Automatic Fuzzy Clustering Using Modified Differential Evolution for Image Classification

Ujjwal Maulik, Senior Member, IEEE, and Indrajit Saha

Abstract—The problem of classifying an image into different homogeneous regions is viewed as the task of clustering the pixels in the intensity space. In particular, satellite images contain landcover types, some of which cover significantly large areas while some (e.g., bridges and roads) occupy relatively much smaller regions. Automatically detecting regions or clusters of such widely varying sizes is a challenging task. In this paper, a new real-coded modified differential evolution based automatic fuzzy clustering algorithm is proposed which automatically evolves the number of clusters as well as the proper partitioning from a data set. Here, the assignment of points to different clusters is done based on a Xie–Beni index where the Euclidean distance is taken into consideration. The effectiveness of the proposed technique is first demonstrated for two numeric remote sensing data described in terms of feature vectors and then in identifying different landcover regions in remote sensing imagery. The superiority of the new method is demonstrated by comparing it with other existing techniques like automatic clustering using improved differential evolution, classical differential evolution based automatic fuzzy clustering, variable length genetic algorithm based fuzzy clustering, and well known fuzzy C-means algorithm both qualitatively and quantitatively.

Index Terms—Differential evolution (DE), fuzzy clustering, genetic algorithm, remote sensing imagery, unsupervised classification.

I. INTRODUCTION

REMOTE sensing satellite images have significant applications in different areas such as climate studies, assessment of forest resources, examination of marine environments, etc. An important task in remote sensing applications is the classification of pixels in the images into homogeneous regions, each of which corresponds to some particular landcover type. This problem has often been modeled as a clustering problem [1]. However, since it is difficult to have an a priori information about the number of clusters in satellite images, the clustering algorithms should be able to automatically determine this value. Moreover, in satellite images it is often the case that some regions occupy only a few pixels while the neighboring regions are significantly large. Thus, automatically detecting regions or clusters of such widely varying sizes presents a challenge in designing clustering algorithms and validity indices. Considerable research has been done in the recent past to evolve the naturally occurring clusters in a complex data set. However, little work has been done to determine the optimal number of clusters at the same time for remote sensing satellite images.

Clustering [2] is a useful unsupervised learning technique that 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 $K \times n$ partition matrix $U(X)$ of the given data set $X$, consisting of $n$ patterns, $X = \{x_1, x_2, \ldots, x_n\}$. The partition matrix may be represented as $U = \{u_{k,j}\}, k = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, n$, where $u_{k,j}$ is the membership of pattern $x_j$ to the $k$th cluster. Now $u_{k,j}$ is used to define the crisp (or hard) and fuzzy (or soft) clustering. For crisp clustering of the data, $u_{k,j}$ will be 0 or 1, meaning that the clusters are disjoint and non-overlapping in nature. Any pattern may belong to only one class in this case. However, in case of fuzzy clustering of the data, $0 \leq u_{k,j} \leq 1$, i.e., $u_{k,j}$ denotes the degree of belongingness of pattern $x_j$ to the $k$th cluster. It indicates that a pattern may belong to all the classes with different fuzzy membership grades. Popular crisp clustering approaches [3], [4] do not consider the overlapping of classes that occur in many practical image segmentation problems. For example, in remote sensing satellite images, a pixel corresponds to an area of the land space, which may not necessarily belong to a single type of land cover. This in turn indicates that the pixels in a satellite image can be associated with a large amount of imprecision and uncertainty. Therefore, applying the principles of the fuzzy set theory appears to be natural and appropriate in such domains.

In 1995, a new floating point encoded an evolutionary algorithm for global optimization, called Differential Evolution (DE) [5], [6] was proposed. Differential Evolution uses a special kind of differential operator, and recently, it has been applied in different fields of engineering and science [7], [8]. In [7], Omran et al. devised a non-automatic crisp clustering scheme based on DE and illustrated the application of the algorithm to image segmentation problems. Das et al. proposed automatic crisp clustering using improved DE and applied that for image segmentation in [8]. In this article, we have provided a new version of DE algorithm, called modified differential evolution (MoDE), and applied it to the automatic fuzzy clustering problem. The modified variant of differential evolution differs from the classical differential evolution in the process of mutation. While doing the mutation, it uses three vectors: one representing the local best, the second representing the global best which are adaptive in nature, and the third one selected...
randomly. This modified mutation process is governed by the adaptive parameter $\alpha$. It helps in increasing the exploration and exploitation capabilities of MoDE when solving optimization problems. The objective of this paper is twofold. First, it aims at the automatic determination of the optimal number of clusters using differential evolution based automatic fuzzy clustering (DEAFC). Second, it attempts to show that differential evolution (DE), with a modification scheme name as modified differential evolution based automatic fuzzy clustering (MoDEAFC), can provide very promising results if applied to the clustering problem. The technique uses a masker along with the initial population of DE or MoDE, which contains 0’s and 1’s. The value 1 in the masker cell indicates that the encoded center in the same position of the vector is valid; otherwise, it is not. The automatic fuzzy clustering scheme uses the Xie–Beni index as in [9]. For comparison, the automatic crisp clustering using improved DE [8] has been extended in the framework of fuzzy clustering by measuring the fuzzy validity index at the time of fitness computation.

Clustering results are reported for two numeric remote sensing data. The efficiency of the proposed algorithm has also been verified in image classification. Indian remote sensing (IRS) satellite images of parts of the cities of Calcutta and Mumbai have been used for demonstrating the effectiveness of the developed modified differential evolution automatic based fuzzy clustering (MoDEAFC) technique for classifying the images into an unknown number of regions. From the ground truth available for the images, the effectiveness of the method in automatically identifying the different landcover types present in the images has been verified. The superiority of the proposed technique has been demonstrated over automatic fuzzy clustering using improved DE [8] has been extended in the framework of fuzzy clustering. The automatic fuzzy clustering scheme uses the Xie–Beni index [12], proposed in [9] is defined as a function of the ratio of the total variation $\sigma$ to the minimum separation $sep$ of the clusters. Here, $\sigma$ and $sep$ can be written as

$$\sigma(U, ZX) = \sum_{k=1}^{K} \sum_{i=1}^{n} u_{k,i}^2 D^2(z_k, x_i)$$

(4)

$$sep(Z) = \min_{k \neq j} \|z_k - z_j\|^2$$

(5)

where $\|\cdot\|$ is the Euclidean norm. The XB index is then written as

$$XB(U, ZX) = \frac{\sigma(U, ZX)}{n \times sep(Z)}.$$  

(6)

Note that when the partitioning is compact and good, the value of $\sigma$ should be low while the value of $sep$ should be high, thereby yielding lower values of the XB index. The objective is therefore to minimize the XB index for achieving proper clustering.

C. $I$ Index

A cluster validity index $I$, proposed in [13] is defined as follows:

$$I(K) = \left( \frac{1}{K} \times \frac{E_1}{E_K} \times D_K \right)^p$$

(7)

where $K$ is the number of clusters. Here

$$E_K = \sum_{k=1}^{K} \sum_{i=1}^{n} u_{k,i} \|z_k - x_i\|$$

(8)

$$D_K = \max_{k \neq j} \|z_k - z_j\|.$$  

(9)

In this article, we have taken $p = 2$. A larger value of $I$ index implies better solution. Note that for computing the
III. PROPOSED DIFFERENTIAL EVOLUTION BASED
AUTOMATIC FUZZY CLUSTERING

A. Differential Evolution Algorithm and Its Modification

1) Classical Differential Evolution Algorithm: Differential Evolution is a relatively recent heuristic designed to optimize problems over continuous domains. In DE, each decision variable is represented in the vector by a real number. As in any other evolutionary algorithm, the initial population of DE is generated randomly, and then evaluated. The kth individual vector of the population at time-step (generation) t has d components (dimensions), i.e.,

\[ G_k(t) = [G_{k,1}(t), G_{k,2}(t), \ldots, G_{k,d}(t)] \]  

(10)

For each target vector \( G_k(t) \) that belongs to the current population, three randomly selected vectors from the current population is used. In other words, the \( d \)th component of each trial vector is generated as follows:

\[ \vartheta_{k,d}(t + 1) = G_{i,d}(t) + F (G_{n,d}(t) - G_{m,d}(t)) \]  

(11)

Here, \( F \) is a mutation factor. In order to increase the diversity of the perturbed parameter vectors, crossover is introduced. To this end, the trial vector

\[ Q_k(t + 1) = [Q_{k,1}(t + 1), Q_{k,2}(t + 1), \ldots, Q_{k,d}(t + 1)] \]  

(12)

is formed, where

\[ Q_{k,j}(t + 1) = \begin{cases} \vartheta_{k,j}(t + 1), & \text{if } \text{rand}_j(0, 1) \leq \text{CR} \\ G_{k,j}(t), & \text{if } \text{rand}_j(0, 1) > \text{CR} \\ \text{and } j \neq \text{rand}(k). \end{cases} \]  

(13)

In (13), \( \text{rand}_j(0, 1) \) is the \( j \)th evaluation of a uniform random number generator with outcome \( \in [0, 1] \). \( \text{CR} \) is the crossover rate \( \in [0, 1] \), which has to be determined by the user. \( \text{rand}(k) \) is a randomly chosen index \( \in \{1, 2, \ldots, d\} \), which ensures that \( Q_k(t + 1) \) gets at least one parameter from \( \vartheta_k(t + 1) \). The following condition decides whether it should become a member of the next generation (t + 1):

\[ G_k(t + 1) = \begin{cases} Q_k(t + 1), & \text{if } f(Q_k(t + 1)) > f(G_k) \\ G_k(t), & \text{if } f(Q_k(t + 1)) \leq f(G_k) \end{cases} \]  

(14)

where \( f(.) \) is the objective function to be minimized. The processes of mutation, crossover, and selection are executed for a fixed number of iterations. The best vector seen up to the last generation provides the solution to the clustering problem. Note that classical DE algorithm is also used for finding the optimal number of clusters in the DEAFC algorithm, whereas representation of vector and masker for the initial population and the fitness computation are similar with the MoDEAFC algorithm.

2) Modified Differential Evolution: In Modified Differential Evolution (MoDE), an approach has been introduced during mutation to push the trial vector quickly toward the global optima. In the mutation process, which deals with three vectors, two of them represent the global best (GBest) and the local best (LBest), respectively

\[ \vartheta_i(t + 1) = G_{\text{GBest}}(t) + \alpha (G_{\text{LBest}}(t) - G_k(t)) \]  

(15)

The crossover and selection are the same as the original DE. In each generation, one of the process is selected depending on \( \alpha \), computed as \( 1/[1 + \exp(-(1/generation))] \). While one process uses the mutation as in the original DE, the other generates a new mutant vector using (15). Note that as the number of generation increases, the value of \( \alpha \) decreases in the range between \([1, 0.5]\), which results in a lower probability of using modiMutation. Also, when modiMutation is used in the initial stage, LBest (best vector in current population) has more contributions for evolving the mutant vector than in the later stage. As the contribution of LBest for the mutant vector decreases with generation, the contribution of GBest (best vector evaluated till the current generation) increases. This fact provides more exploration and exploitation with the knowledge of local and global best. This MoDE is extended for automatic fuzzy clustering, called MoDEAFC, with the concept of representation of vector and masker for the initial population.

B. Automatic Fuzzy Clustering Algorithm Based on DE and MoDE

1) Vector and Masker Representation for Automatic Fuzzy Clustering: Here, for \( n \) data points, in \( d \) dimension, and for a user-specified maximum number of clusters \( K_{\text{max}} \), a vector is represented as a real number of length \( l = K_{\text{max}} \times d \), whereas a masker has \( K_{\text{max}} \) entries of binary numbers 0 and 1, which indicates that the corresponding cluster is either valid (i.e., to be really used for classifying the data) or not. Note that the number of maskers used is the same as the size of the population. Each generation masker is updated with random binary values except the best one of the previous generation. Fig. 1 shows a vector that contains six cluster centers in 3-D space and a randomly generated masker of size that is the same as the number of cluster centers. The valid centers have circled in Fig. 1, where the value of the corresponding masker cell is 1. The rest of the centers in that vector are not participated during fuzzy clustering.

2) Fitness Function: Once the encoded centers of the vectors are activated by the masker, they are used for fitness computation. Let \( \{z_1, z_2, \ldots, z_K\} \) be the set of \( K \) valid cluster centers encoded in a vector. The fitness of that vector (XB) is computed using (6). Subsequently, the centers activated in a vector are updated using (3). The objective is therefore to minimize the XB index for achieving proper clustering. Accordingly, the fitness of the vectors in each generation are computed. Note that if the number of activated points contained in a masker is less than two, then it is reinitialized. Fig. 2 describes the different steps of the MoDEAFC algorithm. However, the crossover and selection processes are identical with classical DE.
Fig. 1. Vector encoding scheme in the proposed method. A total of six cluster centers have been encoded for a data set. Only the valid cluster centers have been shown as circles.

IV. EXPERIMENTAL RESULTS

In order to establish the effectiveness of the proposed DEAFC and MoDEAFC algorithms, their performance is compared to the automatic fuzzy clustering using improved differential evolution, variable length genetic algorithm based fuzzy clustering, and an iterated version of the FCM. In the iterated FCM (IFCM), the FCM algorithm is run for different values of $K$ starting from 2 to $\sqrt{n}$, with $n$ being the number of data points. For each $K$, it is executed 50 times and the run corresponding to the best $J_m$ value is taken. Among these best solutions for different $K$ values, the solution producing the minimum XB index value is chosen as the best partitioning. The AFCIDE and VGFAC algorithms are implemented in the same way as in [8] and [9]. In case of AFCIDE, the only change is made in the fitness computation process. These two methods are also trying to find the optimal number of clusters automatically. Moreover, to investigate the effects of the changes made in the classical DE algorithm, we have compared the MoDEAFC with a classical DE-based automatic fuzzy clustering (DEAFC) method, which uses the same vector representation scheme and fitness function as the MoDEAFC.

Fig. 2. Algorithm of MoDEAFC.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>with known partitioning</th>
<th>using clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$J_{avg}$</td>
<td>$J_{BH}$</td>
</tr>
<tr>
<td>SPOT</td>
<td>1.0742</td>
<td>4.7994</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LANDSAT</td>
<td>1.0348</td>
<td>3.3766</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE I

AVERAGE J-M DISTANCE ($J_{avg}$) AND BHATTACHARYYA BOUND ($J_{BH}$) FOR DIFFERENT DATA SETS

A. Numeric Remote Sensing Data Sets

1) SPOT: This is a 3-D data set [corresponding to green, red, and near-infrared (NIR) bands] consisting of 932 samples partitioned into seven distinct classes of turbid water (TW), pond water (PW), concrete (Concr.), vegetation (Veg), habitation (Hab), open space (OS), and roads (including bridges) (B/R). For the purpose of illustration, Fig. 3(a) shows the scatter plot of the data set from where it can be seen that the clusters are highly complex and overlapping in nature.

2) LANDSAT: This data set has 795 samples and four bands (features): green, red, NIR, and IR. Since the features are highly correlated, the feature space is reduced to two by using the principal component analysis. The data set contains five classes, viz., Manda Granite, Romapahari Granite, Vegetation, Black Phillite, and Alluvium. Fig. 3(b) shows the scatter plot of the data set.

Note that the statistical separability of different clusters for above two data sets has been reported in Table I. In this regards, the average J-M distance ($J_{avg}$) [15] and the average Bhattacharyya Bound ($J_{BH}$) [15] are taken into consideration. As reported in Table I, the MoDEAFC algorithm produces better $J_{avg}$ and $J_{BH}$ values compared to the other algorithms both for SPOT and LANDSET data.

B. Input Parameters

The MoDEAFC, AFCIDE and DEAFC algorithms are executed for 100 generations with a population size of 20. The crossover probability (CR) used for DEAFC and MoDEAFC are 0.8. The mutation factor $F$ and $\alpha$ for DEAFC and MoDEAFC are set to 0.8 and $1/[1 + \exp((-1/generation))]$.  

respectively. However, the crossover probability (CR) and the mutation factor $F$ for AFCIDE are taken by the same procedure as described in [8]. The VGA-based fuzzy clustering algorithm is also executed for 100 generations with a population size of 20. The crossover and mutation probabilities are 0.8 and 0.3, respectively. The FCM algorithm is executed until it converges to the final solution for different numbers of clusters.

Note that the performance of the proposed method for two numeric remote sensing data sets is measured by the Minkowski Score (MS) [16], Kappa Index [17] and % correct [17]. For MS, the optimum score is 0, with lower scores being “better.” However, a higher value of kappa (close to 1) indicates better accuracy.

### C. Results

Tables II and IV provide the result of 50 consecutive runs of the algorithms for different data sets. Table II gives the number of clusters found by the different algorithms as well as the number of time PC found on the experimental data sets.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>AC</th>
<th>PC by Different Algorithms/No. of Times to get PC out of 50 runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPOT</td>
<td>7</td>
<td>MoDEAFC 7/47, AFCIDE 6/43, DEAF 6/39, VGAFC 8/34, IFCM 9/38</td>
</tr>
</tbody>
</table>

Here, AC denotes actual number of clusters present in the data and PC denotes the predicted number of clusters.

Table III gives the average Kappa index and % correct for different data sets.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Algorithm</th>
<th>Kappa Index</th>
<th>% correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPOT</td>
<td>MoDEAFC</td>
<td>0.8203</td>
<td>85.1231</td>
</tr>
<tr>
<td></td>
<td>AFCIDE</td>
<td>0.7407</td>
<td>78.5126</td>
</tr>
<tr>
<td></td>
<td>DEAF</td>
<td>0.6262</td>
<td>69.4172</td>
</tr>
<tr>
<td></td>
<td>VGAFC</td>
<td>0.6285</td>
<td>65.8061</td>
</tr>
<tr>
<td></td>
<td>IFCM</td>
<td>0.4238</td>
<td>48.0273</td>
</tr>
<tr>
<td>LANDSAT</td>
<td>MoDEAFC</td>
<td>0.8426</td>
<td>87.7481</td>
</tr>
<tr>
<td></td>
<td>AFCIDE</td>
<td>0.7573</td>
<td>78.7331</td>
</tr>
<tr>
<td></td>
<td>DEAF</td>
<td>0.6255</td>
<td>66.0562</td>
</tr>
<tr>
<td></td>
<td>VGAFC</td>
<td>0.5744</td>
<td>62.8037</td>
</tr>
<tr>
<td></td>
<td>IFCM</td>
<td>0.4302</td>
<td>48.8501</td>
</tr>
</tbody>
</table>

Fig. 4. Best Confusion matrix for (a) SPOT and (b) LANDSAT data set out of 50 runs of MoDEAFC.

### Table IV

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Algorithm</th>
<th>XB</th>
<th>MS</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPOT</td>
<td>MoDEAFC</td>
<td>1.0318</td>
<td>0.6027</td>
<td>290.0503</td>
</tr>
<tr>
<td></td>
<td>AFCIDE</td>
<td>1.0864</td>
<td>0.6802</td>
<td>222.8103</td>
</tr>
<tr>
<td></td>
<td>DEAF</td>
<td>1.1361</td>
<td>0.7502</td>
<td>267.6225</td>
</tr>
<tr>
<td></td>
<td>VGAFC</td>
<td>1.4033</td>
<td>0.8077</td>
<td>179.0715</td>
</tr>
<tr>
<td></td>
<td>IFCM</td>
<td>1.7807</td>
<td>0.8751</td>
<td>159.2027</td>
</tr>
<tr>
<td>LANDSAT</td>
<td>MoDEAFC</td>
<td>1.3637</td>
<td>0.6057</td>
<td>2310.5371</td>
</tr>
<tr>
<td></td>
<td>AFCIDE</td>
<td>1.3757</td>
<td>0.7093</td>
<td>2305.6083</td>
</tr>
<tr>
<td></td>
<td>DEAF</td>
<td>1.3863</td>
<td>0.8063</td>
<td>2297.6035</td>
</tr>
<tr>
<td></td>
<td>VGAFC</td>
<td>1.4271</td>
<td>0.8837</td>
<td>2286.5351</td>
</tr>
<tr>
<td></td>
<td>IFCM</td>
<td>1.8152</td>
<td>0.9562</td>
<td>2273.8169</td>
</tr>
</tbody>
</table>

Fig. 5. Boxplot of MS for different clustering algorithms for (a) SPOT and (b) LANDSAT data set.

Fig. 6. Convergence plot of different clustering algorithms for (a) SPOT and (b) LANDSAT data set.

As a result, the IFCM is expected to provide solutions in a wider range compared to GA- and DE-based algorithms. Moreover, due to the better convergence property, the set of solutions provided by the MoDEAFC for different runs are not only better than DEAFC or VGAFC; they also belong to a relatively narrow range. Fig. 5 corroborates these facts. Fig. 6 demonstrates the convergence plot of different algorithms. It is also observed from Fig. 6 that the MoDEAFC converges faster than the other algorithms to the optimal value. As already mentioned, the statistical separability of different clusters has been provided in Table I. It can be observed from Table I that
V. IMAGE CLASSIFICATION

Two IRS satellite images of Calcutta and Mumbai are used for demonstrating unsupervised pixel classification. Each image is 512 × 512 and was acquired from the Indian Remote Sensing Satellite (IRS-1A) [18] using the LISS-II sensor at a resolution of 36.25 m × 36.25 m. The images are contained in four spectral wavelength bands: blue (0.45–0.52 μm), green (0.52–0.59 μm), red (0.62–0.68 μm), and NIR (0.77–0.86 μm). The effectiveness of the proposed MoDEAFC technique is shown by the measure of the XB and I index. Moreover, the clustering can also be verified visually from the clustered images.

A. IRS Image of Calcutta

Figs. 7–12 show the original IRS Calcutta image and the corresponding clustered images using the proposed MoDEAFC, AFCIDE, DEAFC, VGAFC, and IFCM algorithms, respectively. From ground knowledge, it is known that the image has four classes [9]: turbid water (TW determined as deep green), pond water (PW determined as light green), concrete (Concr. determined as yellow), and open space (OS determined as white). The river Hooghly cutting across the middle of the image has been classified as TW, whereas several fisheries observed toward the lower-right portion of the image are correctly identified as PW. It appears from Fig. 8 that the water class has been differentiated into TW (the river Hooghly) and PW (canal, fisheries, etc.) because they differ in their spectral properties. Towards the lower right side of the image, a township, Salt Lake, has come out partially as combination of concrete and open space, which appears to be correct since this particular region is known to have several open spaces. The...
canal bounding Salt Lake from the upper portion has also been correctly classified as PW. Two parallel lines observed toward the upper right hand side of the image correspond to the airstrips in the Dumdum airport, and the airstrip is classified rightly as concrete. The presence of some small areas of PW beside the airstrips is correct again as these correspond to the several ponds around the region. The predominance of concrete on both sides of the river, particularly toward the bottom of the image (i.e., the central part of the city) is also correct. From Fig. 10, it can be noted that the river Hooghly and the city region has been incorrectly classified as belonging to the same class. Another flaw is apparent: the whole Salt Lake city has been put into one class. It is evident that although some portions of Salt Lake such as canals, parts of the airstrip, and fisheries are correctly identified, a significant amount of confusion lies in the IFCM clustering result. Results produced by the AFCIDE and DEAFC are better than the VGAFC and IFCM but poorer compared to the MoDEAFC. Still, all four algorithms are efficient in finding the accurate number of clusters, as shown in Table V. The XB and $I$ indices reported in Table VI show the effectiveness of the proposed method.

### B. IRS Image of Mumbai

Figs. 13 and 14 show the original Mumbai image and the classified image using the proposed MoDEAFC technique. According to available ground knowledge [9], the different clusters are labeled as concrete habitation (Concr. determined as yellow), open spaces (OS1 and OS2 determined as ash and white, respectively), habitation (Hab determined as deep yellow), vegetation (Veg determined as green), and turbid water (TW1 and TW2 determined as deep blue and light blue, respectively). As can be seen, the elongated city area is surrounded on three sides by the Arabian Sea, which is distinguished into two classes—TW1 and TW2. It is evident from figure that the sea water has two distinct regions with different spectral properties. Hence, a clustering result providing two partitions for this region is expected. At the bottom right of the image, there are several islands, including the well known Elephanta Islands. The dockyard is situated on the south eastern part of Mumbai, which can be seen as a set of three fingerlike structures. Note that the classes habitation and concrete share common properties. The islands, dockyard, and several road structures have mostly been correctly identified in the image. Within the islands, as expected, there is a high proportion of open space and vegetation. The southern part of the city, which is heavily industrialized, has been classified as primarily belonging to habitation and concrete. Figs. 15–18 demonstrate the Mumbai image clustered using the AFCIDE, DEAFC, VGAFC, and IFCM techniques, respectively. As shown in the figures, the water of the Arabian Sea has been partitioned into three regions rather than two as obtained in the case of the MoDEAFC. The other regions appear to be classified more or less correctly for this data. As can be seen in Table V for the Mumbai image, the MoDEAFC is able to provide the correct number of clusters, whereas other algorithms failed to do so. As earlier, the result produced by the AFCIDE and DEAFC are better than the VGAFC and IFCM but poorer than the MoDEAFC, which is also evident in Table VI.
working in these directions.

elements of masker in the MoDEAFC can be adaptive in nature using the concept of amount of dominance [19]. Moreover, the techniques. In this context, the IRS satellite images of Calcutta and Mumbai have been classified using the proposed technique and the AFCIDE, DEAFC, VGAFC, and IFCM clustering techniques. In this context, the IRS satellite images of Calcutta and Mumbai have been classified using the proposed technique and compared with other clustering algorithms.

For further research, the MoDE-based algorithm can also be extended for solving multiobjective optimization problems using the concept of amount of dominance [19]. Moreover, the elements of masker in the MoDEAFC can be adaptive in nature rather than choosing it randomly. The authors are currently working in these directions.

VI. CONCLUSION

In this article, a modified differential evolution based fuzzy clustering for automatic cluster evolution technique has been described for IRS image classification. The problem of fuzzy clustering has been modeled as an optimization of a cluster validity measure. A modified version of the differential evolution technique has been used as a stochastic optimization tool. Results on different numeric remote sensing data sets indicate that the MoDEAFC consistently performs better than the AFCIDE, DEAFC, VGAFC, and IFCM clustering techniques. In this context, the IRS satellite images of Calcutta and Mumbai have been classified using the proposed technique and compared with other clustering algorithms.

For further research, the MoDE-based algorithm can also be extended for solving multiobjective optimization problems using the concept of amount of dominance [19]. Moreover, the elements of masker in the MoDEAFC can be adaptive in nature rather than choosing it randomly. The authors are currently working in these directions.

REFERENCES


Ujjwal Maulik (SM’05) received the Ph.D. degree from Jadavpur University, Kolkata, India, in 1997. He is currently a Professor with the Department of Computer Science and Engineering, Jadavpur University. He has worked with the Center for Adaptive Systems Application, New Mexico; the University of New South Wales, Australia; the University of Texas at Arlington; the University of Maryland Baltimore County; the Fraunhofer Institute AIS, Germany; Tsinghua University, Beijing, China; the University of Rome, Rome, Italy; and the German Cancer Research Institute and University of Heidelberg, Heidelberg, Germany. He has coauthored five books and about 160 papers. His main research interests include evolutionary computing, pattern recognition, and bioinformatics.

Indrajit Saha received the M.Tech. degree (with first class honors) in computer science and engineering from Kalyani University, Kalyani, India, in 2007. He is currently working toward the Ph.D. degree at the University of Warsaw, Warsaw, Poland. His research interests include soft and evolutionary computing, bioinformatics, and image processing. Mr. Saha was the recipient of the National Doctorate Fellowship Award from the All India Council of Technical Education and Erasmus Mundus Mobility with an Asia fellowship from the European Union in 2009.