Semi Supervised Image Segmentation by Optimal Color Seed Selection using Fast Genetic Algorithm

L.Sankari
Assistant Professor
Department of computer Science
Sri Ramakrishna CAS for Women
Coimbatore -44, India

Dr.C.Chandrasekar
Reader
Department of Computer Science,
Periyar University
Salem, India

ABSTRACT

Key factors like similarity, proximity, and good Many researchers have mentioned the significance of perceptual grouping and organization in vision and listed various continuation that guide to visual grouping of image. However, even to the present situation, many of the computational factors of perceptual grouping have remained unanswered. As there are several probable partitions of the domain of an image into subsets, however it is significant to choose the correct choice. In general Image segmentation refers to the process of partitioning the input image into several disjoint regions with similar characteristics such as intensity, color, and texture, shape etc. There are several algorithms exist based on supervised, Unsupervised and Semi supervised techniques. But all algorithms has several disadvantages like lack of accuracy, more time, etc, The Existing semi supervised clustering method uses mouse clicks as prior information or certain constraints and then Clustering . In this paper semi supervised clustering using prior information is discussed. The prior information is nothing but selected color seeds using FGA and then EM Clustering. The proposed idea results in better visual appearance and also requires only lesser time when compared to the other segmentation using only GA or by using Kmeans clustering.

Keywords--- EM Clustering, Genetic Algorithm, Fast Genetic Algorithm, Semi Supervised Clustering.

1. INTRODUCTION

Image segmentation [14] is helpful in several applications. With the help of the segmentation results, it is probable to classify regions of interest and objects in the picture that is highly valuable to the subsequent image analysis or annotation. Current work includes a variety of techniques like stochastic model based approaches [6] [7], [8], [9], morphological watershed based region growing [10], energy diffusion [11], and graph partitioning [12]. Quantitative evaluation techniques have also been suggested [13]. However, because of the complex nature of the problem, there are only little automatic techniques which can work effectively on a large variety of data.

The difficulty in image segmentation is mainly due to the image texture. If an image consists only homogeneous color regions, clustering techniques in color space are adequate to deal with the difficulty. In truth, natural scenes are highly in color and texture. It is hard to detect the image regions containing color-texture patterns. The following assumptions need to be considered for segmentation:

- Each region in the image contains a uniformly distributed color-texture pattern.

- The color information in each image region can be represented by a few quantized colors, which is true for most color images of natural scenes.

- The colors between two neighboring regions are distinguishable - a basic assumption of any color image segmentation algorithm.

There are two ways to perform this semi supervised image segmentation.

1) Semi supervised classification

2) Semi supervised clustering

Semi-supervised classification employs more number of labeled data and minimum of unlabeled data. But semi-supervised clustering uses more unlabelled data and less labeled data. The prior information is given either as mouse clicks or as pair wise constraints. There are some examples to aid unsupervised clustering and these are shown in several recent projects like (Wagstaff et al., 2001; Basu et al., 2002; Klein et al., 2002; Xing et al., 2003; Bar-Hillel et al., 2003; Segal et al., 2003).. For automated segmentation with better visual, Genetic Algorithm [15] [16] [21] is used. But GA requires more time for execution. To overcome this problem, this paper uses Fast Genetic method for segmentation.

2. RELATED WORKS

Jun Tang [1] proposed a color image segmentation algorithm based on region growing. In the field of image processing, image segmentation is a common topic. Also it is a more concentrated and most focused in the field of image processing techniques. Color image segmentation methods are more concerned by the researchers due to the growth of computer processing facilities and the increased application of color image. Color image segmentation techniques are an extension of the gray image segmentation technique, but most of the gray image segmentation techniques can not be applied directly to color images. This necessitates the improvement in the technique of real gray image segmentation technique according to the color image. This research introduces a color image segmentation technique of automatic seed region growing on basis of the region with the integration of the watershed algorithm with seed region growing algorithm which based on the conventional seed region growing algorithm.

Juyong Zhang *et al.,* [2] put forth a diffusion approach to seeded image segmentation. Seeded image segmentation is a most admired type of supervised image segmentation in computer visualization and image processing. The images were treated as a weighted graph in the earlier methods of seeded image segmentation technique and reduce an energy function on the graph to build segmentation. In this research, the author proposed to carry out the seeded image segmentation based on the result of a heat diffusion method in which the seeded pixels are regarded as the heat sources and the heat diffuses on the image beginning from the sources. Once the stable state is attained after the diffusion, the image is segmented according to the pixel temperatures. Random Walk algorithm is comprised in this implemented framework for image segmentation which diffuses only along the two coordinate axes. For the purpose of controlling the diffusion this paper incorporates the features of the image into the diffusion method, providing an anisotropic diffusion technique for image segmentation. The experimentation result shows that the implemented anisotropic diffusion process yields better segmentation results. When this technique is tested by using the ground truth dataset of Microsoft Research Cambridge (MSRC), achieves an error rate of 4.42%, which is very lesser than the other modern algorithms.

A novel image segmentation method based on random walk is given by Yi-hua Lan *et al.,* [3]. To reduce the difficulties, image segmentation depended on the random walk model in graph hypothesis can be altered into large-scale sparse linear equations. The absolute result of the equation and the iteration convergence rate is according to the selection of the initial value. If the initial value is chosen randomly to segment the large scale image then it ends in a significant disadvantage. In this research, the author proposed a novel image segmentation technique depended on random walk model. Initially the author down-samples the original large image to the small image, then the small image segmentation guides to sparse linear equations in very small scale. Once the solution is obtained, the possible results will be up-sampling to the up layer and then the sparse linear equations in this layer are solved. On repetition of this up-sampling process until to the top layer where the original image is obtained. The final probability image is segmented at the end with a pre-set threshold. The algorithm is tested by taking two images and the segmentation results are compared with the original random walk algorithm. The segmentation result ensures that this method is far better from other techniques. This algorithm works by taking the low-scale image probability output as the initial value of the high-scale image segmentation process. The author concluded that under the same computation time the segmentation result by our algorithm is much better than that by the original random walk segmentation algorithm.

Ye Hou *et al.,* [4] suggested image segmentation based on GC-CV. This is an integrated method which was formulated and applied to image segmentation. The proposed technique combines the advantages of both the graph cut method and the simplified Mumford-Shah model (C-V model). The proposed approach is tested under its three different operational modes. The first mode directly segments the binary images using GC-CV. The second mode uses recursive GC-CV to segment the multi-region images. In the last mode, both the color images and gray images are segmented with the combination of GC-CV

technique and EM algorithm and also uses the YCbCr color space for the segmentation of color image. With the use of many serious experiments, the feasibility and effectiveness of the proposed GC-CV method is verified. From the experimental results it is concluded that the proposed GC-CV method significantly improves the speed of segmentation and considerably reduces the number of iterations when compared with C-V model.

3. SEMI SUPERVISED IMAGE SEGMENTATION

Semi supervised clustering [17] [19] [22] means Grouping of objects such that the objects in a group will be similar to one another and different from (or unrelated to) the objects in other groups with related to certain constraints or prior information.

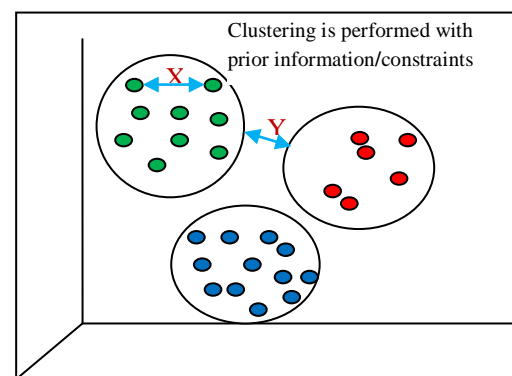The following figure represents the semi supervised clustering model.

Figure 1: Semi supervised Clustering model.

X   →Intra cluster distancce(minimized)
Y   →Inter cluster distancem(maximized)

A set of unlabeled objects, each described by a set of attributes and a small amount of domain knowledge is given. The image is read and the pixels are stored in a matrix. The prior information is nothing but the number of color groups for segmentation. With these information the EM clustering is done. Normally, for EM algorithm the cluster centers are assigned randomly. This method was used in many papers. Since the cluster centers were chosen randomly, for certain images the image regions are not segmented properly. So in the proposed method the initial color centers are obtained using histogram and optimal color centers are chosen. The following section discuss about the proposed method.

3.1 Genetic Algorithm

In order to eliminate local maxima problem of EM clustering the idea of proposed method developed. For optimization of color cluster centers the GA is applied [18] [19] [20] . This algorithm takes various cluster centers but generate only the optimal seeds as center points.By providing the initial cluster centers, GA initially looks for the fitness of the initial cluster centers. Then each cluster center is selected and kept as a parent, then with the parent, crossover and mutation is performed so that new set of cluster centers are generated. All the obtained cluster centers are kept as updated population. Those obtained cluster centers are evaluated with its fitness. Again the obtained cluster centers are kept as a parent and crossover and mutation operations are performed. The resulted

cluster centers are checked for fitness and the process goes on. This process will continue till no more new population is generated and all within the fitness level.

```
begin GA
        Choose initial population as random
        Evaluate initial population
        while not convergence
        do
        begin
Select two parents from the current population
Produce children by the selected parents
Mutate the children
Replace the worst individual of the population by
the best child
            end
        Output the best individual found
end GA
```

Fig 2: Overview of Genetic Algorithm

After the reproduction (crossover and mutation) is performed, Expectation Maximization algorithm is used for secondary level learning. After this learning is performed, the fitness is evaluated for the generated offspring. This will help in better learning in order to obtain better clustering result in lesser time.

3.2 Fast Genetic Algorithm

To reduce the complexity and operation time of GA, a fast genetic algorithm (FGA) is used according to the idea of paper [19] which results in less complexity and less time. Figure 3 shows the proposed FGA. It begins its operations as selecting two chromosome from a dataset using pseudo random number generator, RND( ) after initializing all bits in memory to zero. Selection operator (SO) determines between two selected chromosomes according to its fitness value. The fitness function derives the classifier being balanced between successful recognition and generalization capabilities. Here, the fitness function is defined as the term for the system correctness and successful recognition rate and, the term for class generalization.

The fitness function is used to measure the similarity of each individual with the original image. So let the original image be represented by the following vector x_i where i=1, …, K and the initial population by: Y_i^j where j =1, …, N and i =1,…, K. Then, the fitness value f (Y^j) of each individual of the population is computed as follows:

$$f(Y^j) = \frac{1}{\sum_{i=1}^{k} |Y_i^j - x_i|} , j = 1, …, N \qquad (1)$$

In case of the steady-state GA model, because it replaces two previous parents with two offspring that generated by genetic operator, it has fast convergence speed but it generally lack of diversity and is easy to fall to local minima. To overcome these problems, FGA model devises to compare a worse data of two clusters with the best data among two centers and the migrated data. Good data also increase the entire average fitness value of data. And then, it updates data whenever the fitness is better than worse cluster. In result, the probability of premature convergence increases relatively and the preservation of data is

improved as compared with random selection of survival-based GA.

```
1.Initialize Parent β and best chromosome and initialized to zero
    While (optimal value of Best chrom) do
2. Select two chromosome s from population using two RNs
3. Copy the previous o rent for the other parent
        Parent β =previous Parent α
4. Decide one parent from two chromosomes using fitness value
5. Do Mutation and Crossover
  Crossover (α ; β) : [1-point, 2 point, uniform]
  Childrendata (α ; β) = Crossover (Parent α , Parent β )
Mutation (α ; β) : [single-point, multi-point](Pm)
Childrendata (α ; β) = Mutation (Children α data, Children β
data)
6. Evaluate the fitness on a specific problem
  Childrenfitness (α ; β) = Fitness (Children α data, Children β
data)
7. Update a population using new offspring and Migration data
8. Determine Mutation rate
    H=hamming(New chromfitness, previous Best chrom)
    Mutation rate = pm (H)
9. Migrate data to other GA
    END While
```

Figure 3: The Fast Genetic Algorithm

The crossover operator (XO) exchanges the chromosome strings of two individuals from a crossover n-th index point l_n or random mask. Proposed FGA supports three kinds of crossover patterns such as 1-point, 2-points, , and uniform crossover.

The mutation operator (MO) inverts one bit of the individual and n-point mutation operation means that MO is applied n times. It is used for recovering the genetic characteristics and depends on the mutation probability p_m. Here, p_m is applied to single-point and multi-point for enhancing the schema of GA.

In addition, migration operator imports the best chromosome from external GA for adjusting data diversity. Fitness evaluation with the internal and migrated follows to determine one with higher fitness value between them. Finally, the data is updated by this survived cluster. Empirical change of GA parameters and methods can be added at the end of the procedure in order to obtain higher convergence speed. For the effective implementation of proposed FGA, the iterative characteristics of GA are considered to compromise between the complexity and the speed of computation. For this purpose, it is divided into concurrent parallel groups and each parallel group is connected and process serially on cluster and parallel on chromosome. This shows the FGA will take only less time when compared to the GA and the complexity is also reduced.

Initially, the entire image data is passed to Fast Genetic Algorithm. From the pixel values, random few pixel vales are taken and passed for initialization in Fast Genetic Algorithm. With this pixel color vales, the new population is generated using the crossover and mutation operation. This will results in new color value. This value is passed

for evaluating its fitness using the fitness function. If the fitness is satisfied, then the process is terminated; otherwise, crossover and mutation operation is performed again with the new population. Fast Genetic Algorithm has the capability to perform the selection of optimal color seeds faster than Genetic algorithm as described in the previous algorithm.

Then these generated optimal colors are given as initial values for clustering using Expectation Maximization clustering algorithm.

3.3 Expectation Maximization Clustering

Expectation Maximization algorithm begins with initial parameters X and Y; where Y is the store image and X is the optimum color seed obtained by using Fast Genetic Algorithm.

The joint pdf of the complete data Y, X is written as:

$$p(Y, X; \theta) = \prod_{n=1}^{N} p(Y_n | X_n; \theta) p(X_n; \theta) \quad (2)$$

Where $\theta = [E, P, A]$ with $E = [e_1, e_2, \ldots, e_k]$, $P = [p_1, p_2, \ldots, p_k]$ and $A = [A_1, A_2, \ldots, A_N]$

For the EM algorithm, the expected log-likelihood function is given by:

$$Q(\theta^i; \theta^{i-1}) = E[\log(p(Y, X; Y, \theta^{i-1})] \quad (3)$$

From the above equation, the E (expectation) and M (maximization) steps of the generalized EM algorithm can be derived as follows:

E-Step

$$p(X_n = k | Y_n; \theta^{i-1}) = \frac{p(Y_n | X_n = k; \theta^{i-1}).P(X_n = k; \theta^{i-1})}{\sum_{k=1}^{K} P(X_n = k; \theta^{i-1}) p(Y_n | X_n = k; \theta^{i-1})}$$
(4)

M-Step

$$P(X_n = k; \theta^{i-1}) = \frac{n_k}{N} \quad (5)$$

$$n_k = \sum_{n=1}^{N} p(X_n = k | Y_n; \theta^{i-1}) \quad (6)$$

$$e_k = \frac{1}{W} \sum_{n=1}^{N} p(X_n = k | Y_n; \theta^{i-1}) Y_n \quad (7)$$

where the value of W is a normalization constant to ensure that $\|e_k\| = 1$.

$$A_n = \frac{\left[\sum_{k=1}^{K} p(X_n=k|Y_n;\theta^{i-1})Y_n\right]^2}{\sum_{k=1}^{K} p(X_n=k|Y_n;\theta^{i-1})E[(Y_n - \|Y_n\|e_k)^T(Y_n - \|Y_n\|e_k)]} \quad (8)$$

The E and M steps are repeated until convergence is achieved. This can be
accomplished by monitoring the increment of the log-likelihood function.

4. EXPERIMENTAL RESULTS
The proposed image segmentation technique is evaluated with the help of a bergley's image data set. The comparison is carried out with image segmentation using standard EM , GA algorithm and with proposed FGA. The Error measures GCE,PRI are also given below.
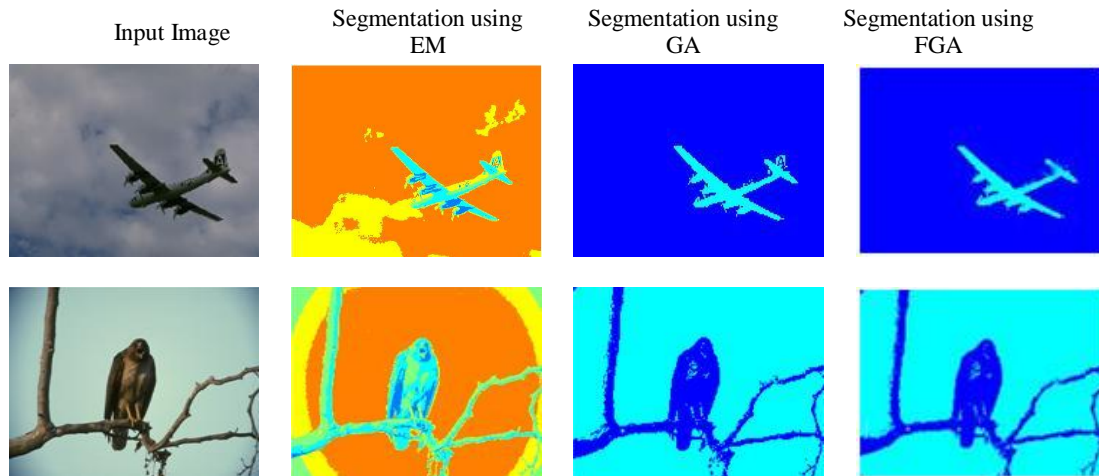
Input Image	Segmentation using EM	Segmentation using GA	Segmentation using FGA

Figure 4: Image Segmentation Comparison among various Techniques

The segmentation result by using various algorithms is provided in figure 4. From the figure it can be clearly observed that the FGA algorithm results in better visual clarity when compared to other segmentation algorithms.

The performance of image segmentation is measured by rand index measure and GCE(Global consistency Error. The following measurement is given for one sample image. The ground truth is given.

4.1 Global Consistency Error (GCE)
The first parameter used for evaluating the proposed segmentation technique is Global Consistency Error (GCE) [27]. This measure is a Region-based Segmentation Consistency which is computed to quantify the consistency among image segmentations of various granularities. The comparison is shown in table 1 and figure 5

Table 1: Comparison of Global Consistency Error

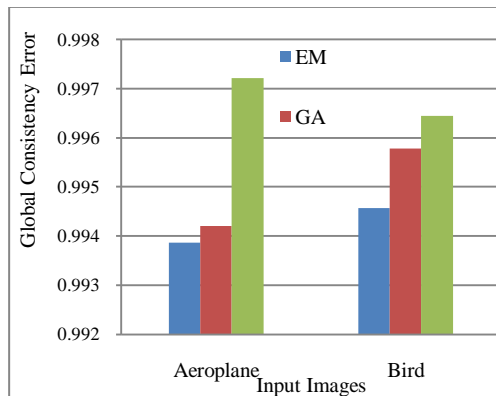| Image	Global Consistency Error		
	EM	GA	FGA
Aeroplane	0.99387	0.99421	0.99721
Bird	0.99457	0.99578	0.99645

Figure 5: Global Consistency Error

4.2 Probabilistic Rand Index (PRI)

Rand Index is the function that converts the problem of comparing two partitions with possibly differing number of classes into a problem of computing pair wise label relationships. PRI [27] counts the fraction of pairs of pixels whose labeling are consistent between the computed segmentation and the ground truth, averaging across multiple ground truth segmentations to account for scale variation in human perception. It is a measure that combines the desirable statistical properties of the Rand index with the ability to accommodate refinements appropriately. Since the latter property is relevant primarily when quantifying consistency of image segmentation results. The comparison is shown in table 2 and figure 6.

Table 2: Comparison of Probabilistic Rand Index

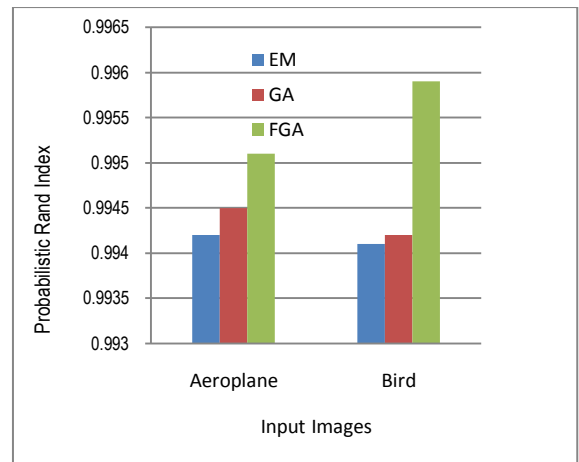| Image	Probabilistic Rand Index		
	EM	GA	FGA
Aeroplane	0.9942	0.9945	0.9951
Bird	0.9941	0.9942	0.9959

Figure 6: Probabilistic Rand Index   measure

4.3 Performance Measure using Time

The performance of FGA for semi supervised image segmentation is shown in the following graph. GA for image segmentation takes more time and this is overcome using FGA.
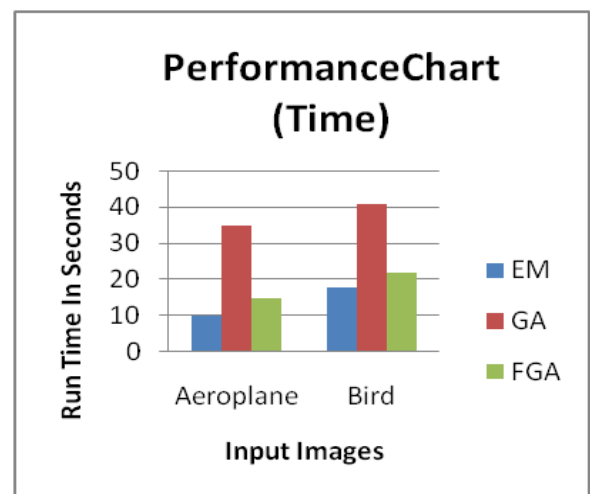

Figure 7: Time taken by various Image segmentation techniques.

The time comparison for various techniques for segmenting an image is provided in figure 7. From this it can be observed that the time required for using the proposed idea is much reduced when compared to GA and standard EM Algorithm for semi supervised image segmentation.

5. CONCLUSION

Image segmentation is widely necessaryn many fields like military, object detection, etc., This method can also be applied for satellite image for segmenting seashore and land or to find minerals near the seashore. In image segmentation, the input image is partitioned into several disjoint regions with similar characteristics such as intensity, color, and texture, shape etc., This separation will be helpful in identifying a particular object. There are several techniques already exists for image segmentation, but all those techniques have several demerits like lesser accuracy, more execution time, etc., For better visual,

Genetic Algorithm is used in which better partition is seen. But the execution time for GA is very large. This problem of image segmentation is solved using Fast Genetic algorithm. In future texture patterns or other constraints may be considered for segmentation.

6. REFERENCES

[1] Yi-hua Lan, Cun-hua Li, Yong Zhang, Xue-feng Zhao, "A Novel Image Segmentation Method Based on Random Walk", PACIIA 2009, Asia-Pacific Conference on Computational Intelligence and Industrial Applications, Vol. 1, pp. 207-210, 2009.

[2] Ye Hou, Bao-long Guo, Jeng-Shyang Pan, Chin-Shiuh Shieh, "Image Segmentation Based on GC-CV", HIS '09, Ninth International Conference on Hybrid Intelligent Systems, Vol. 1, pp. 252 – 256, 2009.

[3] Wenbing Tao, Hai Jin, Yimin Zhang, "Color Image Segmentation Based on Mean Shift and Normalized Cuts", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 37, No. 5, pp. 1382 – 1389, 2007.

[4] S. Belongie, "Color- and Texture-Based Image Segmentation Using EM and its Application to Content-Based Image Retrieval", Proc. of ICCV, Pp. 675-82, 1998.

[5] D.K. Panjwani and G. Healey, "Markov Random Field Models for Unsupervised Segmentation of Textured Color Images", PAMI, Vol. 17, No. 10, Pp. 939-54, 1995.

[6] J.P. Wang, "Stochastic Relaxation on Partitions with Connected Components and its Application to Image Segmentation", PAMI, Vol. 20, No.6, Pp. 619-36, 1998.

[7] S.C. Zhu and A. Yuille, "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation", P Figure 4: Image Segmentation Comparison among various Techniques AMI, Vol. 18, No. 9, Pp. 884-900.

[8] L. Shafarenko, M. Petrou, and J. Kittler, "Automatic Watershed Segmentation of Randomly Textured Color Images", IEEE Trans. on Image Processing, Vol. 6, No. 11, Pp. 1530-44, 1997.

[9] W.Y. Ma and B.S. Manjunath, "Edge flow: a Framework of Boundary Detection and Image Segmentation", Proc. of CVPR, Pp 744-49, 1997.

[10] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation", Proc. of CVPR, Pp. 731-37, 1997.

[11] M. Borsotti, P. Campadelli, and R. Schettini, "Quantitative Evaluation of Color Image Segmentation Results", Pattern Recognition letters, Vol. 19, No. 8, Pp. 741-48, 1998.

[12] D. Comaniciu and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation", Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Pp 750-755, 1997.

[13] Ayhan Demiriz, Kristin P. Bennett and Mark J. Embrechts, "Semi-Supervised Clustering Using Genetic Algorithms", 1999.

[14] U. Maulik, and S. Bandyopadhyay, "Genetic Algorithm-Based Clustering Technique", Pattern Recognition 33, 1999, Pp. 1455-1465.

[15] Yuntao Qian and Wenwu Si, "A Semi-Supervised Color Image Segmentation Method", IEEE International Conference on Image Processing, Vol. 2, Pp. II - 1194-7, 2005.

[16] Franz Pernkopf and Djamel Bouchaffra,"Genetic-Based EM Algorithm for Learning Gaussian Mixture Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 8, Pp. 1344-1348, 2005.

[17] Martínez-Usó, A., Pla, F and Sotoca, J.M., "A Semi-supervised Gaussian Mixture Model for Image Segmentation", 20th International Conference on Pattern Recognition, Pp. 2941-2944, 2010.

[18] Saha, S. and Bandyopadhyay, S., "MRI Brain Image Segmentation by Fuzzy Symmetry Based Genetic Clustering Technique", IEEE Congress on Evolutionary Computation, Pp. 4417-4424, 2007.

[19] Dong-Sun Kim, Member, IEEE, In-Ja Jeon, Seung-Yerl Lee, Phill-Kyu Rhee, and Duck-Jin Chung, Member, IEEE Embedded Face Recognition based on Fast Genetic Algorithm for intelligent Digital Photography

[20] L.Sankari and C.Chandrasekar,"Semi Supervised image segmentation using optimal color seed selection, IEEE 2011

[21] Dr. G.Padmavathi1, Dr. P.Subashini and A.Sumi, "Empirical Evaluation of Suitable Segmentation Algorithms for IR Images", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 2, 2010.

[22] A Yuntao Qian, Wenwu Si, School of Computer Science,Zhejiang University Hangzhou,310027, China,"Semi-supervised Color Image Segmentation Method",2005 IEEE