A GENETIC ALGORITHM BASED FEATURE EXTRACTION TECHNIQUE FOR HYPERSPECTRAL IMAGERY

Byong Tae Ryu*, Choon-Woo Kim*, Hakil Kim*, Kyu Sung Lee**

School of Information and Communication Eng.*, Dept. of Geoinformatic Engineering**,
Inha University, Incheon, Korea
g2041420@inhain.net, {cwkim, hikim, ksung}@inha.ac.kr

ABSTRACT:

Hyperspectral data consists of more than 200 spectral bands that are highly correlated. In order to utilize hyperspectral data for classification, dimensional reduction or feature extraction is desired. By applying feature extraction, computational complexity of classification can be reduced and classification accuracy may be improved. In this paper, a genetic algorithm based feature extraction technique is proposed. Measure from discriminant analysis is utilized as optimization criterion. A subset of spectral bands is selected by genetic algorithm. Dimension of feature space is further reduced by linear transformation. Feasibility of the proposed technique is evaluated with AVIRIS data.

KEY WORDS: Hyperspectral, Feature extraction, Classification, Genetic algorithm

1. INTRODUCTION

Multispectral data provides relatively low spectral resolution. However, new hyperspectral sensors make it possible to generate data whose dimension is over 200. Detailed spectral information is major advantage of hyperspectral data. However, abundance of spectral information results in huge increase of data size and increase in computational complexity and memory resources (Landgrebe, 2003; Chang, 2003).

Recently, various works on hyperspectral image have been reported. They can be divided into preprocessing such as radiometric and atmospheric corrections, feature extraction, classification, data compression, and anomaly detection, etc. Among these research areas on hyperspectral imaging, this paper is focused on feature extraction for classification. Feature extraction for classification represents a process of dimension reduction without deteriorating classification accuracy. Needs for feature extraction of the hyperspectral data arise from the following facts; neighboring bands of hyperspectral data are highly correlated. This indicates that there is a room for dimension reduction. Due to the huge size of hyperspectral data, it is not practical (in terms of computational complexity and memory resources) to process original data without any dimension reduction. This is critical specially for the applications that require on-line classification. Also, often only a limited number of training samples are available. It may cause singularity problem for calculating inverse matrices that are frequently used for classifier.

The first step of feature extraction is to select a criterion for optimization. Ideally, classification accuracy should serve as a criterion for feature extraction to maintain or increase classification accuracy. However, it is very difficult to formulate the classification accuracy as a function of spectral bands even with a simple form of classifier. As alternatives, various measures have been utilized as optimization criteria for feature extraction. They include Bhattacharayya distance, Jeffreys-Matusita distance, optimum index factor, Sheffield index, and divergence, etc. (Fukunaga, 1990).

Reduction of dimension can be achieved by selecting a subset of spectral bands. Thus, it is often called as band selection (Zheng et.al., 1995; Chang et. al., 1999). In this case, feature extraction can be regarded as search problem. Instead of band selection, linear combination of bands with the reduced dimension can be determined. This method is preferred because band selection is a special case of the linear combinations.

Among various optimization techniques that can be employed for feature extraction, a genetic algorithm is chosen as optimization tool in this paper. Recently, genetic algorithm has been widely utilized for remote sensing applications. In (Yao and Tian, 2003), genetic algorithm was applied for feature extraction of the hyperspectral data. Correlation coefficients between the ground reference and extracted features were used for fitness functions. Also, genetic algorithm was applied for the anomaly detection in hyperspectral image (Li et. al, 2005) and the estimation of regularization mixing parameters for feature extraction (Kuo et. al., 2005).

In this paper, feature extraction is performed in two steps. A subset of spectral bands is selected by genetic algorithm. The criterion of DAFE(discriminant analysis feature extraction) (Landgrebe, 2003) is chosen for optimization. Dimension of feature space is further reduced by linear transformation. Feasibility of the proposed technique is evaluated with 10 land cover classes from AVIRIS data having 220 bands.

2. FEATURE EXTRACTION BASED ON GENETIC ALGORITH

2.1 Optimization Criterion

In this paper, criterion of DAFE is chosen for optimisation. Optimization problem can be stated as follows;

$$\text{Maximize } J = tr(S_w^{-1} S_b) \tag{1}$$

where S_w represents within-class scatter matrix. It is defined as

$$S_w = \sum_{i=1}^{L} P_i E\left[(X - M_i)(X - M_i)^T | w_i\right] \tag{2}$$

In Eq (2), L is the number of classes to be classified. P_i denotes priori probability for class i and w_i represents *ith* class. M_i is the mean vector of class i. Also, X represents feature vector. $E[\,]$ denotes expectation operator. In Eq. (1) S_b is the between-class scatter matrix and defined by the following Eq. (3).

$$S_b = \sum_{i=1}^{L} P_i (M_i - M_0)(M_i - M_0)^T \tag{3}$$

where M_0 represents the mean vector of mixture distribution. It is calculated as

$$M_0 = \sum_{i=1}^{L} P_i M_i \tag{4}$$

Suppose X is original vector of size nx1. New feature vector Y of size mx1 where m < n is calculated by the following Eq. (5).

$$Y = A^T X \tag{5}$$

where A^T is mxn matrix representing linear transformation. Criterion J in Eq. (1) can be optimized by selecting the largest m eigenvalues of $S^{-1}{}_w S_b$. The corresponding eigenvectors form the transformation matrix A (Fukunaga, 1990).

2.2 Genetic Algorithm based Optimization

Figure 1 shows the flowchart of genetic algorithm that optimizes the criterion of DAFE in Eq (1). String for genetic algorithm is a vector of size zx1 where z represents the number of spectral bands in hyperspectral image. Its components are either 0 or 1. Number of components having binary value 1 is n. Remaining (z-n) components have the binary value of 0. Strings in initial population are generated randomly. For each string, the value of DAFE criterion is calculated.

The reproduction operator mimics the Darwinian survival of the fittest in the theory of evolution. The objective of the reproduction operator in this study is to increase the value of the DAFE criterion. Assume that *J(k)* is the calculated value of criterion for the *kth* string. Reproduction probability of the *kth* string is calculated by the following Eq. (6)

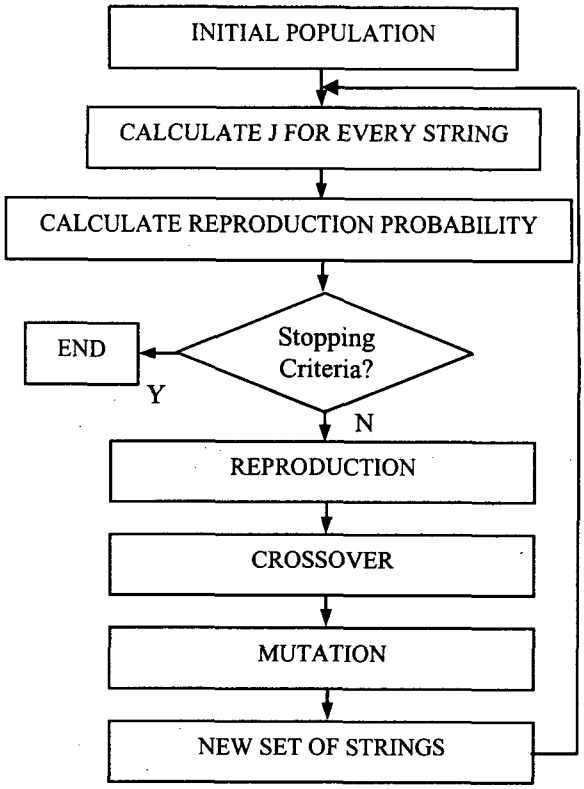

Figure 1. Flowchart of Proposed Feature Extraction

$$prob(k) = \frac{J(k)}{\sum_{i=1}^{P} J(i)} \tag{6}$$

where P is number of strings. Eq. (6) represents the probability of survival of the *k*th string for the next generation or iteration. It indicates that the string with large value of the criterion will have a higher survival probability. Strings for next generation are generated by Russian roulette approach (Goldberg, 1989).

Crossover is performed after reproduction. Crossover operation employed in this study can be explained by an example shown in Figure 2. Assume that a pair of strings A and B is chosen for crossover. Arrow shown in Figure 2 shows crossover position that is selected randomly. Binary values on the left side of crossover position are swapped. After crossover operation, value of 0 or 1 are exchanged so that only n components have the value of 1.

Figure 3 shows an example of mutation operation. In Figure 3, arrows represent positions of mutation that are selected randomly. Two positions, one with the value of 0 and the other with the value of 1, are selected. Values of the selected positions are exchanged.

After three genetic operators are applied to the initial population, a new population with the same number of strings is constructed. The aforementioned procedure is repeated until the stopping condition is satisfied. Number of iterations or difference in the sum of calculated criterions for two consecutive iterations can serve as stopping conditions. Once the stopping condition is satisfied, new features are calculated by Eq. (5).

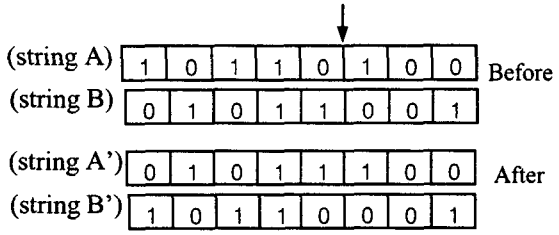

Figure 2. Example of Crossover Operation

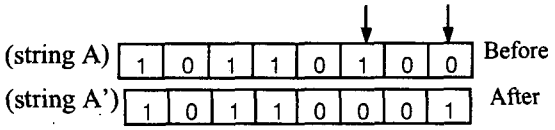

Figure 3. Example of Mutation Operation

3. CLASSIFICATION EXPERIMENTS

3.1 Hyperspectral Image used in experiment

In order to evaluate the performance of the described feature extraction technique, classification experiments are performed. Hyperspectral image used in experiments has 220 spectral bands. It is from a June 1992 AVIRIS data of a mixed agriculture/forestry landscape in the Indian Pine Test site in Northwestern Indiana. It is called as 92AV3C and its size is 145x145 (Landgrebe, 2003). Among 15 classes of ground truth data, 10 classes are utilized for experiments. For each class, 64 samples are chosen for training and testing. Boundaries of the chosen samples are marked as squares on the reconnaissance map in Figure 4 (Landgrebe, 2003).

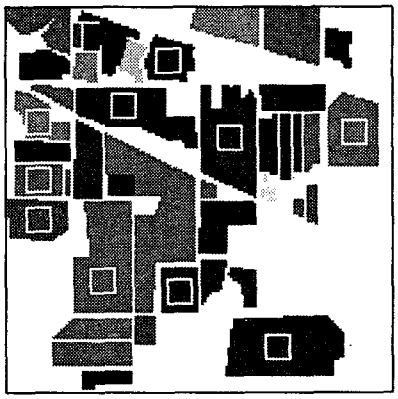

Figure 4. Location of Training/Testing Samples

3.2 Feature extraction

Number of strings for initial population, P in Eq. (6), was 100. Sting is of size 220x1 (z=220) that is the same as the number of spectral bands. In this study, twenty components (n=20) of the string have binary value of 1 and the rest (z-n=200) with the value 0. Locations for binary value 1 are selected randomly. It needs to be mentioned that the value of n should be determined based on the number of spectral bands and number of available training/testing samples to avoid possible singularity problems. For each string, 20 spectral bands corresponding to binary value 1 are extracted and formed an original feature vector X of size 20x1. Based on the ground truth data, the mean vector M_i for each class is calculated. Since pairwise Gaussian classifier is employed in classification experiments, feature extraction is performed for each pair of 10 classes. In other words, genetic algorithm shown in Figure 1 is performed ${}_{10}C_2$=45 times. For each of ${}_{10}C_2$ class pairs, M_0, S_w and S_b are calculated. In Eqs. (2)-(5), L is 2. Size of M_0 is 20x1. S_w and S_b are square matrices of size 20x20.

When difference in the sum of calculated criterions for two consecutive iterations is less than a predetermined threshold value, iteration is stopped. Once iteration is stopped, 20 spectral bands yielding the largest value of the criterion are determined. The selected 20 spectral bands would form the optimum subset for a given class pair among ${}_{220}C_{20}$ possible combinations. Dimension of feature space can be further reduced by the linear transformation specified in Eq. (5).

In this study, new feature vector Y of size 5x1 is calculated by Eq. (5). Among 20 eigenvalues of $S^{-1}{}_wS_b$., the largest 5 eigenvalues are selected. The corresponding eigenvectors form the transformation matrix A in Eq. (5). By the described technique, dimension of feature space is reduced from 220 to 5.

3.3 Pairwise classification

Once 20 spectral bands and their linear transformation matrix are determined for each of ${}_{10}C_2$ class pairs, classification experiments are performed. It is assumed that new feature vectors obey Gaussian distribution. Pairwise Gaussian classifier between class i and j is constructed. It can be formulated as Eq. (7).

$$\frac{1}{2}(Y-M_{Yi})^T\Sigma_i^{-1}(Y-M_{Yi})-\frac{1}{2}(Y-M_{Yj})^T\Sigma_j^{-1}(Y-M_{Yj})$$
$$+\frac{1}{2}\ln\frac{|\Sigma_i|}{|\Sigma_j|}<\ln\frac{P_{Yi}}{P_{Yj}} \qquad (7)$$

where Y is the new feature vector of size 5x1. M_{Yi} and M_{Yj} represent the mean vector of class i and j, respectively. Σ_i and Σ_j denote covariance matrix for class i and j. P_{Yi} and P_{Yj} are priori probabilities. In this study, the number of training/testing samples is 64 for all of 10 classes. Thus, $P_{Yi}=P_{Yj}=0.5$.

Leave-one-out method is employed for classification (Fukunaga, 1990). When feature vector Y of class i is chosen as a testing sample, pairwise classifier between class i and j is constructed using 63 samples from class i excluding the test sample and all of 64 samples from class j. Table 1 lists result of classification experiments when 5x1 feature vectors are utilized. The first column of Table 1 can be interpreted as follows; Among 64 samples of class A(Corn-notill), 62 samples are correctly classified as class A. However, two samples of class A are misclassified, one as class H and the other one as class J. Table 2 shows land coverage associated with

class symbols and classification accuracy for each class. Classes B(Grass/Pasture), C(Grass/Trees), D(Hay-windrowed), and G(Woods) yields the highest classification accuracy of 100%. However, classification for class E(Soybeans-notill) is the smallest value of 84.4%. Overall classification accuracy is 94.7% when feature vectors of 5x1 are utilized in the experiments.

In order to evaluate the effect of feature space reduction by linear transformation, 20x1 feature vectors selected by genetic algorithm are also utilized for the same pairwise classification. In this case, a subset consisting of 20 spectral bands is utilized instead of linear transformation of the spectral bands. Overall classification accuracy for 20x1 vectors is 69.8%. When calculating the inverse of 20x20 covariance matrix in Eq. (7), singularity problems are faced. However, there was no singularity problem when 5x1 feature vectors were utilized. Two experimental results indicate that even with the reduced dimension of 5, high classification accuracy of 94.7% can be achieved by the proposed technique.

Table 1. Results of Pairwise Classification Experiments with 5x1 Feature Vector

	A	B	C	D	E	F	G	H	I	J
A	62	0	0	0	5	4	0	2	0	1
B	0	64	0	0	0	0	0	0	0	0
C	0	0	64	0	0	0	0	0	0	0
D	0	0	0	64	0	0	0	0	0	0
E	0	0	0	0	54	0	0	2	0	0
F	0	0	0	0	0	58	0	0	0	0
G	0	0	0	0	0	0	64	0	0	0
H	1	0	0	0	3	2	0	56	1	4
I	0	0	0	0	0	0	0	0	63	2
J	1	0	0	0	2	0	0	4	0	57

Table 2. Classes and Classification Accuracy

Class Symbol	Class	Classification Accuracy (%) with 5x1 vectors
A	Corn-notill	96.9
B	Grass/Pasture	100.0
C	Grass/Trees	100.0
D	Hay-windrowed	100.0
E	Soybeans-notill	84.4
F	Soybeans-min	90.6
G	Woods	100.0
H	Corn-min	87.5
I	Corn	98.4
J	Soybeans-clean	89.1

4. CONCLUSION

In this paper, feature extraction is performed in two steps. A subset of spectral bands is first selected by genetic algorithm. In this way, it is possible to avoid singularity problem that might happen when only a limited number of samples are available. The criterion of DAFE is chosen for optimization. The dimension of feature space is further reduced by the linear transformation. Feasibility of the proposed technique is evaluated with 10 land cover classes from AVIRIS data having 220 bands. Experimental results indicate that 5x1 feature vectors generated by the proposed technique yields 94.7% overall classification accuracy.

References

Landgrebe, D.A. 2003 *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley and Sons, New Jersey.

Chang, C-I. 2003 *Hyperspectral Imaging : Techniques for spectral detection and classification*, Kluwer Academic, New York

Fukunaga, K. 1990 *Introduction to Statistical Pattern Recognition*, Academic Press, London

Zheng, W. et. al., General and Specific Methods Studying on Bands Selection of Hyperspeectral Remote Sensing Data, *IEEE Int. Geoscience and Remote Sensing Symposium*, Seoul, Korea, pp. 3223-3226.

Chang, C-I., et. al., 1999 A Joint Band Prioritization and Band-Decorrelation Approach to Band Selection for Hyperspectral Image Classification, IEEE Trans. On Geoscience and Remote Sensing, 37(6), pp. 2631-2641.

Yao, H. and Tian, L., 2003 A Genetic-Algorithm-Based Selective Principal Component Analysis(GA-SPCA) Method for High-Dimensional Data Feature Extraction, IEEE Trans. On Geoscience and Remote Sensing, 41(6), pp. 1469-1478.

Li, N. et. al.,2005 Independent Component Analysis Based on Improved Quantum Genentic Algorithm: Application in Hyperspectral Images, *IEEE Int. Geoscience and Remote Sensing Symposium*, Seoul, Korea, pp. 4323-4326.

Kuo, B.C., et.al., 2005 Estimating Mixing Parameters of Regularized Feature Extractions by Genetic Algorithm, *IEEE Int. Geoscience and Remote Sensing Symposium*, Seoul, Korea, pp. 3902-3904.

Goldberg D.E., 1989 *Genetic algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, New York.

Acknowledgements

This research was supported by the ADD, Korea through the Image Information Research Center at KAIST.