

Dynamic Image Segmentation using Fuzzy C-Means based Genetic Algorithm

Amiya Halder
Department of Computer
Science and Engineering
St. Thomas College of
Engineering and Technology
Kolkata-23, West Bengal

Soumajit Pramanik
Department of Computer
Science and Engineering
St. Thomas College of
Engineering and Technology
Kolkata-23, West Bengal

Arindam Kar
Department of Computer
Science and Engineering
St. Thomas College of
Engineering and Technology

ABSTRACT

This paper describes an evolutionary approach for unsupervised gray-scale image segmentation that segments an image into its constituent parts automatically. The aim of this algorithm is to produce precise segmentation of images using intensity information along with neighborhood relationships. In this paper, fuzzy c-means clustering helps in generating the population of Genetic algorithm which there by automatically segments the image. This technique is a powerful method for image segmentation and works for both single and multiple-feature data with spatial information. Validity index has been utilized for introducing a robust technique for finding the number of components in an image. Experimental results shown that the algorithm generates good quality segmented image.

Keywords

Clustering, Image Segmentation, Fuzzy C-means, Genetic Algorithm.

1. INTRODUCTION

Segmentation refers to the process of partitioning a digital image into multiple segments or regions. The goal of segmentation is to simplify the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries in images [1]. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. Image segmentation is a very important field in image analysis object recognition, image coding and medical imaging. Segmentation is very challenging because of the multiplicity of objects in an image and the large variation between them. Image segmentation is the process of division of the image into regions with similar attributes. In many object based image segmentation applications, the number of cluster is known a priori, but our proposed scheme is automatically determined the number of cluster which is produced the segmentation of images. The proposed technique should be able to provide good results whereas K-means algorithm which may get stuck at values which are not optimal [2]. Some of the several unsupervised clustering algorithms developed include K-means [3,4], ISODATA [5], self-organizing feature map (SOM) [6], Particle Swarm Optimization (PSO) [7], learning vector quantizers (LVQ) [8], GA based Clustering [9,10] etc.

One natural view of segmentation is that we are attempting to determine which components of a data set naturally "belong

together". Clustering is a process whereby a data set is replaced by clusters, which are collections of data points that "belong together". Thus, it is natural to think of image segmentation as image clustering i.e. the representation of an image in terms of clusters of pixels that "belong together". The specific criterion to be used depends on the application. Pixels may belong together because of the same color or similarity measure. The result of this algorithm produced a better result to compare with other techniques. Various segmentation techniques have been developed for image segmentation [11-14].

This paper presents in fact a possible extension of the previously presented concept of GA based segmentation [13,14] which was based on the GA based clustering of gray level images. In previous technique, the chromosome of GA is selected by randomly chosen in the image of the pixel. In this paper, fuzzy c-means clustering helps in generating the population of Genetic algorithm which there by automatically segments the image. This paper describes the concepts of optimal clustering technique which gives the better results compare to other state-of-the-art image segmentation methods [2,15,16] and existing segmentation techniques [11,13,17].

The rest of this paper is organized as follows: - Section 2, the concepts of clustering is provided. Section 3 gives the concepts of Genetic algorithm. Section 4 gives overview of FCM based Genetic Algorithm clustering approach. Section 5 describes the proposed algorithm and section 6 described the experimental results and section 7 concludes the paper.

2. CLUSTERING

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. By clustering, one can identify dense and sparse regions and therefore, discover overall distribution patterns and interesting correlations among data attributes.

Clustering may be found under different names in different contexts, such as unsupervised learning (in pattern recognition), numerical taxonomy (in biology, ecology), typology (in social sciences) and partition (in graph theory) [3]. By definition, "cluster analysis is the art of finding groups in data", or "clustering is the classification of similar objects into different groups, or more precisely, the partitioning of a data into subsets (clusters), so that the data in each subset (ideally) share some common trait-often proximity according to some defined distance measure" [4]. Clustering is a challenging field of

research as it can be used as a stand-alone tool to gain insight into the distribution of data, to observe the characteristics of each cluster, and to focus on a particular set of clusters for further analysis. Alternatively, cluster analysis serves as a pre-processing step for other algorithms, such as classification, which would then operate on detected clusters.

Hierarchical agglomerative clustering techniques start with as many clusters as there are unique values. Then pairs of cluster are successively merged till the optimal number of clusters is reached, depending on the termination condition. Termination condition is to be chosen carefully; else the hierarchical agglomerative clustering technique will ultimately yield one cluster containing all the values [18].

Clustering is a useful unsupervised data mining technique which partitions the input space into K regions depending on some similarity/dissimilarity metric where the value of K may or may not be known a priori. The main objective of any clustering technique is to produce a K×n partition matrix U(X) of the given data set X, consisting of n patterns, $X=(x_1, x_2, \dots, x_n)$ [19].

3. GENETIC ALGORITHM

Genetic Algorithm (GA) is a population-based stochastic search procedure to find exact or approximate solutions to optimization and search problems. Modeled on the mechanisms of evolution and natural genetics, genetic algorithms provide an alternative to traditional optimization techniques by using directed random searches to locate optimal solutions in multimodal landscapes [20]. Each chromosome in the population is a potential solution to the problem. Genetic Algorithm creates a sequence of populations for each successive generation by using a selection mechanism and uses operators such as crossover and mutation as principal search mechanisms - the aim of the algorithm being to optimize a given objective or fitness function. An encoding mechanism maps each potential solution to a chromosome. An objective function or fitness function is used to evaluate the ability of each chromosome to provide a satisfactory solution to the problem. The selection procedure, modeled on nature's survival-of-the-fittest mechanism, ensure that the fitter chromosomes have a greater number of offspring in the subsequent generations. For crossover, two chromosomes are randomly chosen from the population. Assuming the length of the chromosome to be *l*, this process randomly chooses a point between 1 and *l*-1 and swaps the content of the two chromosomes beyond the crossover point to obtain the offspring. A crossover between a pair of chromosomes is affected only if they satisfy the crossover probability.

Mutation is the second operator, after crossover, which is used for randomizing the search. Mutation involves altering the content of the chromosomes at a randomly selected position in the chromosome, after determining if the chromosome satisfies the mutation probability.

In order to terminate the execution of GA, a stopping criterion is specified. Specifying the number of iterations of the generational cycle is one common technique of achieving this end.

4. FCM BASED GENETIC ALGORITHM CLUSTERING

The searching capability of GAs can be used for the purpose of appropriately clustering a set of n unlabeled points in N-

dimension into K clusters [9]. In this proposed scheme, the same idea can be applied on image data. We consider a gray level image of size m×n. The basic steps of the GA-clustering algorithm for clustering image data are as follows:

4.1 Encoding

Each chromosome represents a solution which is a sequence of K cluster centers. For an N-dimensional space, each cluster center is mapped to N consecutive genes in the chromosome. For image datasets each gene is an integer representing an intensity value.

4.2 Population initialization

In Genetic Algorithm, the population size of P is needed. In this proposed method, the FCM is run P times for generating these P chromosomes; each chromosome is of size K. So, each of the P chromosomes is obtained by the FCM algorithm which is described below. Each chromosome of the population is a potential solution by FCM algorithm with no. of clusters $c=K$. The Fuzzy C-Means (FCM) algorithm assigns pixels to each category by using fuzzy memberships. Let $X=(x_1, x_2, \dots, x_N)$ denotes an image with N pixels to be partitioned into c clusters, where x_i represents multispectral (features) data. The algorithm is an iterative optimization that minimizes the cost function defined as follows:

$$J = \sum_{j=1}^N \sum_{i=1}^c u_{ij}^m \|x_j - z_i\|^2 \text{ ----- (1)}$$

Where u_{ij} represents the membership of pixel x_j in the i^{th} cluster, z_i is the i^{th} cluster center, $\| \cdot \|$ is a norm metric, and m is a constant. The parameter m controls the fuzziness of the resulting partition, and $m=2$ is used in this study. The cost function is minimized when pixels close to the centroid of their clusters are assigned high membership values, and low membership values are assigned to pixels with data far from the centroid. The membership function represents the probability that a pixel belongs to a specific cluster. In the FCM algorithm, the probability is dependent solely on the distance between the pixel and each individual cluster center in the feature domain. The membership functions and cluster centers are updated by the following:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_j - z_i\|}{\|x_j - z_k\|} \right)^{\frac{2}{m-1}}} \text{ ----- (2)}$$

and

$$z_i = \frac{\sum_{j=1}^N u_{ij}^m x_j}{\sum_{j=1}^N u_{ij}^m} \text{ ----- (3)}$$

Starting with an initial guess for each cluster center, the FCM converges to a solution for z_i representing the local minimum or a saddle point of the cost function. Convergence can be detected by comparing the changes in the membership function or the cluster center at two successive iteration steps.

4.3 Fitness computation

The fitness computation is accomplished in two steps. First, the pixel dataset is clustered according to the centers encoded in the chromosome under consideration, such that each intensity value x_i , $i = 1, 2, \dots, m \times n$ is assigned to cluster with center z_j , $j = 1, 2, \dots, K$.

$$\text{if } \|x_i - z_j\| < \|x_i - z_p\|, p = 1, 2, \dots, K, \text{ and } p \neq j \text{ --- (4)}$$

The next step involves adjusting the values of the cluster centers encoded in the chromosome, replacing them by the mean points of the respective clusters. The new center z_i^* for the cluster C_i is given by

$$z_i^* = \frac{1}{n_i} \sum_{x_j \in C_i} x_j, i = 1, 2, \dots, K \text{ --- (5)}$$

Subsequently, the clustering metric M is computed as the sum of Euclidean distances of each point from their respective cluster centers given by

$$M = \sum_{i=1}^k M_i, i = 1, 2, \dots, K \text{ --- (6)}$$

and

$$M = \sum_{x_j \in C_i} \|x_j - z_i\| \text{ --- (7)}$$

The fitness function is defined as

$$f = \frac{1}{M} \text{ --- (8)}$$

A low value of intra-cluster spread is a characteristic of efficient clustering. Hence our objective is to minimize the clustering metric M i.e. maximize f .

4.4 Selection

This fitness level is used to associate a probability of selection with each individual chromosome. We apply Roulette Wheel selection, a proportional selection algorithm where the no. of copies of a chromosome, that go into the mating pool for subsequent operations, is proportional to its fitness. If f_i is the fitness of individual C_i in the population, its probability of being selected is,

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j} \text{ --- (9)}$$

Where N is the number of individuals in the population.

4.5 Crossover

In this paper, a single-point crossover with a fixed crossover probability of μ_c is used. The procedure followed is the same as that described in section 3.

4.6 Mutation

Each chromosome undergoes mutation with a fixed probability μ_m . A number δ in the range $[0, 1]$ is generated with uniform distribution. If the value at a gene position is v , after mutation it becomes

$$v \pm \delta * v, v \neq 0$$

$$v \pm \delta, v = 0$$

4.7 Termination criterion

We execute the processes of fitness computation, selection, crossover, and mutation for a predetermined number of iterations. In every generational cycle, the fittest chromosome till the last generation is preserved - elitism. Thus on termination, this chromosome gives us the best solution encountered during the search.

The algorithm is a two pass process in each iteration. In the first pass standard FCM algorithm is used to generate the population. In the second pass, Genetic algorithm is applied on the population generated by the FCM algorithm. Then cluster validity index for the fittest chromosome for particular value of K is computed using equation (10), in order to determine the validity of the clustering on the given dataset and the fittest chromosome for that K value is preserved. GA is performed for $K=2$ to K_{max} .

5. Validity Index

The cluster validity measure used in the paper is the one proposed by Turi [11]. It aims at minimizing the validity index given by the function,

$$V = y \times \frac{\text{intra}}{\text{inter}} \text{ --- (10)}$$

The term *intra* is the average of all the distances between each pixel x and its cluster centroid z_i which is defined as

$$\text{intra} = \frac{1}{N} \sum_{i=1}^K \sum_{x \in C_i} \|x - z_i\|^2 \text{ --- (11)}$$

This term is used to measure the compactness of the clusters. The *inter* term is the minimum distances between the cluster centroids which is defined as

$$\text{inter} = \min(\|z_i - z_j\|^2), \text{ --- (12)}$$

Where $i=1, 2, \dots, K-1$, $j=i+1, 2, \dots, K$. This term is used to measure the separation of the clusters. Lastly, y is given as

$$y = c \times N(2,1) + 1 \text{ --- (13)}$$

Where c is a user specified parameter and $N(2,1)$ is a Gaussian distribution function with mean 2 and standard deviation 1, where the variable is the cluster number and is given as

$$N(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(k-\mu)^2}{2\sigma^2}} \text{-----(14)}$$

Where k is the cluster number and $\mu=2$ and $\sigma=1$ as per Turi's thesis on clustering [11]. This validity measure serves the dual purpose of

- minimizing the intra-cluster spread, and
- maximizing the inter-cluster distance.

Moreover it overcomes the tendency to select a smaller number of clusters (2 or 3) as optimal, which is an inherent limitation of other validity measures such as the Davies-Bouldin index or Dunne's index.

6. EXPERIMENTAL RESULTS

The proposed algorithm has been applied to well known natural images such as Lena, mandrill and peppers etc. All the results have been reported in Table I. Column 2 of Table I represent the optimal range for the number of clusters for the images of Lena, mandrill and peppers has also been copied from [11] which is based on visual analysis by a group of ten expert people. These results have been compared to those of SOM, snob [11] and DCPSO [17]. The results tabulated here for each image is the mean of 10 simulations. Figure 2(e), (j) and (o) are shown the segmented output images using our proposed method. The result of Table I is shown better than SOM and snob and always find a solution within the optimal range. The performance of the proposed algorithm with dynamic GA [13] and DCPSO produced comparable segmentation results [17]. The assumptions used for the implementation of the proposed algorithm are given as follows. The value of the parameter, c,

for the validity index referred to from [11], is set to 25. The size of the population, P, is taken as 30, crossover rate, μ_c , as 0.9 and mutation rate, μ_m , as 0.01 [20]. The algorithm uses number of iterations as the terminating condition and it is set to 20. The value of K_{max} is empirically set for the individual images. The maximum intensity level for each picture element in the image is taken as 255.

Also we have been compared to the results of validity index with the state-of-the-art image segmentation methods to those of K-means [2], FCM [15], and FHNN [16] in Figure 1. We obtain the no. of optimal cluster from the proposed method and this same no. of clustering is used as an input of above state-of-the-art image segmentation methods. The result shows that the validity index of the proposed method is lowest than K-means, FCM, and FHNN. This technique gives the better results to compare with others methods.

Figure 2 is shown the segmented images using K-means, FCM, FHNN, Dynamic GA and Proposed Method.

7. CONCLUSIONS

This paper presented a new approach for unsupervised image segmentation algorithm based on clustering technique which determines the optimal clustering of an image dataset, with minimum user intervention. In this paper, that the user does not need to predict the optimal number of clusters, required to partition the dataset, in advance. Comparison of the experimental results with that of other clustering methods, show that the technique gives satisfactory results when applied on well known natural images. Moreover results of its use on images from other fields (MRI, Satellite Images) demonstrate its wide applicability.

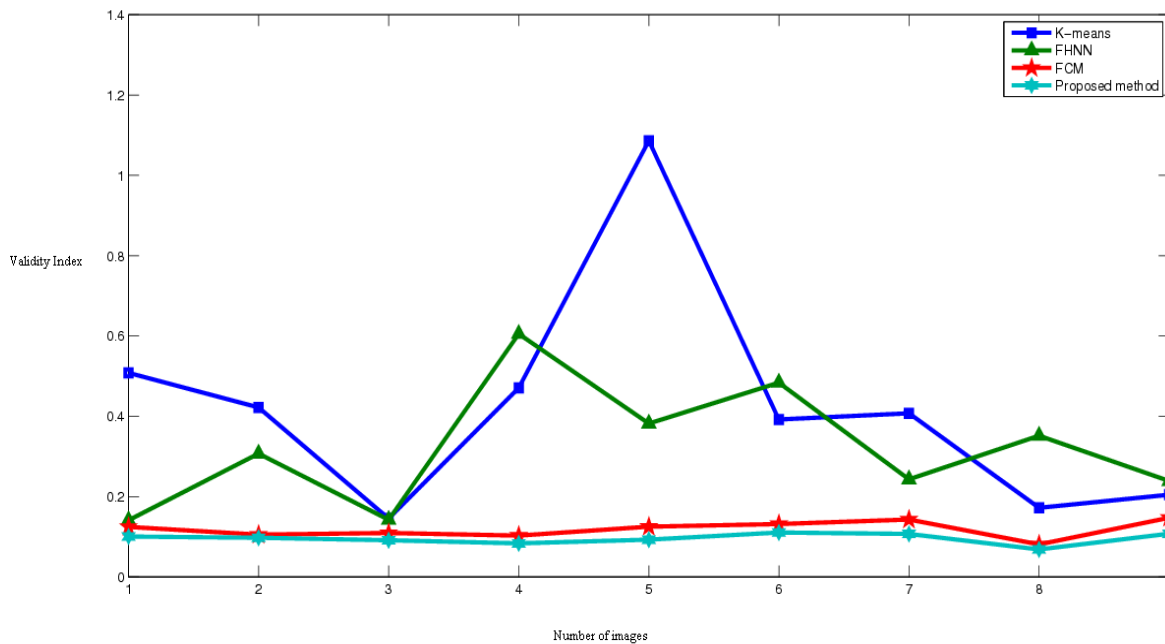


Figure 1: Experimental results (Validity index) on Natural images

Table 1. Experiments on Natural images (Number of cluster)

Image	Optimal Range	Proposed Method	SOM	snob	DCPSO using V	Dynamic GA
Lena	5 to 10	6.5	20	31	6.85	6.29
Mandrill	5 to 10	6.2	20	42	6.25	6.29
peppers	6 to 10	6.9	20	39	6.0	6.74
Jet	5 to 7	5.8	14	22	5.3	6.0
MRI	3 to 7	5.0	-	-	3.4	5.1
Kolkata	4 to 8	6.1	-	-	5.1	6.14
Cameraman	-	7.3	-	-	4.8	6.2
Pangong	-	6.1	-	-	5.1	5.2
Sunset	-	6.8	-	-	4.0	7.1

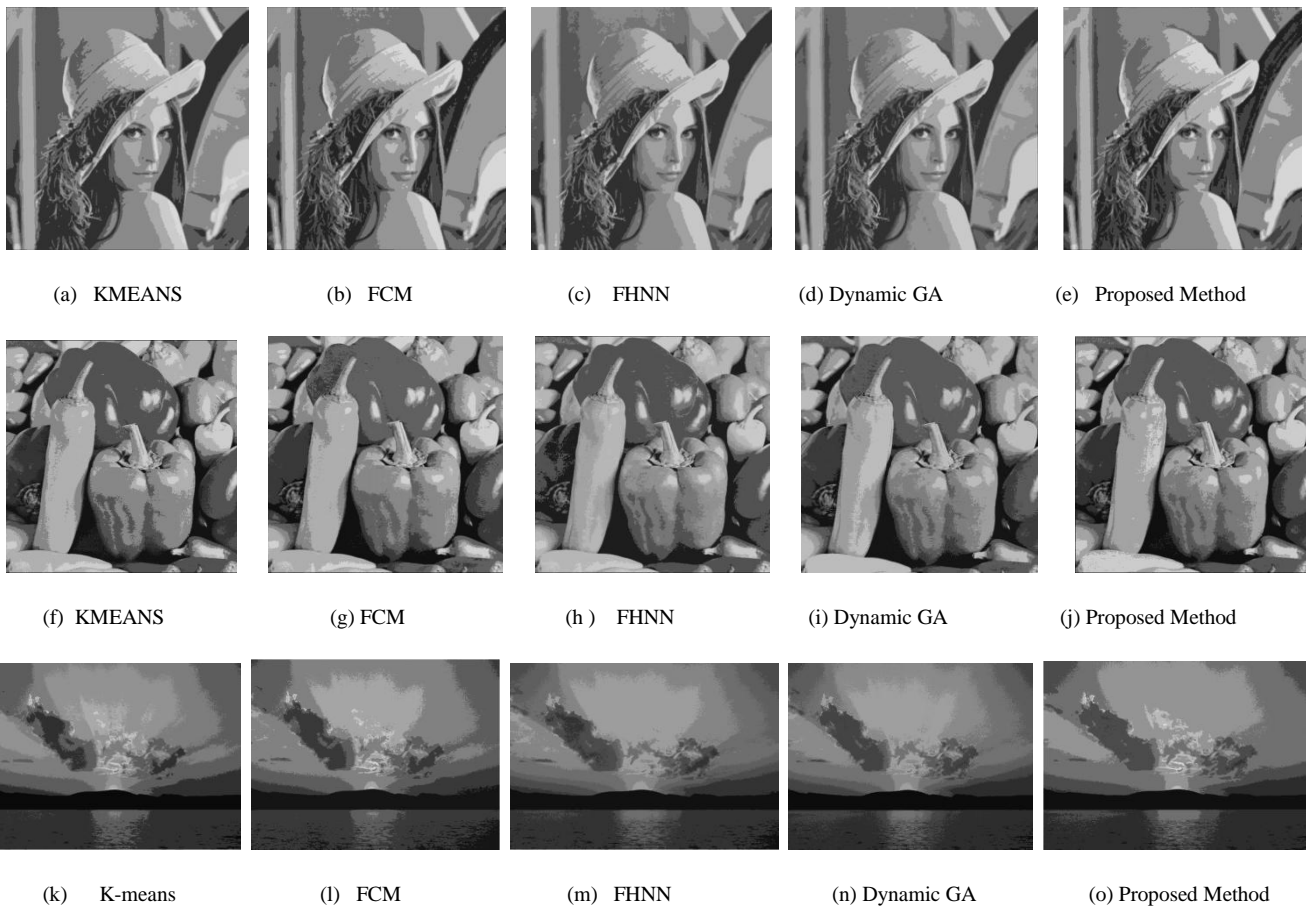


Figure 2: Segmented image using K-means, FCM, FHNN, GA based clustering and Proposed Method

8. REFERENCES

- [1] Rafael. C. Gonzalez, Richard. E. Woods, Digital Image Processing, Pearson Education, 2002.
- [2] S. Z. Selim, M. A. Ismail, K-means Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality, IEEE Trans. Pattern Anal. Mach. Intell. 6, (1984), 81-87.
- [3] E.Forgy, Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification, Biometrics,21: 1965.
- [4] J. A. Hartigan, Clustering Algorithms, John Wiley Sons, New York, 1975.
- [5] G. Ball, D. Hall, A Clustering Technique for Summarizing Multivariate Data, Behavioral Science, 12: 1967.

- [6] T. Kohonen, Self-Organizing Maps, Springer Series in Information Sciences, Vol 30, Springer-Verlag, 1995.
- [7] DW. Vander. Merwe, AP Engelbrecht, Data Clustering using Particle Swarm Optimization.
- [8] LV Fausett, Fundamentals of Neural Networks, Prentice Hall, 1994.
- [9] Ujjwal Maulik and Sangamitra Bandyopadhyay, Genetic Algorithm based clustering technique, Elsevier Science Ltd., 1999.
- [10] Hwe Jen Lin, Fu-Wen Yang and Yang-Ta Kao, An Efficient GA-based Clustering Technique, Tamkang Journal of Science and Engineering 8(2), 2005.
- [11] R.H.Turi, Clustering-Based Color Image Segmentation, PhD Thesis, Monash University, Australia, 2001.
- [12] Mofakharul Islam, John Yearwood and Peter Vamplew, Unsupervised Color Textured Image Segmentation Using Cluster Ensembles and MRF Model, Advances in Computer and Information Sciences and Engineering, 323-328, 2008.
- [13] Dipak Kumar Kole and Amiya Halder, An efficient dynamic Image Segmentation algorithm using Dynamik GA based clustering, International Journal of Logistics and Supply Chain Management, 2(1), pp. 17-20, 2010.
- [14] Amiya Halder and Nilavra Pathak, An Evolutionary Dynamic Clustering Based Colour Image Segmentation, International Journal of Image Processing (IJIP), Volume (4): Issue (6), pp. 549-556, 2011.
- [15] Keh-Shih Chuang , Hong-Long Tzeng , Sharon Chen, Jay Wu, Tzong-Jer Chen, Fuzzy c-means clustering with spatial information for image segmentation, Computerized Medical Imaging and Graphics Vol. 30, pp. 9–15 ,2006.
- [16] Jzau-Sheng Lin, Kuo-Sheng Cheng, Chi-Wu Mao, A Fuzzy Hopfield Neural Network for Medical Image Segmentation, IEEE Transactions on Nuclear Science, Vol. 43, No. 4, pp.2389-2398, August 1996.
- [17] Mahamed G. H. Omran, Andries P Engelbrecht and Ayed Salman, Dynamic Clustering using Particle Swarm Optimization with Application in Unsupervised Image Classification, PWASET 9:2005.
- [18] Arun K Pujari, Data Mining Techniques, Universities Press, 2003.
- [19] Indrajit Saha, Ujjwal Maulik and Sanghamitra Bandyopadhyay, An Improved Multi-objective Technique for Fuzzy Clustering with Application to IRS Image Segmentation, EvoWorkshops 2009, LNCS 5484, pp. 426–431, 2009.
- [20] M. Srinivas, Lalit M. Patnaik, Genetic Algorithms: A Survey.