Article ID: 1007-4619 (2005)02-0176-10

Feature Extraction and Feature Selection Based on Wavelet and Genetic Algorithm

LIU Zheng-jun[1], WANG Chang-yao[2], ZHANG Ji-xian[1]

(1. *Chinese Academy of Surveying and Mapping*, *Beijing* 100039, *China*;
2. *Laboratory of Remote Sensing Information Science*, *Institute of Remote Sensing Applications*, *Chinese Acadimy of Sciences*, *Beijing* 100101, *China*)

Abstract: Classification and pattern recognition of high dimensional remote sensing data are distinctly different from traditional multi-channel remote sensing classification techniques. In this paper, a newly integrated feature extraction algorithm based on GA and wavelet/wavelet packet (WP) transform is proposed for high dimensional data reduction and classification. The proposed algorithm combines the advantages of GA's global optimization and wavelet's multi-resolution and multi-scale analysis. Hyperspectral signals are firstly transformed to feature domain by using a discrete wavelet or wavelet packet decomposition strategy. Since the discrete wavelet transform (DWT) is a linear transform, the DWT coefficients at specific scales could be directly used as linear features. Followed by the decomposition phase is optimal feature subset selection, in which the optimal feature subset acquired the best divergence is obtained according to interclass/intraclass distance of the training samples. This procedure is implemented by a Genetic Algorithm, with each possible feature subset encoded as chromosome. Fitness scores in GA are calculated and evaluated based on Jeffries-Matusita distance of the selected training samples. Hyperspectral data are classified with maximum likelihood classifier (MLC). Experimental results show that the use of DWT/WP and GA-based feature extraction technique improves the overall classification accuracy by 1.1%—6.5%, as compared to the use of conventional feature extraction techniques, such as principal component analysis (PCA), Discriminant Analysis Feature Extraction (DAFE) and Decision Boundary Feature Extraction (DBFE).

Key words: feature extraction; wavelet and wavelet packet; genetic algorithm

CLC number: TP751.1 **Document code**: A

1 INTRODUCTION

The continuous availability of airborne and space-borne high-resolution spectral remote sensing data provided by the hyperspectral sensors, especially the successful launch of EO-1 satellite with the on board HYPERION sensor[1,2], has greatly promoted the potential application of hyperspectral remote sensing techniques for natural resources management, environment assessment, and precision agriculture[3]. However, 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[4]. In addition, far more training samples than traditional multi-spectral classification are needed to accurately estimate those statistical variables that are used to describe pattern properties existed in feature space regarding to the abrupt increase of spectral dimensionality. 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[5—8]. The commonly used dimension reduction techniques include princi-

Received date: 2004-03-19;

Foundation item: International Cooperation Key Project of Ministry of Science and Technology (2001DFBA0005); the China's Special Funds for Major State Research Project (G2000077900).

Biography: Liu Zhengjun (1974—), male, Doctor and Associate Professor. Research interests: remote sensing of environment, feature extraction, and high resolution image processing. Email: zjliu@casm.ac.cn

pal component analysis (PCA)[9], Discriminant Analysis Feature Extraction (DAFE)[10], Decision Boundary Feature Extraction (DBFE)[5].

Recently, wavelet transform has been an interesting topic in dimensionality reduction study. The multi-resolution analysis (MRA) of the discrete wavelet transform (DWT) results of hyperspectral signals provide the direct insights into both global and fine information in hyperspectral signals, making feature extraction at various scales becoming possible. Bruce *et al.* developed an automated classification system based on wavelet transform and linear discriminant analysis (LDA)[7,8]; the experimental results of this technique on ground-based ASD FieldSpec Pro FR portable spectroradiometer data demonstrate a potential application on hyperspectral data classification and dimensionality reduction. Kaewpijit *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[11]. This is applied on three different AVIRIS scene data. The comparative results with PCA show that DWT has an efficient capability for dimensionality reduction and classification accuracy improvement.

However, these techniques are normally based on the intrinsic statistical property, e. g. energy or entropy, from the training data and thus are independent from the classification process. Therefore, the selection of feature subset from feature space at the same or different decomposition level is data oriented rather than target/object oriented.

This paper focuses 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, providing the flexibility of identifying the target classes using the most useful information from subset of transformed features rather than from all features.

The paper is organized as follows. Section II provides an overview of the wavelet based feature extraction for dimension reduction of hyperspectral data, and the genetic algorithm for feature subset selection. Meanwhile, the framework that combines GA and wavelet's MRA for automatic feature extraction is described in detail. Section III concerns the experimental results of applying the described methodology to AVIRIS data for land cover classification. This is accomplished by investigating the impact of the GA optimized wavelet reduction on classification accuracies, as compared with other feature extraction techniques. Section IV provides our concluding remarks for this work.

2 METHODOLOGY

2.1 Multiresolution Analysis with Wavelet Transform and Wavelet Packet Transform

2.1.1 Overview of wavelet decomposition

Wavelet theory[12] is a sophisticate and powerful mathematical technique that has been widely used in remote sensing application, e. g., data compression, fusion, classification, and multiresolution analysis (MRA). It takes a signal in time domain, and represents it in frequency domain. For remote sensing applications, these signals are normally referred to as spectrum curves which represent the spectral signature of the ground objects. Projecting the hyperspectral signal at each pixel location onto a basis of wavelet functions called "mother wavelet, $\psi(x)$" can separate the fine-scale and large-scale information of this signal. Thus, extracting appropriate features from various information at various scales provides a potential to effectively discriminate hyperspectral signals and recognize specific patterns in feature space, according to Bruce's research[7,8].

Theoretically, once the mother wavelet $\psi(x)$ is fixed, wavelets can be obtained by dilation a and shifts b of the mother wavelet $\psi\left(\frac{x-b}{a}\right)$, where factors a and b can be any value in the continuous wavelet transform. But usually for convenience, the values for a and b is defined as dyadic: $a=2^j$ and $b=k\cdot 2^j$, where k and j are integer values, then the space is sparse. Thus the wavelet basis functions can be mathematically defined as

$$\psi_{a,b}(x)=\frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)=2^{-\frac{j}{2}}\psi(2^{-j}x-k)=\psi_{j,k}^{*}(x) \quad (1)$$

where 2^j and k represents the scaling factor and the translation factor, respectively. $\psi_{j,k}^{*}(x)$ represents the discrete form of basis functions that are linear-independent and orthogonal.

The wavelet transform makes any function as a summation of wavelets. It can be expressed as

$$f(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{a^2} W(a,b)\psi_{a,b}(x)\,\mathrm{d}a\mathrm{d}b \qquad (2)$$

for continuous wavelet transform, and

$$f(x) = \sum_{j=0}^{N}\sum_{k=0}^{N} W(k,j)\psi_{j,k}^{*}(x) \qquad (3)$$

for discrete wavelet transform (DWT). Where for our research purpose, $f(x)$ corresponds to a hyperspectral signal, x is the spectral band or channel number which are equally spaced with equal bandwidth. The respective weights $W(k, j)$ of the wavelets in the summation corresponding to the continuous form $W(a, b)$ are called the wavelet coefficients, which are a measure of the intensity of the local variations for the scale under consideration.

After the mother wavelet is selected, the wavelet transform can be used to decompose a signal into approximation and detail information according to scale. This is typically done iteratively using the dyadic discrete wavelet transform, in which scales $1, 2, 4, \cdots, 2^j$ are utilized.

In practice, the dyadic discrete wavelet transform can be implemented through a binary tree algorithm called Mallat algorithm with highly computational performance when the number of input bands is an integer power of two[13]. In common cases, the spectral length or feature number is usually not an integer power of two, thus a signal extensions technique called zero padding is involved and applied to the original spectral bands[14].

2.1.2 Wavelet packet decomposition

The concept of Wavelet Packet (WP) is proposed by Wickerhauser[15] and Coifman *et al.* [16]. They generalized the linkage between multiresolution analysis and wavelets, and introduced wavelet packets. The basic difference between wavelet and wavelet packet is that, when the wavelet decomposition is computed, we only use one branch of a binary tree. To get better frequency localization we need to divide the detail leaves within parts of higher resolution, so that we can represent the details with even smaller details.

Coifman, Meyer and Wicherhauser[16] found that a conjugated mirror filters can transform an orthogonal wavelet basis in two orthogonal families. By iteratively applying this rule, we can in fact divide a signal f into a complete binary tree[17]. In summary, the procedure of decomposing a signal f into a complete wavelet packet binary tree is in fact a further extension of the concept of DWT decomposition process, different in that WPA decomposes every component, including both of the approximation and the detail portion rather that only the approximation signal, into its approximation and detail portion at a deeper level. In this way, the original spectral information could be more specifically decomposed into more components of signal precisely at different scales or frequencies, until finally a wavelet packet binary tree is established, thus providing the possibility of extracting exactly the information helpful for target recognition from this binary tree.

2.2 Feature Selection by Genetic Algorithm

Genetic Algorithms (GA) offer an efficient search method for a complex problem space and can be used as a powerful feature selection tool[18,19].

GA was first proposed by Holland in 1975[20], who was inspired by the natural selection of biology. It was supposed that individuals with more adaptability in current generation would have better capability of survival and breed in the next generation. One of the most important advantages of GA is that it can make use of the limited search processes to achieve the optimal or near-optimal result. The basic GA process can be shown in Fig. 1. In GA, each possible solution is represented by an individual of the population, namely a chromosome. Each chromosome is composed by several genes. Each gene could be expressed by using a specific encoding strategy, e. g., the Hamming encoding strategy. A Simple GA (SGA) produces the new generation of offspring by its reproduction, crossover and mutation operators[18]. The reproduction operation is simply to copy the current generation of chromosomes as one part of the next generation of individuals, and the number of reproduction operation is generally determined by the individual's adaptability and the selection strategy. The crossover operation is adopted to create a new individual that combines the best features from its parent chromosomes by recombining their genes at random. The mutation operation is to randomly replace one encoded gene by another based on a given probabilities. In this way, new trial solutions are created by making small, random changes in the representation of prior trial solutions. Good genetic operators can avoid convergence into local maximum/minimum value and approach to a globally optimal solution generation by generation.

The framework of optimal feature selection of a hyper-

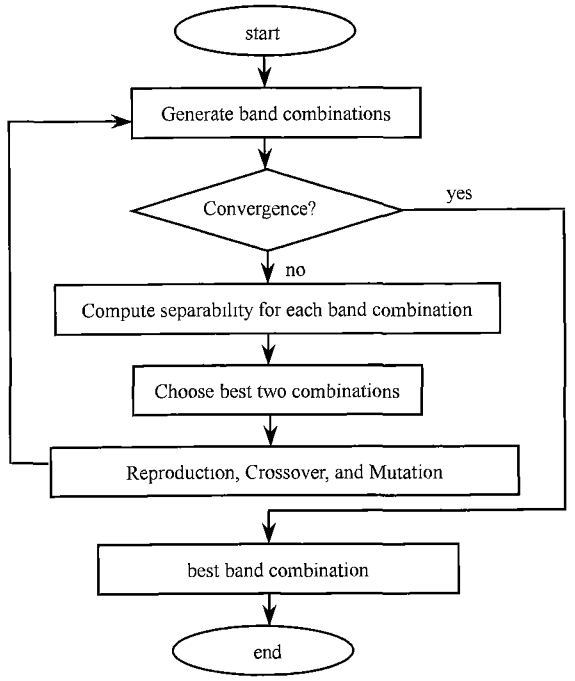

Fig. 1　Optimal feature selection based on genetic algorithm

spectral remote sensing image based on GA proposed in this paper is shown in Fig. 1. This general procedure can also be divided into 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. For example, if we use eight of the 220 AVIRIS channels as a choice of the channel selection, the chromosome length is eight, which means that a total of eight genes are used to represent the feature permutation. The value of each gene is an integer number which represents band number at the range of [1, 220]. Meanwhile, it is assumed that each gene is not equal to any other genes existed in this chromosome. Fig. 2 shows one case of this example.

3	18	7	153	22	64	138	201

Fig. 2　Illustration the encoding of a simple chromosome with length 8

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 weighted J-M distances for each feature combination (chromosome) are calculated. Based on the weighted J-M distances, the fitness of each chromosome is evaluated, and dominant individuals are selected for genetic operation according to the selection strategy, e. g., the Roulette Wheel Selection.

The third step actually performs the reproduction, crossover and mutation operators for producing the new generation of individuals according to its fitness. This evolutionary process continues until the population at a certain generation has converged or the maximal number of iterations is reached. The decision of population convergence is based on evaluate the fitness scores of each individual, as described below.

Fitness score of each individual is calculated and evaluated based on J-M distance of the selected training samples. The mathematical expression of J-M distance could be formulated as Eqs. (4), (5), and (6):

$$B_{ij} = 1/2(M_i - M_j)^{T}\left[\frac{(V_i + V_j)}{2}\right]^{-1}(M_i - M_j) + 1/2\ln\frac{|(V_i + V_j)/2|}{\sqrt{|V_i||V_j|}} \tag{4}$$

$$JM_{ij} = \sqrt{2(1 - e^{-B_{ij}})} \tag{5}$$

$$JM = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(p_i \cdot P_j \cdot JM_{ij}) \tag{6}$$

Where V_i and V_j are the covariance matrix for class i and class j respectively, while M_i and M_j are the correspondent mean vectors, B_{ij} is well known as the Bhattacharyya distance, $p_i > p_j$ are the prior probabilities for class i and j, n is the total class number, JM_{ij} is the J-M distance between class i and j, and the superscript T denotes matrix transpose operation. All J-M distances between each two classes form an $n \times n$ class distance matrix, and JM is the weighted J-M distance of this matrix.

Notice that the J-M class distance matrix we have mentioned above is a symmetrical triangle matrix. Thus Eq. (6) could be reformulated as

$$JM = \frac{2}{n(n+1)}\sum_{i=1}^{n-1}\sum_{i=j+1}^{n}(p_i \cdot p_j \cdot JM_{ij}) \tag{7}$$

According to Eq. (7), the GA fitness function can be simply defined as Eq. (8)

$$f(t) = \frac{JM(t)}{\sum_{t=1}^{N} JM(t)} \tag{8}$$

where $JM(t)$ denotes the weighted JM distance value for a specified chromosome t, N denotes the population size, and $f(t)$ denotes the fitness value for the feature combination that chromosome t represents. Therefore, $f(t)$ can be used to evaluate the goodness of individuals for a given generation of population, resulting in those feature combinations with better JM performance selected for reproduction, mating or mutation. This is done based on Roulette Wheel Selection operator.

2.3 The Integration of Wavelet Analysis and Genetic Algorithm

In this study, the integration of wavelet analysis and GA consists of two steps.

First, a 1-D discrete wavelet transform or wavelet packet transform (DPT) is applied on the hyperspectral imagery in the spectral domain at each pixel location. Both of these two transforms decompose the hyperspectral signature of each pixel 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.

Second, given the number of desired optimal features to be output, the GA-based feature selection techniques are applied on the DWT or DPT transformed hyperspectral signal (composite bands), according to the framework we proposed before. The evolutionary process stops only when the population has converged at a certain generation or the maximal number of iterations is reached. For different number of output features, the process is repeated in the same way.

2.4 Comparative Feature Extraction Techniques

For comparison purpose, other feature extraction techniques are also investigated. These include the principal component analysis (PCA), Discriminant Analysis Feature Extraction (DAFE), and Decision Boundary Feature Extraction (DBFE). Conventionally, these feature extraction methodologies are formed using the first several large-amplitude components, e.g. the PCA coefficients, since most of the energy (or information) of the original signal concentrates in these coefficients. The wavelet-based feature extraction is quite a difference from above-mentioned traditional hyperspectral dimensionality reduction methods since it allows the user to extract features related to scale from the hyperspectral signal [7]. To assess whether this characteristic would help increase classification accuracies, the performance of the proposed method is evaluated and compared with PCA, DAFE, and DBFE.

3 EXPERIMENTAL RESULTS

3.1 Test Data and Classification Method

The AVIRIS hyperspectral data used in this paper are from LARS laboratory, Purdue University, accompanied with the ground truth land cover map[21]. This data is from the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) built by JPL and flown by NASA/Ames on June 12, 1992. It contains 145 lines by 145 pixels (21025 pixels) with 220 spectral bands. The 220 calibrated spectral bands

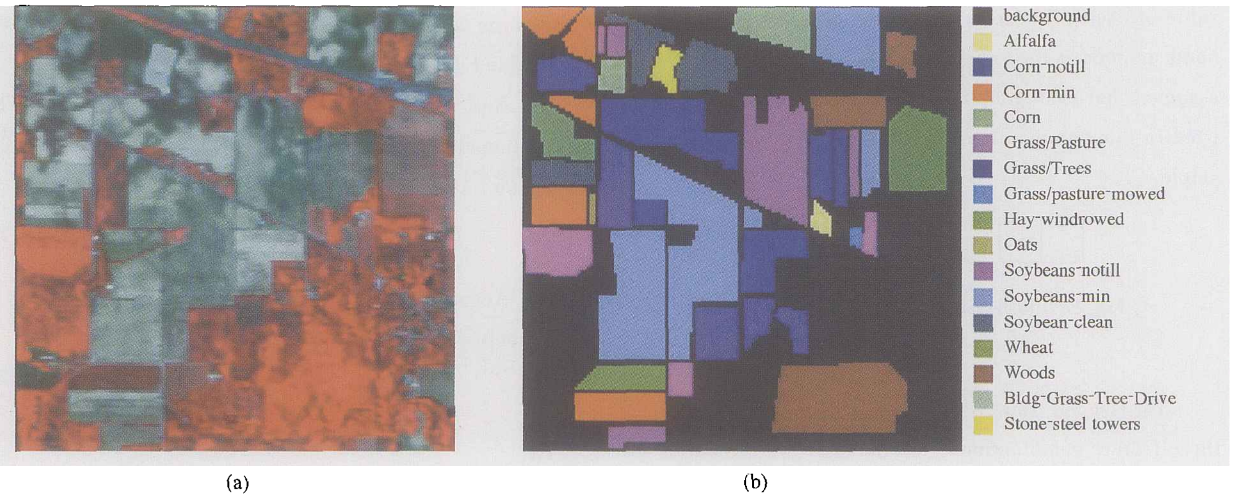

Fig.3 The 220-band AVIRIS test data set (92AV3C) used for feature extraction and land cover classification

Table 1 Number of training and test pixels in AVIRIS data set (92AV3C)

Class No.	Class Name	Training data No. of Pixels	Testing data No. of Pixels
1	Alfalfa	23	46
2	Corn-notill	714	1428
3	Corn-min	415	830
4	Corn	119	237
5	Grass/Pasture	242	483
6	Grass/Trees	365	730
7	Grass/pasture-mowed	28	28
8	Hay-windrowed	239	478
9	Oats	20	20
10	Soybeans-notill	486	972
11	Soybeans-min	1228	2455
12	Soybean-clean	297	593
13	Wheat	103	205
14	Woods	633	1265
15	Bldg-Grass-Tree-Drives	193	386
16	Stone-steel towers	47	93
	Total	5152	10249

extend from 400 nm to 2400 nm in approximately 10 nm bandwidths. The data set is designated 92AV3C9, and calibrated so that the data values in the scene are proportional to radiance. The scene is over an area 6 miles west of West Lafayette, and is a subset of a significantly larger image file, showing as a mixed agriculture/forestry landscape in the Indian Pine Test Site in Northwestern Indiana, as shown in Fig. 3(a). The area is about 2/3 agriculture and 1/3 forest or other natural perennial vegetation[3]. However, due to the early season date of data collection, the cultivated land appears to have very little canopy cover. There is a major dual lane highway and a rail line crossing near the top and a major secondary road near the middle, both in a NW-SE direction. Besides, several county roads are also somewhat apparent. There are 16 classes of land cover types spread over this area, as seen in Table 1.

The classifier adopted here for feature classification is the commonly used maximum likelihood classifier (MLC). A random training of about 50% of the pixels was chosen from the ground truth from each class. The trained classifiers were applied to all of the ground pixels for those classes coverage. The number of training pixels and testing pixels for each class are given in Table 1.

3.2 Evaluate the Combination of Wavelet Transform and GA for Feature Extraction

For the proposed method in this study, the 1-D discrete wavelet transform described in Section 2 are applied on the AVIRIS imagery (92AV3C9 dataset) in the spectral domain at each pixel location, generating a set of composite bands consisting of different approximation and detail information at various scales. According to Bruce *et al.*[9, 10], lower order mother wavelets tend to outperform the higher order mother wavelets and also tend to be more stable than the latter. Therefore, the 2nd order Daubechies function was chosen as wavelet basis.

Meanwhile, the SGA was initialized with one population involving 500 chromosomes. The maximum iteration number was set to 500; crossover probability was set to 0.9; and mutation probability was set to 0.1.

The optimum band/feature combination at different band number by using DWT and GA is shown in Table 2.

From Table 2 we can see that, most of the best classification bands are from the approximate coefficients of wavelet decomposition. As more features are selected, more approximate coefficients are obtained from the first and second level of decomposition. This result shows that a two-level wavelet decomposition of AVIRIS data is sufficient for dimensionality reduction considering the classification performance, as described by Kaewpijit *et al.*[11].

With the increase of feature number, more detail coefficients are involved. However, there are still only a few bands in all selected features, for example, 2 of the 15 channels, 3 of the 16 channels. Most detail coefficients are predominantly from the second level or third level of decomposition. We consider this may be a result of the higher frequent jumble of detail information with high frequency noises hidden in the transformed features during the first level of decomposition, while at coarse scales the noises are more generalized and smoothed. This may also imply that, although wide band remote sensing imageries are disadvantageous for ground object recognition and classification, too fine partition of spectrum into hundreds of narrow bands are also not so much helpful as people may think intuitively, considering the limitations of current pattern recognition models. Thus, the best way is to choose most fit spectral resolution , which will reduce the

Table 2 List of best band combinations by the proposed feature extraction technique base on DWT and GA

No. of Features	Feature No.																		
1	91																		
2	158	188																	
3	131	146	185																
4	14	33	146	158															
5	11	123	179	187	193														
6	9	19	123	131	158	188													
7	9	19	123	130	173	187	195												
8	10	16	19	32	69	123	179	193											
9	11	19	100	128	131	145	155	171	175										
10	17	19	30	118	129	148	172	175	182	193									
11	8	17	19	23	155	179	180	185	188	193	371								
12	91	117	119	121	123	145	171	178	179	181	191	371							
13	73	85	120	121	130	132	158	164	172	173	175	194	377						
14	18	20	40	64	104	116	120	130	158	173	176	178	188	191					
15	13	19	93	115	120	123	130	135	145	174	178	188	191	289	316				
16	12	35	50	69	85	91	117	123	127	135	138	161	174	214	346	371			
17	10	42	45	120	126	142	171	174	175	186	188	191	193	214	292	336	377		
18	19	26	32	35	117	123	134	137	144	147	149	162	172	173	193	194	196	371	
19	30	35	63	85	91	120	121	122	135	138	141	146	163	172	175	188	221	371	373

Note: channel 1—111 are the first level approximate coefficients of DB2 wavelet decomposition; channel 112—168 are the second level approximate coefficients; channel 169—198 are the third level approximate coefficients; channel 199—309 are the first level detail coefficients; channel 310—366 are the second level detail coefficients; channel 367—396 are the third level detail coefficients.

data dimensions without a significant sacrifice of classification accuracy. For this reason, the wavelet transform and wavelet packet transform provide very suitable mathematical tools for spectral resolution judgments.

Fig. 4 shows the changing trends of J-M distance in accordance with different feature numbers by using the proposed feature extraction technique. We can see that the weighted J-M distance increases with the increased number of features monotonically before saturation.

The classification performances of various feature extraction techniques applied on AVIRIS data at different number of features are shown in Fig. 5. 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 conventional feature extraction techniques, such as PCA, DAFE and DBFE. In almost all case when the channel number is greater than three, the proposed method outperforms the other three algorithms, especially when the feature number

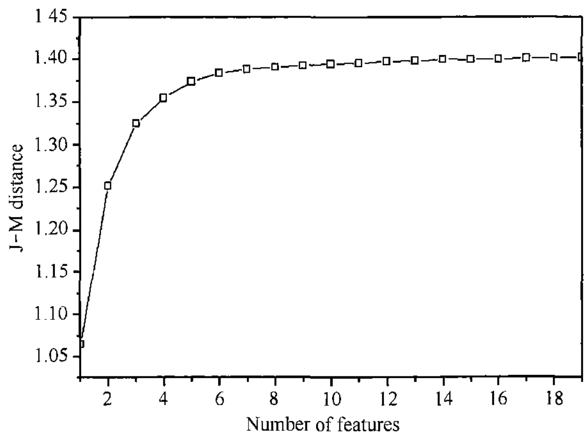

Fig. 4 Changing trends of the J-M distance according to different number of features

is between six and ten. Meanwhile, we found that the accuracy improvement according to increased feature number became saturated when the feature number was greater than fourteen, whatever feature extraction technique was chosen. This could generally be explained as the Hughes phenomena.

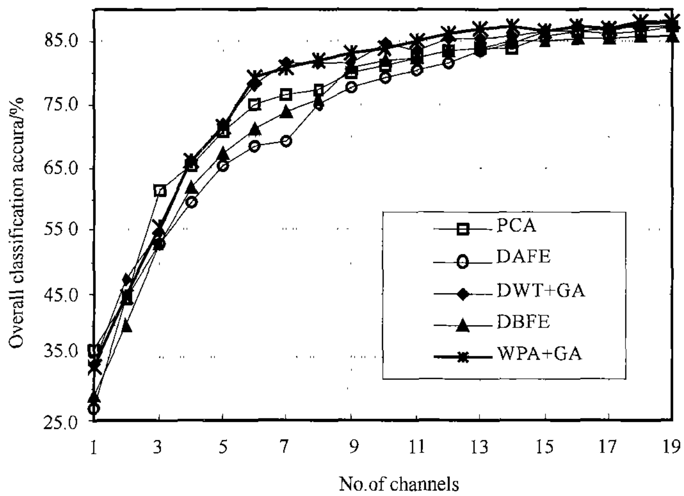

Fig. 5 Comparison of classification results of various feature extraction algorithms applied on AVIRIS data

3. 3 Evaluate the Combination of Wavelet Packet Transform and GA for Feature Extraction

We also investigate the performance of combination of GA with Wavelet Packet Analysis (WPA). For simplicity, the second order Daubechies wavelet packet basis and the Shannon entropy criterion are adopted. The three-level wavelet packet decomposition creates 690 features, followed by a SGA based feature extraction similar to the technique applied on DWT decomposition.

For the SGA, one population with 800 chromosomes was initialized. The maximum iteration number was set to 800; crossover probability was set to 0. 9; mutation probability was set to 0. 1. The best feature combinations at different feature number are classified using the maximum likelihood classifier.

Compare the result from WPA and GA with that from DWT and GA, it can be seen that the accuracy improvement is not prominent when the feature number is small than eight. In this case, the feature number is 2, 3, 4 and 6, respectively; in other case, for example the feature number is 1, 5, and 7, the accuracy is even decreased to a certain extent.

By analysis the results above we believe that there are two possible factors which make the wavelet packet based feature extraction technique does not outperform the DWT based technique.

1) The decomposition of the original signal into 2^n features other than $2n-1$ features by WPA makes the information more fragmentally distributed in all these features. When only a few number of features are selected for classification, those several features with most information are chosen by the classifier, other features with less information are considered as noises and discarded without any help except making the chosen procedure more difficult and wasting more computational resources.

2) The stability of the GA algorithm for feature selection. In fact, since GA is a stochastic search process in the solution space, they may fail in finding the optimal solution when the population size is too small or GA iteration finished with a premature convergence.

When the feature number increases, more fine and useful features decomposed by WPA are automatically evaluated by the GA algorithm, thus provide the possibility of finding more suitable information combined from different features better than the DWT based algorithm, resulting in a certain degree of accuracy improvement for the same classification system.

In general, the 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 (in case of this experiment, 11), the overall classification accuracy becomes more evident, high from 0 to 1. 4%, as shown in Fig. 5.

4 CONCLUSIONS

In this paper, we have presented a new integrated feature extraction technique based on wavelet analysis and genetic algorithm. The proposed algorithm combines the advantages of GA's global optimization and wavelet's multi-resolution and multi-scale analysis. Hyperspectral signals are firstly transformed to feature domain by using a wavelet or wavelet packet decomposition strategy. Followed by an optimal feature subset selection, in which the optimal feature subset acquired the best divergence is chosen according to interclass/intraclass distance of the training samples. This procedure is implemented by a customized Genetic Algorithm. Analytical assessment and experimental results of classification accuracy have proven that the effectiveness of the algorithm for high dimensional data reduction. Specifically, results from 16-class land cover recognition show that the use of DWT and GA-based feature extraction technique improves the overall classification accuracy by 1.1%—6.5%, compared with the use of conventional feature extraction techniques, such as principal component analysis (PCA), Discriminant Analysis Feature Extraction (DAFE) and Decision Boundary Feature Extraction (DBFE).

Additionally, we 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 band number are chosen, the overall classification accuracy becomes more evident, high from 0 to 1.4%.

Meanwhile, we also found that large scale approximate coefficients were selected in most cases. This also proves Bruce *et al.*'s conclusion that larger scale views of the spectral curves (lower spectral resolution data) often were more useful than simply observing the reflectance at the most finely resolved spectral bands[8].

Finally, the proposed feature extraction technique also provides a prototype to implement an automatic classification system for high dimensional remote sensing data, which may be another important issue.

REFERENCES

[1] Pearlman J, Carman S, Segal C, *et al.* Overview of the Hyperion Imaging Spectrometer for the NASA EO-1 Mission [A], IEEE 2001 International Geoscience and Remote Sensing Symposium [C]. July 9—13, 2001, Sydney, Australia.

[2] Barry P, Segal C, Pearlman J, *et al.* Carman Hyperion Data Collection: Performance Assessment and Science Application [A], IEEE 2001 International Geoscience and Remote Sensing Symposium[C]. July 9—13, 2001, Sydney, Australia.

[3] Van Niel T, McVicar T, Campbell S, *et al.* The Coleambally Agricultural Component of Hyperion Instrument Validation [A]. IEEE 2001 International Geoscience and Remote Sensing Symposium[C]. July 9—13, 2001, Sydney, Australia.

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

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

[6] Hsu P H, Tseng Y H. Feature Extraction for Hyper Spectral Image [C]. 20th Asian Conference on Remote Sensing, November22—25, 1999, Hong Kong, China.

[7] Bruce L, Li J. Wavelet for computationally Efficient Hyperspectral Derivative Analysis [J]. *IEEE Transaction on Geoscience and Remote Sensing*, 2002, **39**: 1540—1546.

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

[9] Richards J. Remote Sensing Digital Image Analysis: An introduction, 3rd ed. [M]. Springer-Verlag Berlin Heidelberg, 1997.

[10] Schowengerdt R. Remote Sensing Models and Methods for Image Processing [M]. Academic Press, 1997.

[11] Kaewpijit S, Moigne J, El-Ghazawi T. Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Analysis [J]. *IEEE Transaction on Geoscience and Remote Sensing*, 2003, **41**(4): 863—871.

[12] Mallat S. A Wavelet Tour of Signal Processing, Second Edition [M]. Academic Press, 1999.

[13] Mallat S. A Theory for Multi-resolution Signal Decomposition: The Wavelet Representation [J]. *IEEE Trans. Pattern Anal. Machine Intell.*, 1989, **11**: 674—693.

[14] Barnard H, Weber J, Biemond J. Efficient Signal Extension for Subband/Wavelet Demoposition of Arbitrary Length Signals [J]. *Proc. SPIE*, 1993, **2094**: 966—975.

[15] Wickerhauser M. INRIA Lectures on Wavelet Packet Algorithms [R]. Lecture Notes, Department of Mathematics, Yale University, New Haven, 1991.

[16] Coifman R, Meyer Y, Wickerhauser M. Wavelet Analysis and Signal Processing [A]. (Ruskai B *et al.*) In Wavelets and their Applications, Jones and Bartlett Publishers. 1992: 153—178.

[17] The MathWorks, Inc., Wavelet Toolbox User's Guide [M], The MathWorks, Inc., 2001.

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

[19] Mitchell M. An Introduction to Genetic Algorithms [M]. Cambridge, MA: MIT Press, 1996.

[20] Holland J. Adaptation in Natural and Artificial Systems [M]. Ann Arbor: The University of Michigan Press, 1975. (Second edition printed in 1992 by MIT Press, Cambridge, MA)
[21] URL: 220-band AVIRIS data set (92AV3C). Available at: ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C
[22] Landgrebe D. Multispectral Data Analysis: A Signal Theory Perspective [R]. Purdue University, 1998, Available at: http://www.ece.purdue.edu/~biehl/MultiSpec/Signal_Theory.pdf

基于小波与遗传算法的特征提取与特征选择

刘正军[1],王长耀[2],张继贤[1]

(1. 中国测绘科学研究院,北京 100039;
2. 中国科学院遥感应用研究所遥感科学国家重点实验室,北京 100101)

摘 要: 高维遥感数据的分类与识别与传统的多光谱遥感分类技术具有明显的区别。本文提出了一种基于遗传算法和小波/小波包分析相结合的特征提取方法用于高维遥感数据降维与分类。该方法综合了遗传算法的全局优化和小波/小波包分析的多尺度、多分辨率的特点。首先,通过离散的小波变换(DWT)或小波包变换(WP)将高光谱信号变换到特征域进行光谱分解。由于DWT变换是一种线性变换,不同尺度的DWT系数可作为线性光谱特征。然后,对这些线性光谱特征利用遗传算法结合训练样本计算类内/类间距离搜索最优分类子集,其具体染色体编码取可能的特征号,适应度函数基于样本平均Jeffries-Matusita距离计算。所用的分类器采用最大似然分类器。试验结果表明该方法与常规特征提取算法如主成分变换(PCA)、判别分析特征提取(DAFE)、决策边界特征提取(DBFE)相比,能提高分类精度约1.1%—6.5%。

关键词: 特征提取;小波与小波包;遗传算法