Hierarchical Clustering Algorithm for Land Cover Mapping Using Satellite Images

J. Senthilnath, *Student Member, IEEE*, S. N. Omkar, V. Mani, Tejovanth N, P. G. Diwakar, and Archana Shenoy B

***Abstract*—This paper presents hierarchical clustering algorithms for land cover mapping problem using multi-spectral satellite images. In unsupervised techniques, the automatic generation of number of clusters and its centers for a huge database is not exploited to their full potential. Hence, a hierarchical clustering algorithm that uses splitting and merging techniques is proposed. Initially, the splitting method is used to search for the best possible number of clusters and its centers using Mean Shift Clustering (MSC), Niche Particle Swarm Optimization (NPSO) and Glowworm Swarm Optimization (GSO). Using these clusters and its centers, the merging method is used to group the data points based on a parametric method (k-means algorithm). A performance comparison of the proposed hierarchical clustering algorithms (MSC, NPSO and GSO) is presented using two typical multi-spectral satellite images – Landsat 7 thematic mapper and QuickBird. From the results obtained, we conclude that the proposed GSO based hierarchical clustering algorithm is more accurate and robust.**

***Index Terms*—Glowworm swarm optimization, mean shift clustering, niche particle swarm optimization.**

I. Introduction

IN LAND COVER mapping problem, we need actual information regarding the features of land to make good use of it. Using satellite images, we can accurately plan and use land efficiently. Satellite images offer a method of extracting this temporal data that can be used in gaining knowledge regarding land use. Recent advances in the realm of computer science have allowed us to perform this "intelligent" job. This has established a vast research area in solving the land cover mapping problem for city planning and land-usage [1].

Manuscript received September 27, 2011; revised December 28, 2011; accepted January 31, 2012. Date of publication March 19, 2012; date of current version June 28, 2012. A preliminary version of this paper was presented at the IGARSS 2011.

J. Senthilnath is with the Department of Aerospace Engineering, Indian Institute of Science, Bangalore 560012, India (e-mail: snrj@aero.iisc.ernet.in).

S. N. Omkar is with the Department of Aerospace Engineering, Indian Institute of Science, Bangalore 560012, India (corresponding author, e-mail: omkar@ aero.iisc.ernet.in).

V. Mani is with the Department of Aerospace Engineering, Indian Institute of Science, Bangalore 560012, India (e-mail: mani@aero.iisc.ernet.in).

Tejovanth N is with the Department of Electrical and Electronics Engineering, National Institute of Technology, Surathkal, India (e-mail: tejovanth.n@ieee.org).

P. G. Diwakar is with the Earth Observation System, ISRO Headquarters, Bangalore, India (e-mail: diwakar@isro.gov.in).

A. Shenoy B is with the Department of Electronics and Communication Engineering, National Institute of Technology, Surathkal, India (e-mail: archana.shenoyb@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2012.2187432

Unsupervised techniques can be used for grouping distinct land cover regions, provided there is a lack of ground truth information [2]. Based on certain similarity metric, the data is sub-divided into clusters [3], [4], using unsupervised methods where the number of clusters is not known a priori [5]. The objective is to maximize the inter-cluster distances while the intra-cluster distances are minimized. The clustering problems can be studied using hierarchical approach [6], by breaking a large cluster and merging smaller groups into their closest centroid [7]. Two approaches are used in this hierarchical clustering method: (i) divisive methods, where a large cluster is split into several small clusters; (ii) agglomerative methods, where many small clusters are merged to form a large cluster.

The grouping of the same clusters is regarded as a fundamental task in land cover mapping problem, which transforms the remotely sensed images to generate thematic land-use/land-cover maps [5]. Several methods to compute a single-band gradient function from satellite images have been studied previously by Tarabalka *et al.* including pixel-wise classification methods [4], [8]. Studies show that hierarchical step-wise optimization and spectral clustering have given good results for analyses of satellite images [9]. And these results have been improved by a combination of probabilistic classification and hierarchical step-wise optimization algorithm [10].

In the literature different methods have been developed to cluster data sets by splitting and merging [6]. Broadly, they can be classified into parametric and non-parametric methods. In parametric methods such as K-means clustering [11], prior assumptions of the number of clusters are made. This is essentially a function minimization technique, where the objective function is the squared error distance measure. In non-parametric methods such as Mean Shift Clustering (MSC) [12], [13], no prior assumptions are made on the number of clusters. This is a procedure for locating the maxima of a mapped function given a set of discrete data points sampled from that function. It is useful for detecting the modes of density given a density function. Conventionally, mean shift clustering uses single point for locating modes (local maxima).

Recently, researchers are interested in locating multiple local optima of a given multi-modal function in a d-dimensional search space. For this purpose nature inspired techniques are used. Brits *et al.* [14] developed Niche-Particle Swarm Optimization (NPSO) which is a variant of Particle Swarm Optimization (PSO) [15], Krishnand *et al.* [16], [17] developed Glowworm Swarm Optimization (GSO); these methods are useful for simultaneous capture of multiple local maxima/minima of a given function. It is observed in the literature that GSO is more efficient in finding local maxima/minima by comparing with other nature inspired techniques [17].

In our hierarchical clustering algorithm the non-parametric techniques – MSC [12], NPSO [14], and GSO [16] are used for the land cover mapping problem. These methods are used to split the complex large data set into a number of cluster centers by satisfying Bayesian Information Criteria (BIC), which is commonly used in model selection [18]. The cluster centers are used for merging the data set (agglomerative approach) to their respective group. The challenge here is in optimizing how the clusters are split and merged to group the data set to their respective classes. Two typical multi-spectral satellite images – Landsat thematic mapper of South India [19], and QuickBird [20] of the central Bangalore, India are used to analyze the parameters and the performance of the proposed hierarchical clustering algorithm.

II. Cluster Splitting and Merging

The cluster analysis forms the assignment of data set into clusters based on some similarity measure. In this study, hierarchical splitting (MSC, NPSO and GSO) and merging method (k-means) using clustering algorithm is applied. These hierarchical splitting techniques make use of kernel functions for locating maxima for a given set of discrete data points.

A. Splitting Methodology

Initially, for a given large data set, the number of clusters to be generated is difficult to predict, as this is data dependent. We require a platform from where we can pick an optimal number of cluster centers for a given data set. Bayesian Information Criteria (BIC) is a model fitting approach, which provides the optimal number of clusters. The splitting of data set using BIC into number of clusters is given by [18], [21]

$$BIC \approx L(\theta) - \frac{1}{2} * k_j * \log(n) \tag{1}$$

where $L(\theta)$ is the log-likelihood measure for the data set; k_j is the number of free parameters for the specific number of clusters; and n is the number of data points for a given data set.

1) Mean Shift Clustering: Let $x_1, x_2, \ldots, x_n$, where n is the number of data points in d-dimensional space. In the mean shift clustering, each data point is shifted to the average of the other data points in its neighborhood. This is done by using a Gaussian kernel, based on Euclidean distance between two data points (r) which is given by

$$K(r) = e^{-\|r\|^2}. \tag{2}$$

The main idea behind mean shift clustering is to treat the points in the d-dimensional feature space as a probability density function. The dense regions in the feature space correspond to the local maxima of the underlying distribution. We perform a gradient ascent procedure on the local estimated density until convergence is reached. The stationary points of this procedure represent the modes of the distribution [13].

The mean shift for point x_i is

$$m(x_i) = \frac{\sum_{j=1}^{n} x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{j=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} \tag{3}$$

where h is a bandwidth to limit the density within the kernel function and $g = -k'(x)$, x is the mean of all n data points. A complete discussion about various aspects of mean shift clustering and its applications are given in [12], [13].

2) Splitting the Data Set Using Mean Shift Clustering:

Step 1. Place the kernel on point x_i.

Step 2. Calculate the mean shift vector using (3).

Step 3. Kernel density estimation window is translated using $x_i(t+1) = x_i(t) + m(x_i)$.

Step 4. Repeat till the mean shift procedure find the stationary points and also by satisfying (1).

Here, care is taken by setting the parameter value of bandwidth (h) of the kernel function in MSC to split the data set into a maximum number of clusters by satisfying BIC.

3) Niche Particle Swarm Optimization (NPSO): The NPSO [14] is a variant of Particle Swarm Optimization (PSO) [15], based on a flock of birds aimed to locate multiple optima in a multi-modal problem. Each scattered particle 'i' is required to hold the following information about itself: (a) x, the current position, (b) v, the current velocity, and (c) y, the personal best position.

Each particle is updated using velocity as

$$v_{i,j}(t+1) = wv_{i,j}(t) + c_1 r_{1,j}(t)\left(y_{i,j}(t) - x_{i,j}(t)\right) \tag{4}$$

where w is the inertial weight ($0 < w < 1$) that determines how much of the old velocity the particle carries into the next iteration. $c_1 r_{1,j}$ specifies the weight of the random component with which the particle returns to its personal best position [14]. The objective function value of all particles x_i (cluster center) is calculated using the kernel function given in (2). If the objective function value is below a threshold ε_s is named 'sub-swarm leader'.

$$f(x_i) < \varepsilon_s. \tag{5}$$

The swarm is divided into several overlapping 'sub-swarms' in order to detect multiple peaks. Sub-swarm is created with all particles around the local center within a certain radius r_s. These particles are made to converge towards the local best position i.e., the sub-swarm leader's position $\hat{y}_{i,j}(t)$.

$$v_{i,j}(t+1) = wv_{i,j}(t) + \rho(t)(\hat{y}(t) - x_{i,j}(t)) \tag{6}$$

$$x_{i,j}(t+1) = -x_{i,j}(t) + v_{i,j}(t+1) \tag{7}$$

where $-x_{i,j}(t)$ resets the particle position to the local best ($\hat{y}_{i,j}(t)$) within the sub-swarm radius, $wv_{i,j}(t)$ is the search direction, and $\rho(t)$ is the region of better solution. The personal best position '$y_i(t)$' associated with a particle 'i' is the position that gave the highest fitness value for that particle. If 'f' denotes the objective function, then $y_i(t+1)$ is

$$y_i(t+1) = \begin{cases} y_i(t) & \text{if } f\left(x_i(t+1)\right) \geq f\left(y_i(t)\right) \\ x_i(t+1) & \text{if } f\left(x_i(t+1)\right) < f\left(y_i(t)\right) \end{cases}. \tag{8}$$

4) Splitting the Data Set Using NPSO:

Step 1. Initialization – Randomly choose N points as agents.

Step 2. Update velocity of every particle in the main swarm using (4).

Step 3. Find the sub-swarm leaders using (5).

Step 4. For each sub-swarm:

a. Determine the sub-swarm particles using a fixed radius from the swarm leader.

b. Reset the sub-swarm leader's postion using (6).

Step 5. Update positions of the entire particle set in the sub-swarm to move towards the sub-swarm leader using (7)

Step 6. Update the personal best position of all agents; (8).

Step 7. Repeat from Step 2 until stopping criteria are met.

There are four tunable parameters for NPSO: N is the number of particles, it is the iteration number, ρ is the weight with which the sub-swarm follows the leader and r_s is a sub-swarm radius. These parameters have been studied and it is observed that the ρ has the highest influence in picking the optimal value [17].

5) Glowworm Swarm Optimization (GSO): The GSO is a population based method to find the multiple optima of multimodal objective functions based on the foraging behavior of glowworms. In GSO algorithm, agents are randomly distributed in the search space of d-dimension. Agents are thought of as glowworms that carry a luminescence quality, called luciferin, that emit light proportional to this value. The local decision range (r_o) also varies with iteration bounded at the upper and lower end by the sensor range (r_s). This helps in capturing the multiple optima points. Each glowworm is attracted by the brighter glow of other neighboring glowworms within the decision range [16].

The GSO algorithm has three steps – the luciferin update phase, movement phase and the decision range update phase.

The *luciferin update* depends on the objective function value at the glow-worm position where each glow-worm adds a luciferin quantity proportional to the measured value objective function at that point with a proportional decay.

$$l_j(t+1) = \max\left\{0, (1-\rho)l_j(t) + \gamma J_j(t+1)\right\}. \quad (9)$$

Where ρ is the luciferin decay constant ($0 < \rho < 1$) and γ is the luciferin enhancement constant and $J_j(t)$ represents the value of the objective function at agent j's location at time t.

For a point x_i (cluster center) in the *movement phase*, first its neighbor is determined using the decision range r_d^i at the iteration i. Let the number of neighbors for point x_i be K. The objective function at all these K points are determined. Of all these K points, the points K' that have better (more) objective function value than the objective function value at point x_i is selected. Then the point x_i is chosen with one of the K' points based on a probabilistic approach (roulette wheel selection). Let x_j be the point in K' selected for movement. The point x_i is moved towards x_j given as

$$x_i(t+1) = x_i(t) + s\left(\frac{x_j(t) - x_i(t)}{\|x_j(t) - x_i(t)\|}\right) \quad (10)$$

where s is the step size. In order to detect multiple peaks, the *local decision range* and sensor range must be made a varying parameter. The function chosen to adaptively update the local decision domain range of each glowworm is given by

$$r_d^i(t+1) = \min\left\{r_s, \max\left\{0, r_d^i(t) + \lambda(n_t - |N_i(t)|\right\}\right\}. \quad (11)$$

Where λ is a constant parameter, r_s is a sensor range and n_t is used as a threshold parameter to control the number of neighbors. A detailed discussion on the proof of convergence and other issues in GSO algorithm are given in [16], [17].

6) Splitting the Data Set Using GSO:

Step 1. Initialization – Randomly choose N points as agents assign luciferin value l_0 and sensor range r_s for each agent.

Step 2. Luciferin update phase using (9).

Step 3. Move the glowworms using (10).

Step 4. Vary the local decision range using (11).

Step 5. Repeat Steps 2 to 5 till convergence is achieved.

There are four tunable parameters for the GSO algorithm: N is the number of agents, it is the number of iteration, s is the step size and r_s is the sensor range. Among these parameters, step size s has been observed to be effective in detecting the optimal value [17].

B. Merging Methodology

The cluster centers generated using MSC, NPSO and GSO is grouped using agglomerative approach. These cluster centers are used for initializing k-means to perform agglomerative clustering [21]. Here k-means algorithm (parametric method) is used to group the data points to closest cluster centers using similarity metric.

1) Merging Data Set Algorithm:

Step 1: Results obtained as cluster centers from MSC, NPSO and GSO is given to k-means clustering.

Step 2: Merge data points to the closest clusters.

Step 3: Use voting method for each data points in the cluster.

Step 4: Clusters are grouped agglomerative using labels.

Step 5: Assign each data point to one of the class.

We explain further steps 4 and 5 from the merging data set algorithm. To group a cluster we obtain class labels (ground truth information) for the data set and count the number of data points for each label belonging to a cluster. Then, that cluster is classified to a particular label, if the number of data points in that cluster is maximum. For example, consider a cluster of 10 data points in which 8 data points belong to class A and other data points belong to class B. In voting method, maximum score in the cluster is used for labeling. In this example, class A has maximum score and hence the cluster is labeled as class A. Thus, data points aid in labeling the cluster centers. Assignment of label to the cluster centers enables merging the clusters into their respective group in single step. All clusters generated using hierarchical clustering is labeled and then class label wise clusters are grouped into their respective classes.

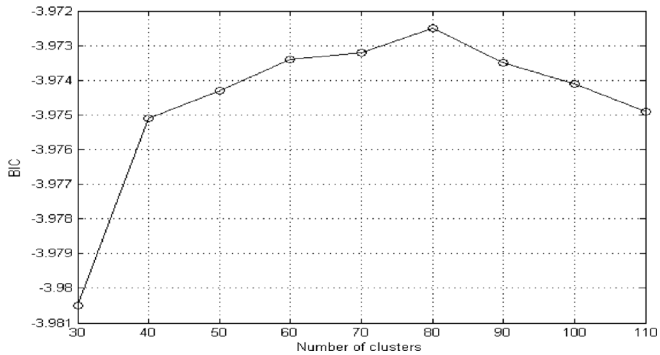

Fig. 1. BIC for Landsat image.

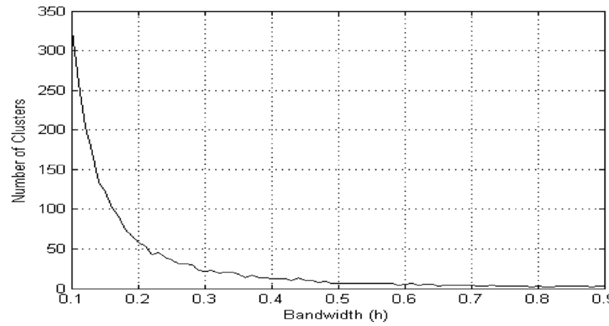

Fig. 2. Effect of bandwidth in mean shift clustering for Landsat data set.

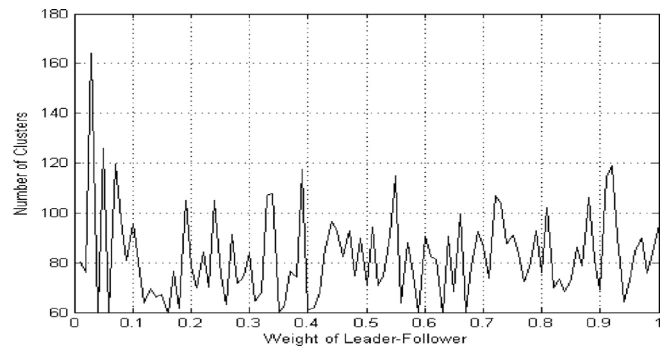

Fig. 3. Effect of weight of leader-follower in niche-particle swarm optimization for Landsat data set.

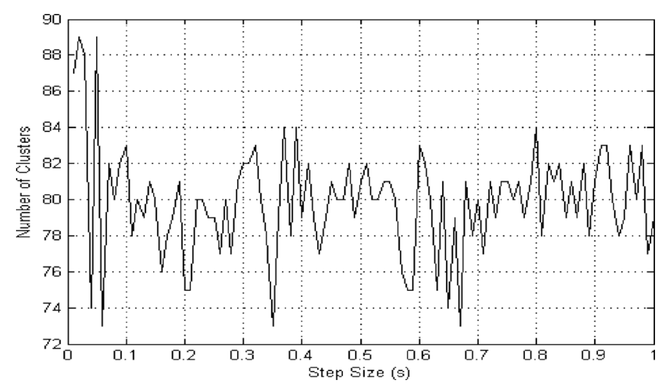

Fig. 4. Effect of step size in glowworm swarm optimization for Landsat data set.

III. Performance Measures

We evaluate the performance of hierarchical clustering methodology (split using MSC, NPSO, and GSO, and merge using K-means) with the classification matrix Q, of size $n_c \times n_c$ where n_c is the number of clusters. A typical entry q_{ij} in the classification matrix shows how many samples belonging to class i have been classified into class j [19]. For a perfect classifier, the classification matrix is diagonal. However due to misclassification, we get off-diagonal elements.

The performance measures considered are: individual class efficiency (η_i), average efficiency (η_a) and the overall efficiency (η_o). These are defined as

$$\eta_i = \frac{q_{ii}}{\sum_{j=1}^{n_c} q_{ji}} \quad \eta_a = \frac{1}{n_c}\sum_{i=1}^{n_c}\eta_i \quad \eta_o = \frac{1}{N}\sum_{i=1}^{n_c} q_{ii} \tag{12}$$

where q_{ii} is the number of correctly classified samples, n_c is the number of samples for the class c_i and N is the number of samples in the data set.

IV. Results and Discussion

In this section, we present the results obtained for two multispectral satellite images for land cover mapping problem. The first Landsat thematic mapper image contains 9 classes, of the southern region of India [19]. The second QuickBird four band image contains 6 classes of the central Bangalore, India [20].

A. Image 1 – Landsat Image

In this study, the Landsat image used is $15 \times 15.75\ \mathrm{km}^2$ (500×525 pixels) and has 30 *m* spatial resolution. The aim of the study is to develop an unsupervised technique to distinguish the 9 land cover regions using Landsat image. Senthilnath *et al.* [5] provides a detailed description of the data set, original and ground truth images. There are 9 level-II land cover region for this image which includes deciduous (C_1), deciduous-pine (C_2), pine (C_3), water (C_4), agriculture (C_5), bare ground (C_6), grass (C_7), urban (C_8) and shadow (C_9).

From Fig. 1, we can observe that maximum cluster centers generated based on BIC for this data set should be 80. The bandwidth in MSC, weight of leader-follower in NPSO and step size in GSO is set by limiting to 80 cluster centers. To classify by merging to the nearest cluster center is done using k-means technique.

The parameter values play an important role to generate 80 cluster centers for this data set based on BIC estimator. The parameter value for MSC is bandwidth $(h) = 0.18$; for NPSO (N, it, ρ, r_s) are (150, 500, 0.09, 0.1) and for GSO (N, it, s, γ, λ) are (150, 500, 0.05, 0.6, 0.01). The bandwidth study for MSC shown in Fig. 2, which shows a smooth curve and the number of clusters is 80 at 0.18. The weight of leader-follower (ρ) for NPSO was observed to be more dominant than other parameters. Fig. 3 gives the variation of number of clusters with varying ρ. The abrupt variation is caused due to the high degree of randomness involved. Similar characteristic is obtained when number of clusters is plotted against varying step size for GSO as shown in Fig. 4.

The algorithms were tested on a computer with Core-i7 processor, and 6 GB RAM, Matlab 7.12, and the execution time for each dataset was recorded. The time of execution for the

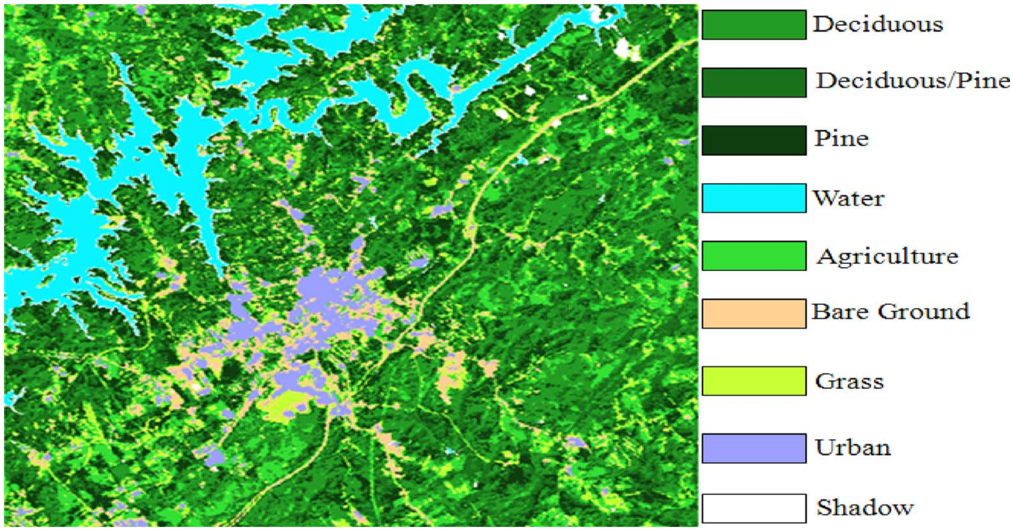

Fig. 5. Landsat image classification using GSO.

TABLE I
PERFORMANCE MEASURE FOR K-MEANS, MSC, NPSO AND GSO IN LANDSAT DATA SET GROUPING.

Classification Efficiency	K-means	MSC	NPSO	GSO
η_1	82.1	85.9	85.0	87.2
η_2	68	81.0	81.3	85.6
η_3	53.9	69.3	82.6	84.1
η_4	92.3	92.4	92.4	92.4
η_5	76.3	76.2	77.9	80.2
η_6	35.6	38.6	70.7	71.4
η_7	67.3	69.2	72.2	75.7
η_8	39.9	41.7	70.8	78.8
η_9	28.4	66.8	77.8	77.5
η_a	60.4	69.0	78.9	81.4
η_o	70.8	78.1	81.8	84.7

Landsat image for MSC is 5.0172 s, GSO took 205.1024 s and NPSO took 509.3871 s.

From Table I, we can observe that the performance measure of individual efficiency using GSO based hierarchical clustering and classification is better in comparison with that of NPSO, whereas NPSO is better than the traditional clustering technique – MSC and k-means for Landsat data set. For the samples belonging to water (C_4) the methods has 92% individual efficiency. In the literature, researchers have observed that water region in multi-spectral satellite image provides better grouping than other land cover regions [19], [20]. The overall efficiency of the GSO based hierarchical clustering is better in comparison with that of k-means, MSC and NPSO technique. Fig. 5, shows the classification result obtained for Landsat image using GSO based hierarchical clustering and classification.

B. *Image 2 – Quickbird Image*

The study area lies in the central region around M.G. Road of Bangalore city in Karnataka, India. QuickBird multi-spectral image used is *3.2977* $\times$ *2.4337* km^2 *(1375* $\times$ *1017 pixels)* with the resolution of *2.4 m*. In Image 2 a Level-I hierarchical clustering methods was carried out by dividing the image into six classes: vegetation (D_1), water (D_2), Road (D_3), Urban (D_4), Shadow (D_5) and Barren/Waste land (D_6). The details of the land cover region and ground truth information are given in Omkar *et al.* [20].

Similarly, using BIC for this data set we observed the maximum cluster centers generated should be 60. Here also, care has to be taken to set the bandwidth in MSC, swarm radius in NPSO and step size in GSO to generate 60 cluster centers. To classify by merging to the nearest cluster center is done using k-means technique. The parameter value for MSC is bandwidth $(h) = 0.14$; for NPSO (N, it, ρ, r_s) are (100, 1000, 0.12, 0.5) and for GSO ($N, it, s, \gamma, \lambda$) are (100, 1000, 0.05, 0.6, 0.01). As observed earlier, MSC was the fastest with the time of execution for the Quickbird image being 8.1938 s, GSO took longer at 493.2918 s and NPSO was 2019.1028 s.

From Table II, we can observe that for this data set also the performance measure of individual efficiency using GSO based hierarchical clustering is better in comparison with that of k-means, MSC and NPSO. The individual efficiency of samples belonging to road (D_3) in all the methods is not promising. In the literature, researchers have used extracting road (D_3) from satellite image as a separate problem [22]. The overall efficiency of the GSO based hierarchical clustering is better than k-means, MSC and NPSO based hierarchical clustering. From the performance of classification efficiency we can infer that the NPSO is a better clustering technique when compared to k-means and MSC whereas GSO is better in comparison with that of NPSO. Fig. 6, shows the classification result obtained

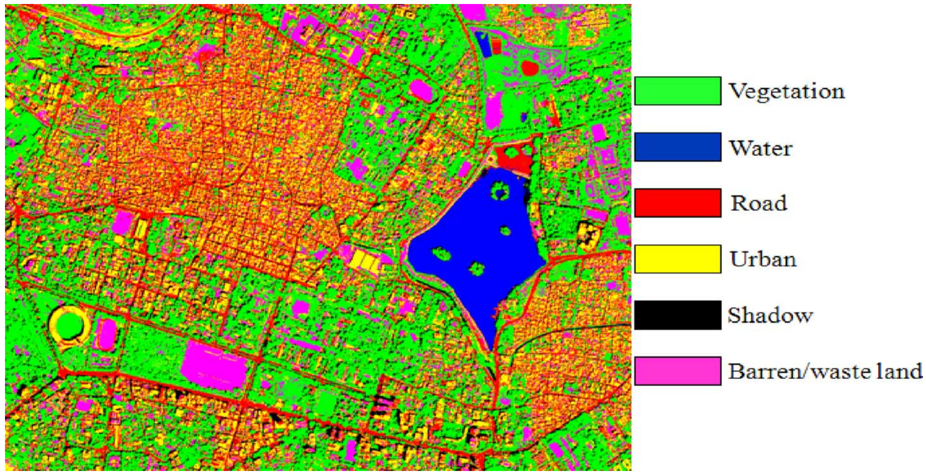

Fig. 6. QuickBird image classification using GSO.

TABLE II
PERFORMANCE MEASURE FOR K-MEANS, MSC, NPSO AND GSO IN QUICKBIRD DATA SET GROUPING.

Classification Efficiency	K-means	MSC	NPSO	GSO
η_1	89.4	89.8	91.7	93.6
η_2	85	86.4	92.8	93.4
η_3	40.4	41.4	68.9	76.9
η_4	78.9	90.7	89	88.1
η_5	70.5	72.9	71.9	94
η_6	71.7	77.5	84.7	87.5
η_a	72.7	76.5	83.2	88.9
η_o	76.4	80.8	85	89.6

for QuickBird image using GSO based hierarchical clustering and classification.

V. CONCLUSION

In this paper, we have presented the hierarchical clustering methods for land cover mapping problem. The satellite images used for this study are Landsat and QuickBird. The hierarchical technique adopts MSC, NPSO and GSO algorithm for splitting the data set by satisfying BIC and k-means algorithm is used to merge the data set. The randomness in GSO and NPSO hinders picking proper parameters for the clustering problem, unlike MSC where number of clusters is predictable. The performance is measured using classification efficiency—individual, average and overall efficiency of the three algorithms using Landsat and Quickbird images. We observe that though computationally GSO is slower than MSC, the former algorithm gives much better classification results. The NPSO on the other hand, is slower than GSO, and is also less efficient.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their comments which were useful in the revision of this paper.

REFERENCES

[1] L. David, "Hyperspectral image data analysis as a high dimensional signal processing problem," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, 2002.

[2] S. Bandyopadhyay and U. Maulik, "Genetic clustering for automatic evolution of clusters and application to image classification," *Pattern Recognit.*, vol. 35, no. 6, pp. 1197–1208, 2002.

[3] J. Senthilnath, S. N. Omkar, and V. Mani, "Clustering using firefly algorithm: Performance study," *Swarm and Evolutionary Computation*, vol. 1, no. 3, pp. 164–171, 2011.

[4] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, 2009.

[5] J. Senthilnath, S. N. Omkar, V. Mani, N. Tejovanth, P. G. Diwakar, and S. B. Archana, "Multi-spectral satellite image classification using glowworm swarm optimization," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, 2011, pp. 47–50.

[6] A. K. Jain and R. C. Dubes*, Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[7] C. Ding and X. He, "Cluster merging and splitting in hierarchical clustering algorithms," in *Proc. IEEE ICDM*, 2002, pp. 1–8.

[8] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, 2010, DOI: 10.1016/j.patcog.2010.01.016.

[9] Y. Tarabalka, J. C. Tilton, J. A. Benediktsson, and J. Chanussot, "A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2012, DOI: 10.1109/JSTARS.2011.2173466, to be published.

[10] Y. Tarabalka and J. C. Tilton, "Spectral-spatial classification of hyperspectral images using hierarchical optimization," in *Proc. Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2011, pp. 1–4.

[11] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp.*, 1967, pp. 281–297.

[12] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[13] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.

[14] R. Brits, A. P. Engelbrecht, and F. van den Bergh, "A niching particle swarm optimizer," in *Proc. 4th Asia-Pacific Conf. Simulated Evolution and Learning*, 2002, pp. 692–696.

[15] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Networks*, 1995, pp. 1942–1948.

[16] K. N. Krishnanand and D. Ghose, "Detection of multiple source locations using a glowworm metaphor with applications to collective robotics," in *Proc. IEEE Swarm Intelligence Symp. (SISS2005)*, Pasadena, CA, 2005, pp. 84–91.

[17] K. N. Krishnanand and D. Ghose, "Theoretical foundations for rendezvous of glowworm-inspired agent swarms at multiple locations," *Robotics and Autonomous Systems*, vol. 56, no. 7, pp. 549–569, 2008.
[18] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
[19] S. Suresh, V. Mani, S. N. Omkar, and N. Sundararajan, "Multispectral satellite image classification using an evolving neural network approach," *J. Aerosp. Sci. Technol. Aero. Soc. India*, vol. 58, no. 4, pp. 287–304, 2006.
[20] S. N. Omkar, K. M. Manoj, D. Mudigere, and D. Muley, "Urban satellite image classification using biologically inspired techniques," in *Proc. IEEE Int. Symp. Industrial Electronics*, 2007, pp. 1767–1772.
[21] H. Li, K. Zhang, and T. Jiang, "The regularized EM algorithm," in *Proc. 20th Nat. Conf. Artificial Intelligence*, 2005, pp. 807–812.
[22] J. Senthilnath, M. Rajeswari, and S. N. Omkar, "Automatic road extraction using high resolution satellite image based on texture progressive analysis and normalized cut method," *J. Indian Soc. Remote Sens.*, vol. 37, no. 3, pp. 351–361, 2009.

J. Senthilnath is a research scholar at the Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India. His research interests include nature inspired computational techniques, satellite image processing, machine learning and computer vision.

S. N. Omkar is a Principal Research Scientist at the Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India. His research interests includes nature inspired computational techniques, satellite image processing and parallel computing

V. Mani is a Professor at the Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India. His research interests include distributed computing, queuing networks, evolutionary computing, and neural computing.

Tejovanth N is currently pursuing his undergraduate degree in electrical and electronics at the National Institute of Technology, Karnataka, India.

P. G. Diwakar is currently Director of Earth Observation System, ISRO, Bangalore, India.

Archana Shenoy B is currently pursuing her undergraduate degree in electronics and communication engineering at National Institute of Technology, Karnataka, India.