PSO based Selection of Spectral Features for Remotely Sensed Image Classification

Agrawal R.K.
G.H.Raisoni College of Engineering, Nagpur,India

N.G.Bawane, Ph.D
S. B. Jain Institute of Technology, Management & Research, Nagpur, India

1. ABSTRACT
Spectral feature used in remotely sensed image classification are recorded in narrow, adjacent frequency bands in the visible to infrared spectrum. Due to narrow spacing, these features are highly correlated and provide some redundant information which may reduce classification accuracy. Hence discriminative feature selection technique is required for better classification. In this paper, we present particle swarm optimization based technique to select best spectral features for remotely sensed image classification. The pixels intensity in selected best spectral band is used to construct the feature vector for that pixel. Each pixel in multispectral imagery is classified into various land cover types like water, vegetation, road and urban area etc. We employed ANN for supervised classification of the image pixel. The accuracy obtained with proposed algorithm is compared with that of traditional classifiers like MLC and Euclidean classifier. The performance of the proposed system is evaluated quantitatively using *Xie-Beni* and β indexes. The result shows the superiority of the proposed method to the conventional one.

General Terms
Pattern Recognition,

Keywords
ANN, Classification, Feature selection, PSO, Multispectral image.

2. INTRODUCTION
Remotely sensed multispectral satellite images consist of images of Earth's surface taken into few spectral bands. These images are used to derive landcover map for various applications. The classification approach can be broadly classified into two categories: unsupervised and supervised [1]. The unsupervised clustering techniques are used when no priory information exist about land cove types. The statistical approaches for supervised classification are distance classifiers and maximum likelihood classifier (MLC). They assume same form of distribution for the data to be classified which may not be correct and difficult to obtain due to higher dimensionality of the data. Also they are subjected to "Hughes phenomenon" [2]. In nonparametric approach, artificial neural networks are suitable choice due to their ability to learn the data distribution though training.

In our survey, we found that different neural network architectures are used for unsupervised and supervise classification [3-9]. For supervised classification, feed forward neural network with only one hidden layer is widely used. The performance of classification system depends on the set of features used [10]. A variety of features based on texture, shape etc. have been used. Our target is the classification system based on original spectral features. The researchers used the all available spectral bands for preparing the feature vector. We believe that use of all spectral bands is not necessary since these images are recorded in narrow, adjacent frequency band and therefore may have same redundant information. For classification distinctive information is required. Thus there is need of a methodology to find best spectral band. The aim is to find subset of the original spectral features that have distinctive information.

Feature selection procedure should perform search over candidate solutions to select the optimal subset [11]. In this context, Evolutionary algorithm with population based search on can provide effective solution. Particle swam optimization is population based stochastic search method inspired by the social behavior of animals. Each particle (a candidate solution) of a given population can benefit from the past experiences of all other individuals in the same population. In this work we present the particle swarm optimization based selection of best spectral features among the available spectral bands. The block diagram of the system is shown in Figure 1. The rest of the paper is organized as follows. Section 3 briefly describes the particle swam optimization technique. The proposed algorithm is explained in section 4. In section 5, experimental setup is discussed. Finally, result and conclusion are discussed in section 6.

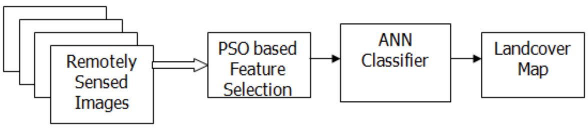

Figure 1. Block diagram of the system

3. PARTICLE SWARM OPTIMIZATION
PSO is population (called as swarm) based search method invented by Kennedy and Eberhart [12]. It is stochastic optimization technique inspired by the social behavior of animals. Each particle (a candidate solution) of a given population can benefit from the past experiences of all other individuals in the same population. During the iterative search process, each particle will adjust its velocity and position according to its own experience as well as those of the other particles in the swarm. Let us consider a swarm of particle of size S i.e. $P_i (i = 1,2,...S)$. Let $P_t(t)$ be the current position, $V_i(t)$ be its velocity at iteration t and $P_{bi}(t)$ the best position identified. Let P_g be the best global position found over all trajectories traveled by the particles of the swarm. Position optimality is measured by the fitness functions defined based on the given optimization problem.

During the search process, the particles move according to the following rule:

$$V_i(t+1) = wV_i(t) + c_1 r_1 (P_{bi}(t) - P_i(t)) + c_2 r_2 (P_g(t) - P_i(t)) \quad \ldots(1)$$

$$P_i(t+1) = P_i(t) + V_i(t+1) \quad (2)$$

where r_1 and r_2 are random variables drawn from a uniform distribution in the range $[0,1]$, c_1 and c_2 are two acceleration constants with respect to the best global and local positions respectively. These parameters determine the relative bias of the best position of the particle (self experience) and the global best position (experience of group members). The inertia weight w is used as a tradeoff between the global and local exploration capabilities of the swarm. Equation (1) allows the computation of the velocity at iteration $(t+1)$ for each particle in the swarm and the particle position is updated with Eq. (2).

These equations are iterated until maximum number of iterations is completed or the best value of the adopted fitness function is reached. Since in this application particle have discrete binary values of 1's and 0's, velocity value will indicate the probability of bit taken the value 1 or 0. Therefore update formula changes as follows.

$$\begin{aligned} P_i(t+1) &= 1 \quad if \quad sig(V_i(t+1)) \geq 0.7 \\ &= 0 \quad \text{else} \end{aligned} \quad (3)$$

This is called as binary PSO.

4. PSO BASED FEATURE SELECTION

The important issue in PSO setup is to determine the particle structure and fitness function.

4.1 Particle Structure

Since our aim is to detect best spectral band, the particle structure will be a vector that encodes the spectral bands in term of Boolean value. If d is the total number of spectral band available then,

$$\begin{aligned} f(i) &= 1 \ \text{ if } i \text{ th featue is selected} \\ &= 0 \ \text{ f } i \text{ th featue is not selected} \\ \text{where } i &= 1 \ to \ d \end{aligned}$$

The structure of particle is as shown in Figure 2.

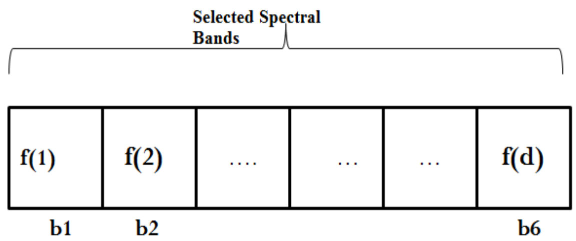

Figure 2. PSO Structure

4.2 Fitness Function

During optimization process, goodness of the particle is evaluated by function called as fitness or objective function. The lower value indicates better fitness of the particle. We have selected mean squared error (MSE) on test samples (N) as fitness function.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (X_i - D_i)$$

The low MSE means less difference in desired output (D) and actual output (X). Hence more will be the accuracy.

4.3 Algorithm Description

The steps in PSO based spectral band selection algorithm are as follows.

1. Initialization
   a. Generate randomly an initial swarm of size S.
   b. Initialize each particle position $P_i (i = 1,2,\ldots S)$ as follows: Choose the number of detected features randomly in the interval [1, d]. Choose randomly feature detection coordinates: set them to 1, while fix all other feature detection coordinates to 0. [d is the total number of spectral bands]
   c. Set to zero the velocity vectors $V_i(t)$ associated with the S particles.
   d. Set the best position of each particle with its initial position, i.e., $P_{bi} = P_g$
   e. For each position of the particle P_i from the swarm, train an ANN classifier and compute the corresponding fitness function i.e MSE
2. Search process
   a. Detect the best global position in the swarm exhibiting the minimal value of the considered fitness function over all explored trajectories.
   b. Update the speed of each particle using equation 1.

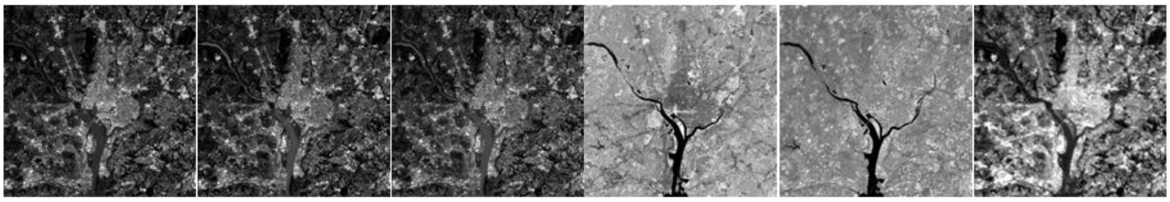

Figure 3. Multispectral Images

c. Update the position of each particle using equation 3.

d. For each candidate particle P_i , train an ANN classifier and compute the corresponding fitness function.

e. Update the best position P_{bi} of each particle if its current position has a smaller fitness function.

3. Convergence: If the maximum number of iterations is not yet reached, return to step 2.

4. Classification

a. Select the best global position in the swarm and train an ANN classifier fed with the subset of detected features as encoded in that best global particle structure.

b. Classify all pixels using selected spectral features and the trained ANN classifier.

5. EXPERIMENTAL SETUP

5.1 Multispectral Image Data

We have used images of Washington DC city area taken by Landsat satellite [13]. The data set consist of six images of size 512×512 taken in six spectral bands: b1: visible blue (450 – 520 nm), b2: visible green (520 – 600nm), b3: visible red (630 – 670 nm), b4: near infrared (760 – 900 nm), b5: middle infrared (1550 – 1750 nm) & b6: thermal infrared (10,400 – 12500 nm). The land cove types visually identified are: road, river, vegetation and urban area. Figure 3 shows the images corresponding to all bands.

5.2 Neural Network

For supervised classification we have used feed forward neural network with one hidden layer. The number of nodes in input layer is equal to the dimension of feature vector of the pixel to be classified. The pixels intensity in spectral band is used as feature vector. Thus the number of input layer nodes depends upon the number of best spectral bands selected by proposed PSO based algorithm.

The number nodes in the output are equal to the land cover types to classify. Hence there are four output nodes corresponding to four classes: road, river, vegetation and urban area. The number of nodes in hidden layer is equal to square root of the product of the number of input- and output-layer nodes. The network is first trained by back propagation algorithm using selected training samples and then used as a classifier in forward direction [14].

5.3 Training and Test Data Set

These data sets are constructed by visual inspection of the images in Matlab software. Total 50 samples of each class are randomly selected and divided equally divided into two set: one for training and other for testing.

5.4 Performance Evaluation Indexes

Following two indexes are used to evaluate the performance of different classification algorithms.

5.4.1 β Index

It is *index of homogeneity* and defined as the ratio of the total variation and within-class variation [15]. For given image total variation remains constant, therefore β value is dependent on within class variation. The higher be homogeneity within class, the lower would be variation within the class and hence higher would be the β value. Mathematically, it is given by,

$$\beta = \frac{\sum_{i=1}^{C}\sum_{j=1}^{M_i}\left(X_{ij} - \overline{X}\right)^2}{\sum_{i=1}^{C}\sum_{j=1}^{M_i}\left(X_{ij} - \overline{X}_i\right)^2}$$

where $\overline{X}$ is mean of pattern vectors, X_{ij} is the j th pattern vector ($j = 1...M_i$) of the class i ($i = 1...C$), $\overline{X}_i$ mean of pattern vector of the i th class. For better classification, higher β value is desirable.

5.4.2 Xie-Beni Index

It is defined as the ratio of compactness and separation [16]. It measures the within cluster compactness and separation between clusters. For good classification, smaller value of *Xie-Beni* index is desirable. Mathematically, it is expressed as,

$$XB = \frac{1}{N} \frac{\sum_{i=1}^{c}\sum_{j=1}^{N} \mu_{ij} \left\| V_i - X_j \right\|^2}{\min_{i \neq k} \left\| V_i - V_k \right\|^2}$$

where μ_{ij} is the membership value of j th pixel for i th class, v_i is the centroid of i th class, x_j is j th pattern and N is the total number of patterns.

6. RESULT & CONCLUSION

We ran the algorithm for different population size: 5, 10 and for different epoch 100, 1000 and 1500. The result obtained is shown in Table 1. Figure 4 indicates trajectory MSE during execution of algorithm. It is observed that only three spectral bands are sufficient for the classification. Among the red, green and blue wavelength, any one is getting selected in different run of the algorithm. Also from near infrared and middle infrared, any one is being selected. The thermal infrared band is selected in all runs.

The visual inspection of these images validates the result of our algorithm. Since red, green and blue band images are visually similar and provide same information to the classifier. Both infrared images are also visually looks similar and hence provide same information. And thermal band is distinct and hence provides distinct information for classification. Thus as per the result, we selected the three spectral bands b3, b4, and b6 for making feature vector for each pixel. These vectors are then applied to the trained neural network for classification. The classified gray scale image is shown in Figure 5. As observed all details are well classified.

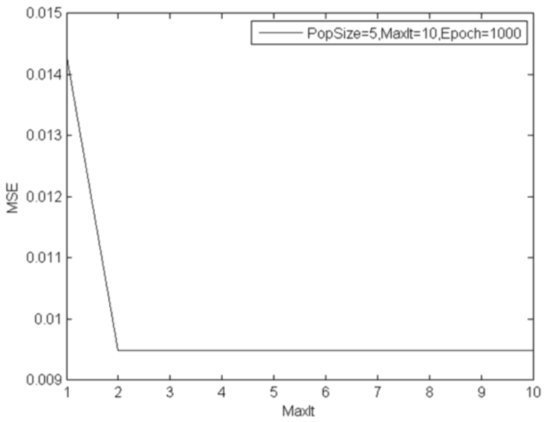

Figure 4. MSE trajectory

The proposed work is compared with traditional classifier like Euclidean classifier and MLC using *Xie-Beni* and β indexes. The result shown in Table 2 indicates the superiority of the algorithm. Although the accuracy obtained by MLC is comparable to that obtained by our algorithm, but qualitatively classification provided by our algorithm is much better than that of MLC. MLC fails to classify finer details in the image.

Table 1. Result of different run of algorithm

Epoch	Population Size	Detected Spectral Bands	MSE
100	5	b3b5b6	0.013
	10	b3b4b6\ b2b5b6	0.014
1000	5	b3b4b6 \b2b3b6	0.009
	10	b3b4b6	0.010

Table 2. Comparison with different algorithm

Classifier	Accuracy (%)	*Xie-Beni*	β
Euclidean	90	3.5	2.2
MLC	94	3	2
Proposed Algorithm	94	0.9	2.3

Thus quantitatively as well as qualitatively our algorithm provides significant improvement in classification compared to both traditional classifiers. Hence our proposed algorithm selects the best spectral band and helps to achieve good classification accuracy using subset of the available spectral bands. Obviously this also results in less computational cost during classification phase.

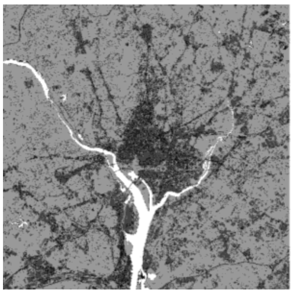

Figure 5. Classified Gray scale Image

7. ACKNOWLEDGMENTS

The author would like to thanks Rafael .C. Gonzalez, Richard .E. Woods & Steven L. Eddins for permitting us to use multispectral image data set.

8. REFERENCES

[1] P. M. Mather. 1999. Computer Processing of Remotely Sensed Images. John Wiley, UK.

[2] D.A. Landgrebe. 2003. Signal Theory Methods in Multispectral Remote Sensing, Wiley, New York.

[3] P. M. Atkinson, Tattnall.997. Neural network in remote sensing. Int. J. Remote Sensing 18 (4) 699-709.

[4] J. A. Benediktsson, Swain P. H., Erosy O. K..1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data. IEEE Transactions on Geosciences and Remote Sensing 28 540-552.

[5] G. M. Foody, M. K. Arora. 1997. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. Int. J. Remote Sensing 18 (4) 799–810.

[6] H. M. Chee, O. K. Ersoy. 2005. A statistical self-organizing learning system for remote sensing classification. IEEE Transactions on Geosciences and Remote Sensing 432 (8) 1890–1900.

[7] M. Acharyya, R. K. De, M. K. Kundu. 2003. Segmentation of remotely sensed images using wavelet features and their evaluation in soft computing framework. IEEE Transactions on Geosciences and Remote Sensing 41 (12) 2900-2905.

[8] B. Mannan Joy, A. K. Ray. 1998. Fuzzy ARTMAP supervised classification of multispectral remotely sensed images. Int. J. Remote Sensing 19 (4) 767-774.

[9] Saroj K. Meher, B. Uma Shankar, and Ashish Ghosh. 2007. Wavelet Feature Based Classifier for emotely sensed Images. IEEE Transactions on Geosciences and Remote Sensing 45 (06) 1881-1886.

[10] L. O. Jimenez, D. A. Landgrebe.1998. Supervised classification in high dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data. IEEE Trans. Syst., Man, Cybern. A, 28, 39–54.

[11] A. Jain, D. Zongker. 1997. Feature selection: Evaluation, application, and small sample performance. IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 2, pp. 153–158.

[12] J. Kennedy, R. C. Eberhart. 2001. Swarm Intelligence. Morgan Kaufmann, San Mateo.

[13] R.C. Gonzalez, R. E. Woods, S. L. Eddins. 2002. Digital Image Processing using MATLAB, Pearson, Singapore.

[14] S. Haykin. 1997. Neural Networks: A Comprehensive Foundation, Prentice Hall, England.

[15] S. K. Pal, A. Ghosh, and B. U. Shankar.2000. Segmentation of remotely sensed images with fuzzy thresholding and quantitative evaluation. Int. J. Remote Sens., vol. 21, no. 11, pp. 2269–2300.

[16] X.L. Xie, G. Beni. 1991. A validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (8), 841–847.

Rajesh K. Agrawal is currently pursuing his Ph.D. degree in the Department of Electronics Engineering at G.H. Raisoni College of Engineering, Nagpur, India. He received the B.E. degree from University of Poona, Pune, India in 1992 and M.E. degree from the Kolhapur University, Kolhapur, India, in 2007, all in electronics engineering. He has total teaching experience of more than 17 years. His research interest includes signal processing, soft computing and ANN.

Dr. Narendra G. Bawane is currently working as a Principal, S. B. Jain Institute of Technology, Management & Research, Nagpur, India. He was former Head of Computer Science and Engineering department at G. H. Raisoni college of Engineering, Nagpur & also worked with B.D. College of Engineering, Sewagram and Govt. Polytechnic, Nagpur. He has total teaching experience of more than 23 years. He has completed his B.E. from Nagpur University in 1987 and M. Tech. in 1992 from IIT, New Delhi. He completed his Ph. D. in 2006 at VNIT, Nagpur. His areas of interest are: Artificial Neural Network (ANN), Wavelet analysis, Image processing & Emotion in speech and facial recognition and Hybrid Intelligence.