

# A hybrid FPAB/BBO Algorithm for Satellite Image Classification

Navdeep Kaur Johal  
Lecturer  
RIEIT  
Railmajra, Punjab, India

Samandeep Singh  
Lecturer  
RIEIT  
Railmajra, Punjab, India

Harish Kundra  
Asst. Professor  
RIEIT  
Railmajra, Punjab, India

## ABSTRACT

In the past years, remote sensing has been used for the classification of satellite image on a very large scale. This paper deals with image classification by using swarm computing technique. In this work, we use a new swarm data clustering method based upon flower pollination by artificial bees to cluster the satellite image pixels. The aim of clustering is to separate a set of data points into self-similar groups. Those clusters will be further classified using Biogeography Based Optimization. The results indicate that highly accurate classification of the satellite image is obtained by using the proposed algorithm.

## Keywords

FPAB, Biogeography Based Optimization, Satellite Image Classification

## 1. INTRODUCTION

Swarm intelligence is any attempt to design algorithm or distributed problem-solving devices inspired by collective behavior of social insect colonies or other animal societies [1]. This field deals with how social insects collectively perform some specific tasks in daily life (that include finding food, building or extending nest, efficiently feeding the brood, responding to external challenges, spreading alarm etc.), modeling their behavior, and using the model as a basis upon which artificial variations can be developed, either by tuning the model parameters beyond the biological relevant range or by adding nonbiological features to the model.

### 1.1. Artificial Bees Clustering

Clustering is the division of data into similar groups of objects. Each cluster consists of objects that are similar among themselves and dissimilar to objects of other groups. The clustering in this work is based on flower pollination by artificial bees[4], which is named as FPAB. FPAB does not require any parameter settings and any initial information such as the number of classes and the number of partitions on input data. Initially, in FPAB, bees move the pollens and pollinate them. Each pollen will grow in proportion to its garden flowers. Better growing will occur in better conditions. After some iteration, natural selection reduces

the pollens and flowers and the gardens of the same type of flowers will be formed. The prototypes of each garden are taken as the initial cluster centers for Fuzzy C Means algorithm which is used to reduce obvious misclassification errors. In the next stage, the prototypes of gardens are assumed as a single flower and FPAB is applied to them again.

In FPAB, each individual artificial bee is a simple agent with a one-slot memory. The memory is used to save the growth of carrying pollen in its source. The bees will pick up the pollen of flower with lowest growth and pollinate the pollen where it will grow better. Each pollen grows in proportion to its neighbor flowers and after some iteration the natural selection will select the flower with best growth of one species to survive and will sear others. Using this simple behavior and limited memory, the artificial bees can perform complicated tasks such as clustering.

### 1.2. FPAB Algorithm

Initially, flowers scattered randomly on 2D discrete grid, which is named as jungle. This jungle can be considered as a matrix of  $m \times m$  cells. The matrix is toroidal which allows the bees to fly from one end to another easily and there is no initially more than one flower in each cell of jungle. The size of the jungle depends on the number of flowers. We have used a jungle of  $m \times m$  such  $m^2 = 4n$  that where  $n$  is the total number of flowers (objects) to be clustered.

Initially, the bees are randomly scattered throughout the jungle. We use  $n/3$  bees, where  $n$  is the total number of flowers to be clustered. Each garden defined as a collection of 2 or more flowers (or pollens) and for simplicity of algorithm each garden spatially located in a single cell. Consider a garden  $G$  with  $n_g$  flowers. We define the following parameters:

- The compatibility between two flowers in the garden

$$C(pi, pj) = 1 - D(pi, pj)$$

Where  $D$  is normalized Euclidean distance between the attributes of flowers  $pi, pj$

- The center of garden  $G$

$$O_{center}(G) = \frac{1}{n_g} \sum_{O_i \in G} O_i$$

- The growth of the pollen in a garden

$$G(pi) = C(pi, O_{center})$$

- The flower with minimum growth in a garden G

$$p_{min} = \arg \min G(pi)$$

where  $pi \in G$

The main FPAB algorithm is written below:

1. Randomly scatter bees and flowers into the jungle (atmost one flower per cell)
2. Set the initial growth of all flowers to zero
3. Repeat 4, 5, 6 until stopping criteria
4. for each bee do
  - a) fly the bee in jungle
  - b) if the bee does not carry any pollen then it possibly picks up a pollen else the bee possibly pollinates
5. Next bee
6. Natural Selection

#### Algorithm I: The basic ant based algorithm [4]

Initially, the bees are randomly scattered into the jungle and also randomly pick up or pollinate when they reach a new garden in the jungle. Here, the stopping criterion for the bees is the number of times through the repeat loop.

#### 1.2.1. Picking up pollen

When the bee does not carry any pollen, it looks for possible pollens to pickup looking at 8 neighbor cells around its current location. If a garden or a flower exists there, it could possibly pick up pollen. We consider three cases here: if there was only one flower in the cell, the bee picks up its pollen with probability  $P_{pickup}$ . This probability can be set to any number between [0, 1] and it effects merely on the speed of the algorithm. If there was a garden of two flowers, the bee picks up the pollen of one flower randomly from the garden with probability  $P_{dist}$ , this probability is set the same as  $P_{pickup}$ . If there was a garden of more than two flowers, the bee must pick up the pollen of flower with lowest growth in that garden.

The algorithm for picking up pollen is given below:

1. mark the 8 neighbor cells around the bee as “unexplored”
2. Repeat
  - a) consider the next unexplored cell around the bee
  - b) if the cell contains a garden of 2 flowers then randomly pick up a pollen of ones with probability  $P_{dist}$
  - c) else if the cell contains a garden of more than 2 flowers then pick up the pollen of flower with lowest growth
3. until all 8 neighbor cells have been explored or one pollen has been picked up

#### Algorithm II: Algorithm to pick up an object [4].

#### 1.2.2. Pollination

When the bee carries pollen, it examines 8 neighbor cells around its current location. Three cases may occur: if the cell is empty, the bee do nothing and consider the next neighbor cell. If the bee encounters a flower with lower growth than its carrying pollen growth, the bee discards the carrying pollen and picks up the pollen with lower growth. If the carrying pollen does not exist here and its growth will be better, the bee will pollinate the pollen. The algorithm of pollination is given below:

1. mark the 8 neighbor cells around the bee as “unexplored”
2. Repeat
  - a) consider the next unexplored cell around the bee
    - i) if the cell is empty then continue
    - ii) if the cell contains a flower with lower growth than carrying pollen then discard the carrying pollen and pick up the pollen of flower with lower growth
    - iii) if the carrying pollen has better growth here than its growth in its source garden then pollinate in this garden
  - b) label the cell as “explored”
4. until all 8 neighbor cells have been explored or one pollen has been dropped

#### Algorithm III: Algorithm for Pollination [4].

#### 1.2.3. Natural Selection

After pollinating, there maybe exists more than one instance of one flower in different gardens, which have different growth. Now we use natural selection process to sear instances with lower growth of the same type and allow others to survive. The flowers with the same characteristics are agglomerated in one garden i.e. same type flowers are clustered. We use here hard natural selection, which deletes all flowers of the same type with lower growth than maximum growth of that flower.

The algorithm of natural selection is given below:

1. for all species of flowers find maximum growth of that flower
2. for each garden in jungle
  - a) for each flower in garden  
If the flower is lower than its maximum growth in the same type then delete flower and update growth of garden
  - b) Next flower
3. Next garden.

#### Algorithm IV: Natural selection algorithm [4].

#### 1.2.4. Merge Algorithm

Artificial bees constitute several clusters of very similar flowers in limited iterations and objects will be clustered unsupervised without any parameter settings and any initial information about clusters. For removing obvious misclassification errors such as single flowers which does not belong to any garden, we use FCM and then we find the centroids of these gardens and assume them as single flowers to achieve lower number of gardens until stopping criterion (here after sufficient levels of clustering). The total algorithm of clustering is given below:

1. Randomly scatter flowers and artificial bees into the jungle at most one flower per cell.
2. for  $i=1$  to 100
  - a) for each bee do
    - i. fly the bee in jungle
    - ii. if the bee does not carry any pollen then it possibly picks up pollen
    - iii. else the bee possibly pollinates
  - b) Next bee
  - c) Natural Selection
3. Next  $i$
4. Apply Fuzzy C Means based on the centers of clusters obtained from the previous step
5. Assume the centers of clusters as new single flowers i.e. a single flower per garden
6. Repeat 2-5 to reach appropriate number of clusters

**Algorithm V: The total clustering + merging algorithm [4].**

### 1.3. Biogeography based Optimization

Biogeography-based optimization (BBO) is an evolutionary algorithm that is based on the science of biogeography. Biogeography is the study of the geographical distribution of organisms. Habitats that are well suited as habitats for biological species are said to have a high Habitat suitability index (HSI). Features that correlate with HSI include rainfall, topographic diversity, area, temperature, etc. The variables that characterize these features are called suitability index variables (SIVs). SIVs are the independent variables of the Habitat, and HSI is the dependent variable [6].

Habitats with a high HSI tend to have a large number of species, and those with a low HSI have a small number of species. Habitats with a high HSI have many species that emigrate to nearby islands because of the accumulation of random effects on its large populations. Emigration occurs as animals ride flotsam, fly, or swim to neighboring islands.

Suppose that we have some problem, and that we also have several candidate solutions. A good solution is analogous to an island with a high HSI, and a poor solution is like an island with a low HSI. High HSI solutions are more likely to share their features with other solutions, and low HSI solutions are more likely to accept shared features from other solutions. This approach to problem solving is called biogeography-based optimization. As with every other evolutionary algorithm (EA), each solution also typically has some probability of mutation, although mutation is not an essential feature of BBO [7].

### 1.4. Remote Sensing Image Classification

Satellite images have several properties embedded in them like spatial, spectral and temporal properties etc. Through these properties feature extraction can be performed. The problem of extracting homogeneous regions from an image is viewed as the task of clustering the pixels in the intensity space. In particular, satellite images contain land cover 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 presents a challenging task [2].

In the future, new generation of high-resolution satellite sensors will acquire enormous data quantity, so extracting features from a remote sensed image will become even more difficult. Researchers have been continuously searching for new techniques that can extract maximum information from the remotely sensed image.

Image classification is conducted in three modes: supervised, unsupervised, and hybrid. A supervised classification requires the manual identification of known surface features within the imagery and then using a statistical package to determine the spectral signature of the identified feature. The "spectral fingerprints" of the identified features are then used to classify the rest of the image. An unsupervised classification scheme uses spatial statistics (e.g. the ISODATA algorithm) to classify the image into a predetermined number of categories (classes). These classes are statistically significant within the imagery, but may not represent actual surface features of interest. Hybrid classification uses both techniques to make the process more efficient and accurate.

## 2. PROPOSED WORK

In the proposed work, a hybridization of Flower Pollination by Artificial Bees (FPAB) and Biogeography Based Optimization Algorithm has been used to classify a satellite image. FPAB algorithm has been used to find the clusters of similar land cover features and those clusters are further classified into a particular land cover type using Biogeography based Optimization Algorithm. BBO algorithm is basically used to find the optimal solution of a problem. But satellite image classification is a clustering problem that requires each class to be extracted as a cluster. The original BBO algorithm does not have the inbuilt property of clustering. The BBO based image classification has been discussed in our previous work in [5].

As a real-world case, study area of Alwar (Rajasthan) is considered as it contains a good variety of Landuse/Landcover features. A 7-band satellite image of Alwar region of size 472 X 546 is taken as input for the purpose of image classification.

Since the image used in this work contains 2.5 lakh pixels, these pixels are first clustered into homogeneous groups using FPAB algorithm. The final clusters produced by honey bees are each treated as a species in BBO and is put in the universal habitat. These species are further migrated by using the algorithm

proposed in the next section and hence satellite image is classified.

### 1.5. Proposed Algorithm

The hybrid FPAB/BBO algorithm proposed in this section is inspired from the algorithm proposed in [5]. The modification to the original algorithm [5] has been proposed and the new algorithm is written in Figure 1.

1. Obtain the clusters of pixels of Satellite Image using Flower Pollination by Artificial Bees (FPAB) Algorithm.
2. Take six habitats – One Universal habitat where each species is the cluster obtained by FPAB Algorithm, other five are feature habitats which contains the pixels of water, vegetation, barren, rocky and urban (the features we want to extract from satellite image) provided by the expert.
3. Obtain the HSI of each feature habitat
4. Take one species from universal habitat at a time and recalculate HSI after migrating it to each feature habitat.
5. Species will be absorbed in that habitat where difference in HSI before and after migration is minimum.
6. Algorithm stops when no species is left in the universal habitat.

**Fig 1: A hybrid FPAB/BBO algorithm for satellite image classification**

Initially homogeneous clusters of image pixels are obtained using FPAB algorithm discussed in section B. Clusters obtained by honey bees are treated as species in BBO and are put to universal habitats. There are other feature habitats containing the training pixels of the corresponding feature produced by experts. For example, water habitat initially contains pixels of water and standard deviation of all the training pixels of water is calculated in each of seven bands. The HSI is calculated on these training pixels. Each species is taken from the universal habitat and is migrated to each feature habitat one at a time. The HSI of the habitat is recalculated after migrating the species to it. The difference between the original HSI and recalculated HSI is calculated. And if the difference is minimum in some particular feature habitat, species is migrated to that habitat.

FPAB algorithm obtains very homogenous clusters [4] and thus each cluster obtained by this algorithm contains pixels of one land cover feature. The species that are put to universal habitat initially contains pixels of one type only and by using the theory of migration will be absorbed in most suitable habitat. Thus this algorithm will classify the satellite image in just one iteration.

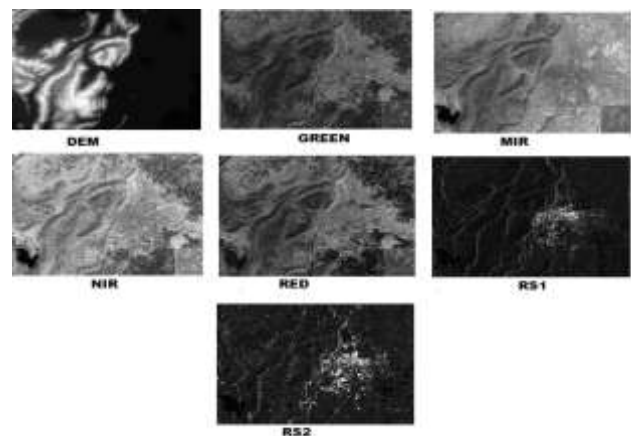
### 3. RESULTS AND DISCUSSION

The algorithm proposed in the previous section has been implemented in Matlab 7.0 [8]. The algorithm has been applied to multi-spectral, multi resolution and multi-sensor 7-band image of Alwar City (in Rajasthan, India). These bands are Red, Green, Near Infra Red (NIR), Middle Infra Red (MIR), Radarsat-1 (RS1), Radarsat-2 (RS2), and Digital Elevation Model (DEM) (figure 2).

Firstly, the clusters of pixels are obtained by FPAB algorithm. Each pixel of satellite image is treated as an object of FPAB algorithm. So total number of objects (n) in FPAB algorithm is equal to total number of pixels in satellite image. In the first iteration of FPAB, honey bees will pick and drop one pixel only. In second iteration, honey bees will pick and drop cluster of pixels obtained in the first iteration. The clusters obtained by FPAB algorithm are further classified by using BBO.

During classification, initially there is one universal habitat and five feature habitats. Each feature habitat initially consists of training sets. Training sets of the image are provided by the experts, which help us to classify the whole image. These are basically some of the pixels of the image, for which an expert is sure about that this pixel is of which land cover region. The training sets used in this study have been stored in the Microsoft excel sheets, with one row representing one pixel and 7-columns representing 7-bands of the image. Decision attribute describes the feature to which the each pixel (row) belongs. Figure 3 shows the training pixels originally present in the rocky habitat.

After calculating the HSI on training pixels, each species in the Universal habitat is migrated to the feature habitat and the HSI of the habitat is recalculated after migration. A given species is absorbed to that feature habitat where it makes minimum variation in HSI after migration.



**Fig 2: Seven Band Image of Alwar (Rajasthan)**

Fig 3: Training Set of Rocky Region

After applying the proposed algorithm to the 7-band of Alwar Image, the classified image is obtained in figure 4. The yellow, black, blue, green, red color represents rocky, barren, water, vegetation, urban region respectively.

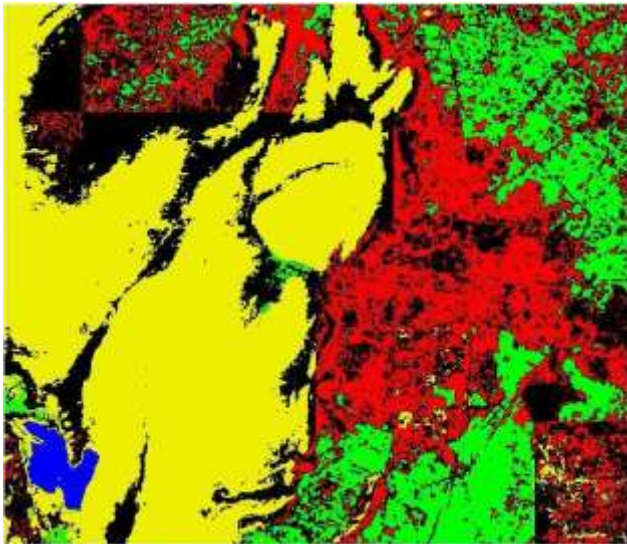


Fig 4: Hybrid FPAB/BBO classification of Satellite Image of Alwar region

### 3.1. Accuracy Assessment

Accuracy assessment is an important step in the classification process. The goal is to quantitatively determine how effectively pixels were grouped into the correct feature classes in the area under investigation. Classification accuracy of our proposed algorithm is expressed using classification *error matrix*. Error matrices compare, on a category-by category basis, the relationship between known reference data (ground truth) and the corresponding results of an automated classification. We took 150 vegetation pixels, 190 urban pixels, 200 Rocky pixels, 70 water pixels, 170 barren pixels from the training set and the error matrix obtained is shown in Table 1. The error matrix's interpretation along column suggests how many pixels are classified correctly by algorithm. For e.g. in the first column, out of total 150 vegetation pixels, 138 pixels were correctly classified into vegetation by the proposed algorithm, 10 were misclassified as

rocky and 2 were miss classified as Barren. All water pixels are correctly classified.

Producer's accuracies result from dividing the number of correctly classified pixels in each category (on the major diagonal) by the number of training set pixels used for that category (the column total). The producer accuracy for vegetation and urban are 92% and 48% respectively whereas in original work [5] these were 85% and 46%. So the vegetation and urban areas are classified better with the proposed algorithm.

Table 1: Error Matrix of hybrid FPAB/BBO based Satellite Image Classification of Alwar Region

	Vegetation	Urban	Rocky	Water	Barren	Total
Vegetation	138	11	0	0	1	150
Urban	0	90	2	0	36	128
Rocky	10	0	181	0	28	219
Water	0	0	0	70	0	70
Barren	2	89	17	0	105	213
Total	150	190	200	70	170	780

The Kappa coefficient of the Alwar image is calculated using following formula described by Lillesand and Kiefer in [3].

$$k = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}$$

r = number of rows in the error matrix (r =5 in our case)

$x_{ii}$  = the number of observations in row i and column i (on the major diagonal)

$x_{i+}$  = total of observations in row i (shown as marginal total to right of the matrix)

$x_{+i}$  = total of observations in column i (shown as marginal total at bottom of the matrix)

N = total number of observations included in matrix. (N=780 in our case)

The Kappa (K) coefficient of the Alwar image is 0.67927 which indicates that an observed classification is 67.927 percent better than one resulting from chance.

#### **4. CONCLUSION AND FUTURE SCOPE**

In this work, we have proposed a hybrid FPAB/BBO based algorithm as an efficient land cover classifier for satellite image. Many of the land cover features are identified much more clearly when proposed algorithm is used. The proposed algorithm produces homogeneous clusters using FPAB algorithm and those clusters are classified in land covers using BBO in just one iteration of algorithm. The Kappa (KHAT) –coefficient has been used as a measure of elicited knowledge and it also provides us a mechanism to compare the knowledge content with other paradigms of supervised classification.

The future scope of the research includes proposing certain modification to the algorithm so that the Kappa coefficient can be improved further. The current system is implemented using a simple heuristic technique namely standard deviation; the system performance can be increased by using other heuristic functions. Also, an unsupervised version of the Biogeography based satellite image classification may be explored by including the context information of clusters used in generation of species granules and then evaluating the HSI.

#### **5. REFERENCES**

- [1]. Bonabeau E., Dorigo M., and Theraulaz G. (1999), “Swarm Intelligence: From natural to artificial systems”, Oxford University press, NY, pp.1-25.
- [2]. Lillesand, T.M., Kiefer, R.W. and Chipman, J.W. (2003), “Remote Sensing and Image Interpretation”, Fifth Edition, Wiley & Sons Ltd., England, pp.586-592.
- [3]. Ma, H. (2009), “An analysis of the behavior of migration models for Biogeography-based Optimization”, Submitted for publication.
- [4]. Kazemian, M., Ramezani, Y., Lucas, C., Moshiri, B. (2006), “Swarm Clustering Based on Flowers Pollination by Artificial Bees”, Studies in Computational Intelligence (SCI), vol. 34, Springer Berlin Publishers, New York, pp. 191 – 202.
- [5]. Panchal, V.K., Singh, P., Kaur, N. and Kundra, H.(2009), “Biogeography based Satellite Image Classification”, International Journal of Computer Science and information Security, vol. 6, no.2, pp.269-274.
- [6]. Simon, D. (2008), “Biogeography-based Optimization”, IEEE Transactions on Evolutionary Computation, vol. 12, No.6, IEEE Computer Society Press. pp. 702-713.
- [7]. Simon, D., Ergezer, M. and Du, D.(2009), “Population Distributions in Biogeography-Based Optimization Algorithms with Elitism”, IEEE Conference on Systems, Man, and Cybernetics, San Antonio, TX, pp. 1017-1022.
- [8]. The MATLAB ver 7, The MathWorks, Inc.