On Combining Genetic Algorithm (GA) and Wavelet for High Dimensional Data Reduction

Zhengjun Liu
Chinese Academy of Surveying and Mapping, Beijing 100039, China
liuzj@lab.irsa.ac.cn

Changyao Wang
LARSIS, the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101, China

Jixian Zhang
Chinese Academy of Surveying and Mapping, Beijing 100039, China

Qin Yan
Chinese Academy of Surveying and Mapping, Beijing 100039, China

Abstracts: In this paper, we present a new algorithm for high dimensional data reduction based on wavelet decomposition and Genetic Algorithm (GA). Comparative results show the superiority of our algorithm for dimensionality reduction and accuracy improvement.

Keywords: Feature Extraction, Wavelet, Genetic Algorithm

1. Introduction

Classification and pattern recognition of high dimensional remote sensing data are distinctly different from traditional multi-channel remote sensing classification techniques due to the Hughes phenomena [1]. Thus, the exploitation of reliable and efficient feature extraction techniques for high dimensional data has been one of the most desirable research topics in remote sensing pattern classification and recognition fields [2, 3, 4], in which one of the most widely used dimension reduction techniques include principal component analysis (PCA), Discriminant Analysis Feature Extraction (DAFE) [5], Decision Boundary Feature Extraction (DBFE) and Wavelet dimensionality reduction.

The use of wavelet transform in dimensionality reduction has recently been an interesting topic in hyperspectral remote sensing study. Bruce *et al.* developed an automated classification system based on wavelet transform and linear discriminant analysis [4]; Sinthop *et al.* suggested an automatic decomposition level selection and wavelet reduction scheme based on spectral correlationship between the original spectrum and the reconstructed spectrum from specific decomposition level [6]. However, these techniques are normally based on the intrinsic statistical characteristics, e.g. energy or entropy, from the training data and thus independent from the classification process. Therefore, the choose of features from feature space at the same or different decomposition level is data adaptive rather than both data and methodology adaptive.

This paper investigates issues on methodology and applications of feature extraction and image classification for high dimensional remote sensing data. A new feature extraction algorithm based on Genetic Algorithm and wavelet transform is proposed for high dimensional data reduction and classification accuracy improvement. The proposed algorithm combines the advantages of GA's global optimization and wavelet's multi-resolution and multi-scale analysis.

2. Methodology

1) Wavelet decomposition and Wavelet Packet Decomposition

The principle of our method is to apply a Discrete Wavelet Transform (DWT) or Wavelet Packet Analysis (WPA) to hyperspectral data in the spectral domain and at each pixel location [4, 6]. In this way, the DWT or WPA decompose the original signals into subband of approximate signals and detail signals at different scale, characterized by its approximate coefficients and detail coefficients. Feature extraction and classification are thus based on these coefficients.

For the wavelet transform, we choose the 2nd order Daubechies function as wavelet basis; for WPA, the 2nd order Daubechies wavelet basis and the Shannon entropy criteria are used.

2) Feature selection by genetic algorithm

Genetic Algorithms offer an efficient search method for a complex problem space and can be used as a powerful feature selection tool [7].

The use of genetic algorithm for feature space feature selection consists of three major phases. The first phase is to decide the representation of band (feature) combination, i.e., whether we use a binary strings form or directly use a string consisting of integer numbers to represent the corresponding feature identifications; The second step is the evaluation on the fitness of these feature combinations, this is based on estimating the class separability of the current feature combination through decoding each genome and computing its fitness function; The third one is applying the evolutionary process such as selection, crossover, and mutation operations by a genetic algorithm according to its fitness. The evolution stops when the iteration is greater than a predefined number or the population has converged. This procedure could be shown as Figure 1.

Fitness scores in GA are calculated and evaluated based on Jeffries-Matusita distance of the selected training samples. The mathematical expression of J-M distance could be formulated as Eqs. (1), (2), and (3)：

$$B_{ij} = 1/2\left(M_i - M_j\right)^T \left[\frac{\left(V_i + V_j\right)}{2}\right]^{-1}\left(M_i - M_j\right) + 1/2\ln\frac{\left|\left(V_i + Vj\right)/2\right|}{\sqrt{\left|V_i\right|\left|V_j\right|}} \tag{1}$$

$$JM_{ij} = \sqrt{2\left(1 - e^{-B_{ij}}\right)} \tag{2}$$

$$JM = \frac{2}{n(n+1)}\sum_{i=1}^{n-1}\sum_{i=j+1}^{n}(p_i \cdot p_j \cdot JM_{ij}) \tag{3}$$

Where V_i and V_j are the covariance matrix between class i and class j respectively, while M_i and M_j are the correspondent mean vectors of class i and j, B_{ij} is well known as the Bhattacharyya distance，p_i、p_j are the prior probabilities for class i and j, JM_{ij} is the J-M distance between class i and j, and JM is the average J-M distance. The superscript T denotes matrix transpose operation.

For GA, one population with 500 chromosomes was initialized. The maximum iteration number was set to 500; crossover probability was set to 0.9; mutation probability was set to 0.1.

3. Experimental Results

The AVIRIS hyperspectral data are from LARS laboratory, Purdue University, accompanied with the ground truth land cover map [8]. There are 16

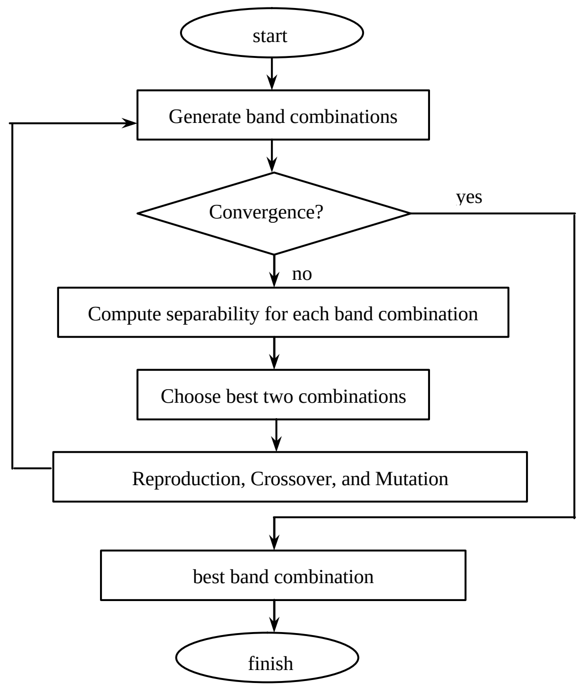

Fig.1. Optimal feature selection based on genetic algorithm

classes of land cover types spread over this area. The data are classified with maximum likelihood classifier (MLC). The classification performances of various feature extraction techniques applied on AVIRIS data at different number of features are shown in Figure 2. Experimental results show that the use of DWT and GA-based feature extraction technique improves the overall classification accuracy by 1.1-6.5%, as compared to the use of original hyperspectral signals or conventional feature extraction techniques, such as PCA, DAFE and DBFE.

We also investigate the performance of combination of GA with Wavelet Packet Analysis (WPA). Preliminary results shows that combining GA and WPA does not significantly improved classification accuracy other than the combination of GA and wavelet transform when selective limited band number are chosen. However, when increased number of bands are chosen, the overall classification accuracy becomes more evident, as shown in Figure 2.

4. Conclusions

In this paper, we have presented a new feature extraction technique based on wavelet analysis and genetic algorithm. Analytical assessment and experimental results of classification accuracy have proven that the effectiveness of the algorithm for high dimensional data reduction. Besides, the proposed feature extraction technique also provides an approach to design an automatic classification system for high dimensional remote sensing data.

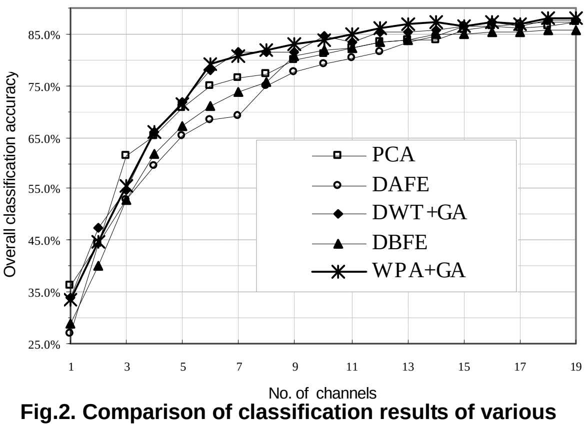

Fig.2. Comparison of classification results of various feature extraction algorithms applied on AVIRIS data

References:

[1] Hughes, G.F., 1968. On the Mean Accuracy of Statistical Pattern Recognizers. *IEEE Trans. Inform. Theory* IT-14. pp. 55-63.

[2] Lee C. and D. Landgrebe 1993. Feature Extraction and Classification Algorithms for High dimensional Data, TR-ENVIRONMENT 93-1, Purdue University.

[3] Hsu, Pai-Hui, Tseng Yi-Hsing, 1999. Feature Extraction for Hyper Spectral Image. Asian Conference On Remote Sensing , November 22 - 25, 1999, Hong Kong, China

[4] Bruce, L. M., Koger C. L., and Li J., 2002. Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction. *IEEE Transaction on Geoscience and Remote Sensing*, vol.40(10): pp.2331-2338.

[5] Richards, J.A., 1997. *Remote Sensing Digital Image Analysis: An introduction*, Springer-Verlag Berlin Heidelberg, Third Edition.

[6] Sinthop, K., Jacqueline, L. M., and Tarek, E-G, 2003, Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Analysis. *IEEE Transaction on Geoscience and Remote Sensing*, vol.41(4): pp.863-871.

[7] Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, New York.

[8] **URL:** 220-band AVIRIS data set (92AV3C). Available at：ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C