

Image size	128*128
No. of iteration	30
Number of cluster	10
Error value	1.41576730294866E-05

Figure (4): the implementation of hybrid clustering algorithm on the above image

Also, this system is tested on different true images and the behavior for these clustering methods can be shown in figure (5). Figures a, b, c and d show the behavior of the SSGA, whereas e, f, g and h show the behavior of the hybrid algorithm.

(a)   *(b)*   *(c)*   *(d)*

(e), c=8   *(f), c=8*   *(g), c=5*   *(h), c=5*
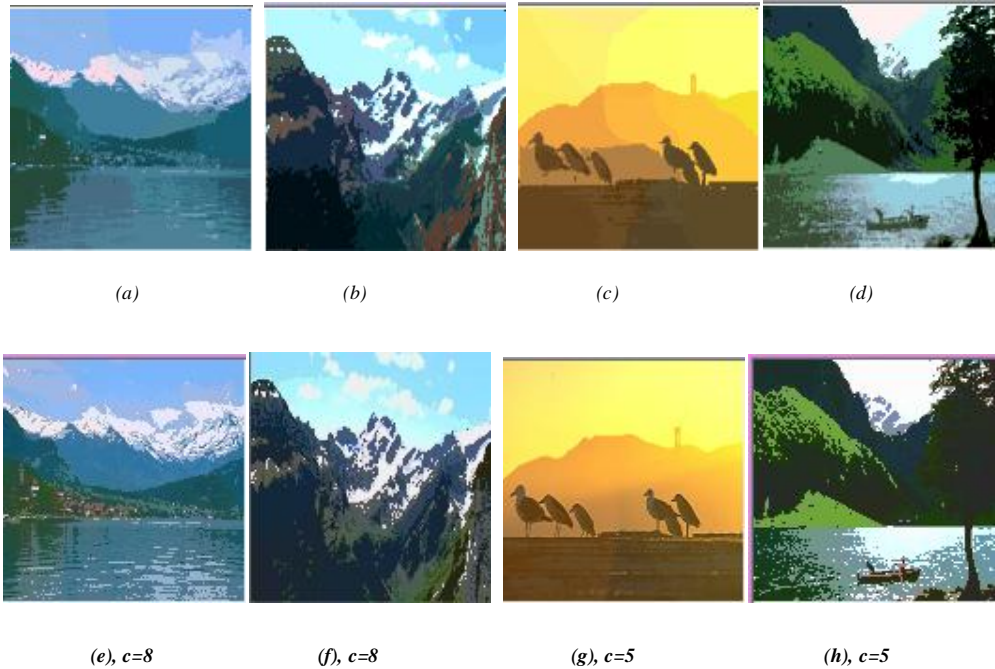
Figure (5): The implementation for the proposed system for both methods for different cluster numbers

8. Conclusion

a- The relation between the distance and the fitness function is revered. This means
(I.e. fitness=1/distance).

b- The number of the clusters depends on the image's nature and complexity and not on the size of the image.

c- The resulted cluster images depend on the nature and complexity of the image.

d- The increasing of the size of the population and the length of the chromosome obtained the best resulted in the SSGA algorithm.

e- The error value of the homogeneous color image is greater than non-homogenous color image although the clustering process in the homogenous is better.

f- The three methods of crossover (1X, UX, PMX) which have been used in this paper have the same result; this refers to the power of the fitness function.

g- The hybrid k-harmonic clustering method provided suitable results in a less number of iterations; also the root mean square error is small which reflects the goodness of this method.

h- The result presented in this paper showed that the hybrid k-harmonic clustering method is more suitable than the other methods, which gave a good result in small number of clustering.

i- Using of RGB color space in the first method and converting RGB to YUV color space in the second method provide good results.

_____

References

[1] A. Jain, M. Murty and P. Flynn. "Data Clustering: A Review. ACM Computing Surveys", vol. 31, no. 3, pp. 264-323, 1999.

[2] A. Jain, R. Duin and J. Mao. "Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 22, no.1, pp. 4-37, 2000.

[3] G. Hamerly and C. Elkan, "Alternatives to the K-means Algorithm that Find Better Clustering's". In Proceedings of the ACM Conference on Information and Knowledge Management (CIKM-2002), pp. 600-607, 2002.

[4] G. Hamerly and C. Elkan. "Learning the K in K-means". In The Seventh Annual Conference on Neural Information Processing Systems, 2003.

[5] G. Coleman and H. Andrews. "Image Segmentation by Clustering". In Proceedings of IEEE, vol. 67, pp. 773-785, 1979.

[6] A. Jain and R. Dubes. "Algorithms for Clustering Data". Prentice Hall, New Jersey, USA, 1988.

[7] P. Pachunde, V. R. Ingle , M. A. Gailwad, "Segmentation of Color Images Using Genetic Algorithms: A Survey", IOSR Journal , ISSN: 2278-1676 Volume 1, Issue 6 (July-Aug. 2012), PP 09-12.

[8] Y. Gui, X. Zhang and Y. Shang, "SAR image segmentation using MSER and improved spectral clustering", EURASIP Journal on Advances in Signal Processing 2012.

[9] L. Sankari and C. Chandrasekar ,"Semi Supervised Image Segmentation Using Optimal Color Seed Selection", IACSIT, Vol. 4, No. 6, December 2012.

[10] A. Al-Guwaizani, "Variable Neighborhood Search Based Heuristic for K-Harmonic Means Clustering", PHD. Thesis, Brunel University, London, May, 2011,

[11] M.Hauschild, Bhatia, M. Pelikan, "Image Segmentation using a Genetic Algorithm and Hierarchical Local Search", GECCO'12,July 7–11, 2012, Philadelphia, Pennsylvania, USA.

[12] K.Deb, "Genetic algorithm in Search and optimization the technique and applications", proceedings of International workshop soft computing and intelligent systems, Calcutta, 85, (1998).

[13] D.E Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesely, Massachuchusetts. 1998.

[14] M.Mitchell,"An Introduction to genetic algorithms", HIT Press, London, 1996.

[15] M. Schmidt, T. Stidsen, "Hybrid Systems: Genetic algorithms", Neural Networks and Fuzzy Logic, DAMI IR, 1997.

[16] Panrony X., "Image Compression by Wavelet Transform", East Tennessee State University, Master Thesis, 2001.

[17] Plataniotis K. and Venetsanopoulos, "Color Image Processing and Application ", Springer, 2000.