Adaptive classification method for multispectral remote sensing data

Hongzhi Zhao and Mita D. Desai

Division of Engineering, University of Texas at San Antonio
6900 North Loop 1604 West
San Antonio, TX 78249
Hongzhi_zhao_utsa@hotmail.com
Mdesai@nsf.gov

ABSTRACT

Based on local features, a new adaptive classification approach for multispectral remote sensing data is presented. Typical classification techniques based on global features tend to degrade because all classes are projected along the same direction, e.g. the principal component direction for Principal Component Analysis (PCA) and minimum momponent direction for Minimum Component Analysis (MCA). The typical methods are under the assumption that class separability is uniform for all directions, which is not always true. The new method overcomes that disadvantage by selecting features, which give the maximum class separability, based on local information of the classes instead of global information. In the new method, a projection matrix for every class is first sought based on making its training examples well separated from the others. Every input vector is then linearly transformed into another space by every projection matrix. In the transformed spaces, it can be classified or labeled to different class by Maximum Likelihood Classification (MLC). In order to reduce computation cost, adaptive dimension reduction is also introduced. Good performance of the new method can be shown from the experimental results on the Kennedy Space Center (KSC) TM images.

Index Terms--Classification, projection pursuit, feature extraction, multispectral images, maximum likelihood classification

1. INTRODUCTION

Remotely sensed images have been used in a broad range of applications. A short list of applications includes environmental mapping, global change research, geological research, wetlands mapping, assessment of trafficability, plant and mineral identification and abundance estimation, crop analysis, and so on. In all of these applications, the common topic is the requirement for classification of each pixel in the scene. To make the classification well, feature extraction becomes very important [1]. In the processing of multispectral remote sensing images, linear feature extraction can be viewed as finding a set of vectors that represent an observation while reducing the dimensionality. So in classification, it is desirable to extract features that are focused on maximizing dissimilarity among classes. Although a reduction in dimensionality is desirable, the error increment due to the reduction in dimensionality must be constrained to be adequately small. Finding the minimum number of feature vectors that represent observations with reduced dimensionality without sacrificing the discriminating power of pattern classes along with finding the specific feature vectors has been one of the most important problems of the field of classification [2].

Principal Component Analysis (PCA), which is also known as Karhunen-Loeve transformation, is commonly used in feature extraction. As a result of PCA, the data coordinates will be rotated along with the direction of the maximum variance of the data matrix so that the significant information of the data can be prioritized in accordance with the magnitude of the eigenvalues of the data covariance matrix [3]. Minimum Component Analysis (MCA) is another method for feature extraction proposed recently [4]. In MCA, its assumption is that the principal component is not always the best choice since it has a high possibility for a class to overlap with other classes in the principal component direction. So feature extraction is based on projection of the input data along the minimum component direction. Two disadvantages arise from the PCA and MCA approaches. The first one is that the pixels in the PCA-transformed data or MCA-transformed data are still mixing of spectral signatures with unknown abundances. The second disadvantage is that there is no guarantee they are optimal in the terms of class discrimination and separability [5], e.g. PCA is only optimal in the meaning of minimum mean square error (or reconstruction criteria). Moreover, an assumption for PAC and MCA is that the separability is uniform for all component directions, which is not always true [6].

Many attempts have been made to find the best features for classification based on criterion functions [7][8]. Canonical Analysis (CA) is the most typical. In it, a vector is found to maximum the ratio of a between-class scatter matrix to a within-class scatter matrix of a projected data. So CA is based on global information [9][10]. In this paper, an adaptive feature extraction method for classification is proposed. It is very similar to CA, but it is based on local information. In the new method, different classes are projected along different vectors, which are significant to the classes individually. Moreover, adaptive dimension reduction is also introduced in order to reduce computation cost.

2. ADAPTIVE CLASSIFICATION METHOD FOR MULTISPECTRAL REMOTE SENSING DATA

A. Previous Work: CA [3][11]

Within-class scatter matrix S_W is defined:

$$S_W = \sum_{i=1}^{M} P(i)\Sigma_i \tag{1}$$

Between-class scatter matrix S_B is defined:

$$S_B = \sum_{i=1}^{M} P(i)(m_i - m)(m_i - m)' \tag{2}$$

In Eqs. (1) and (2), P(i) is the prior probability of i-th class, M is the number of classes in input data, Σ_i is the i-th class sample convariance, m_i is i-th class sample mean, and m is the total sample mean.

Let P=[p1,p2,…,pJ] is a projection matrix, where pi (i=1,2,…,J) is column vector. Here criterion function is defined:

$$Q(P) = \frac{\det(P'S_B P)}{\det(P'S_W P)} \tag{3}$$

The optimal projection matrix is the one that makes training sample well separated into the set of classes M. In other words, the optimal projection matrix can maximize Eq. (3).

To maximize Eq. (3), the generalized eigenvalues and right eigenvectors of the matrix pair (S_B, S_W) are computed that satisfy $S_B A = S_W A \Lambda$ [12]. The J generalized eigenvectors corresponding to J largest generalized eigenvalues of the matrix pair S_B, S_W form the optimal projection matrix. It is optimal in the terms of class discrimination and separability.

B. Fundamental Idea of Random Process and Mathematical Model for Classification

Suppose $X_1, X_2, \ldots, X_n$ are independent normally distributed random variables such that , for k=1, 2,…., n, X_k has mean m_k and variance s_k^2. The theory of random process [13] shows that the linear combination of $X_1, X_2, \ldots, X_n$, e.g. $Y = a_1 X_1 + a_2 X_2 + \ldots. + a_n X_n$, is also normally distributed with mean m_Y and variance s_Y^2, where

$$m_Y = \sum_{k=1}^{n} a_k m_k \tag{4}$$

$$s_Y^2 = \sum_{k=1}^{n} a_k^2 s_k^2 \tag{5}$$

That is the mathematical basis of newly adaptive classification method proposed here.

For classification, suppose there are N classes in the input data, which are independent normally distributed random variables with different means and different variances. That is the assumption for the new approach. According to the theory of random process above, the input data can be treated as two independent normally distributed random variables with different means and different variances: one is corresponding to a particular class and the other is corresponding to the left. After linear transformation or projection, that is also true.

C. Feature Extraction

In Section A, it is showed that CA is based on a global measure of separability. Because of the selection of global feature, local information of different classes may not be taken into account, which may provide more separability between a particular class and the others when compared to the same feature for all the classes. Due to that, classification accuracy may not be satisfied. Here new adaptive classification approach is based on local information. In other words, different classes have their own projection matrices. Its diagram is shown in Figure 1.

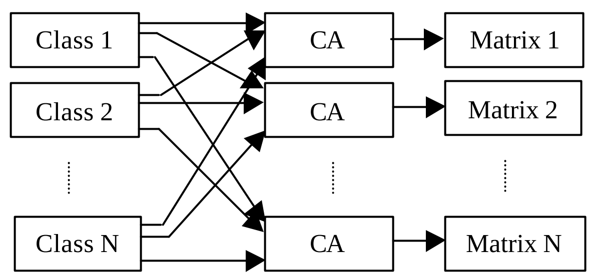

Fig. 1. N projection Matrices Obtained by Training

For ith class, the whole input data can be divided into two classes: ith class and not ith class. According to CA, there existed an optimal projection matrix I, which can make the data belonging to ith class well separated from the other data.

If the number of classes in input data is N, there are N projection matrices corresponding to N classes, one to one. For every vector X in multispectral input data, N features can be extracted by projecting it along N projection matrices.

Moreover, adaptive dimension reduction is also introduced in feature extraction in order to reduce computation cost. Because generalized eigenvectors corresponding to generalized eigenvalues of the matrix pair S_B, S_W ordered in decending order form the optimal projection matrix, the largest J eigenvectors are selected to make the ratio of sum of their eigenvalues to that of all eigenvalues not less that a threshold. So the numbers of eigenvectors in every class projection matrix is adaptively determined.

D. Adaptive Classification

After applying N projection matrices, every input vector X has N features, F_1, F_2, …., F_N, corresponding to N different classes. According to the mathematical model discussed in Section B, every vector in multispectral input data belongs to one of the N normally distributed random variables before and after its linear projection, if it is supposed that every class is corresponding to a normally distributed random variable and the N classes are independent with different means and different variances. So it is easily classified or labeled by MLC, that is

$$X \in w_i \quad \text{if} \quad p(F_i \mid w_i)p(w_i) > p(F_j \mid w_j)p(w_j) \quad \text{for all } j \neq i \qquad (6)$$

and

$$p(F_i \mid w_i) = (2\pi)^{-\frac{N}{2}} \left| \Sigma_i \right|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(F_i - m_i)' \Sigma_i^{-1} (F_i - m_i)\} \qquad (7)$$

where m_i and Σ_i are the mean vector and covariance matrix of the data in class w_i.

A block diagram of the classifier is shown in Figure 2. Let X is an input vector to be classified. X is first projected using projection matrix 1 for the first class. Then projection y_1 is obtained. For the first class, the value $p(F_1 \mid w_1)$ is computed at the point $F_1 = y_1$ by Eq. (7). Then X is projected using projection matrix 2 for the second class, yielding y_2. The value $p(F_2 \mid w_2)$ is computed at the point $F_2 = y_2$ by Eq. (7), too . This scheme is repeated until the value $p(F_N \mid w_N)$ is computed at the point $F_N = y_N$. According to Eq. (6), X is classified or labeled. This procedure is applied to all input vectors in the input data.

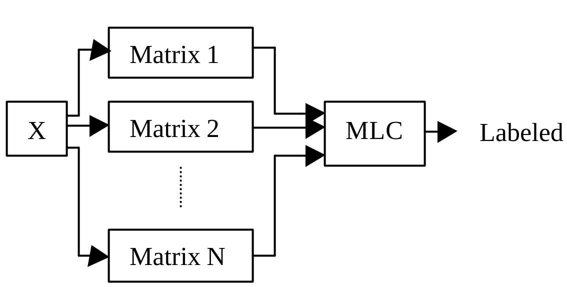

Fig. 2. The Classifier for Input Vector X

3. EXPERIMENTAL RESULTS

The newly presented classification approach is implemented on the Kennedy Space Center (KSC) 1987 Landsat Thematic Mapper (TM) imagery. The study area chosen for this experiment includes a considerable variety of ground cover and there are 20 classes in it. All seven bands of original TM images are used. Band 3 of them is shown in Figure 3. Ground truth image of the same region is also shown in Figure 4. To evaluate the performance of the new method, some traditional methods are also used to same data sets. These traditional methods are MLC, MCA, and CA. Pixel-by-pixel classification accuracy of these kinds of methods are given in Table 1. It is shown that higher classification accuracy for most classes can be obtained by the new adaptive method.

From the Table 1, it is also shown that there is similar tendency between MLC and the new adaptive method. Especially, they produce lower classification accuracy for classes with small prior probability than that with large prior probability. Here in class 1 and class 15, there are only 47 and 10 pixels, respectively. So they have very small prior probability. According to the principle of MLC and the new method, it is easy to understand why MLC and the new methods give so low classification accuracy for these two classes.

4 CONCLUSION

In this paper, an adaptive classification method for multispectral remote sensing data is presented. The proposed approach consists of two stages: feature extraction by using projection matrices for classes in the first stage followed by MLC. The main idea is that the feature extraction is based on local information, which gives the maximum class separability. Moreover, adaptive dimension reduction is also introduced to reduce computation cost. Good performance of the new method can be shown from the experimental results on the Kennedy Space Center (KSC) TM images.

5 ACKNOWLEDGEMENT

This work is financially supported by National Aeronautics and Space Administration (NASA) under Grant NAG10-01SS.

6 REFERENCES

[1] Te-Ming Tu, Chin-Hsing Chen, Jiunn-Lin Wu, and Chein-I Chang, "A fast two-stage classification method for high-dimensional remote sensing data," IEEE Trans. Geosci. Remote Sensing, vol. 36, no. 1, pp. 182-191, 1998.
[2] Chulhee Lee and David A. Landgrebe, "Feature extraction based on decision boundaries," IEEE Trans.

Pattern Anal. Machine Intell., vol. 16, no. 4, pp.388-400, 1993.
[3] John R. Jensen and Xiuping Jia: An Introduction: Remote sensing Digital Image Analysis, 3rd Edition, Springer,New York, 1999.
[4] Guohui He, Mita D. Desai, and Xiaoping Zhang, “Minimum component eigen-vector based classification technique with application to TM images,” ICASSP 1999, pp. 3533-3536, 1999.
[5] Chein-I Chang, Xiao-Li Zhao, Mark L. G. Althouse, and Jeng Jong Pan, “Least squares subspace projection approach to mixed pixel classification for hyperspectral images,” IEEE Trans. Geosci. Remote Sensing, vol. 36, no. 3, pp. 892-912, 1998.
[6] Balaji Mittapalli and Mita D. Desai, “Adaptive local feature based classification of multispectral images,” ICASSP 2000
[7] R. O. Duda and P. E. Hart: Pattern Classification and Scene Analysis. New York: Wiley, 1973.
[8] D. H. Foley and J. W. Sammon, “An optimal set of discriminant vectors,” IEEE Trans. Comput., vol. C-24, no.3, pp.281-289, 1975.
[9] Sadao Fujimura and Senya Kiyasu, “A method for object-oriented feature extraction from hyperspectral data generation of new channel by fusion of data,” IGARSS’97, vol. 2, pp. 975-977, 1997.
[10] Fred E. Shaudys and Todd K. Leen, “Feature selection for improved classification,” IJCNN’92, vol. 4, pp. 697-702, 1992.
[11] C. R. Rao: Linear Statistical Inference and Its Applications, John Wiley and Sons, New York, 1965.
[12] G. H. Golub and C. F. Van Loan: Matrix Computations, The Johns Hopkins Univ. Press, Baltimore, 1983.
[13] Edward R. Dougherty: Random Processes for Image and Signal Processing, IEEE Press, New York, 1999

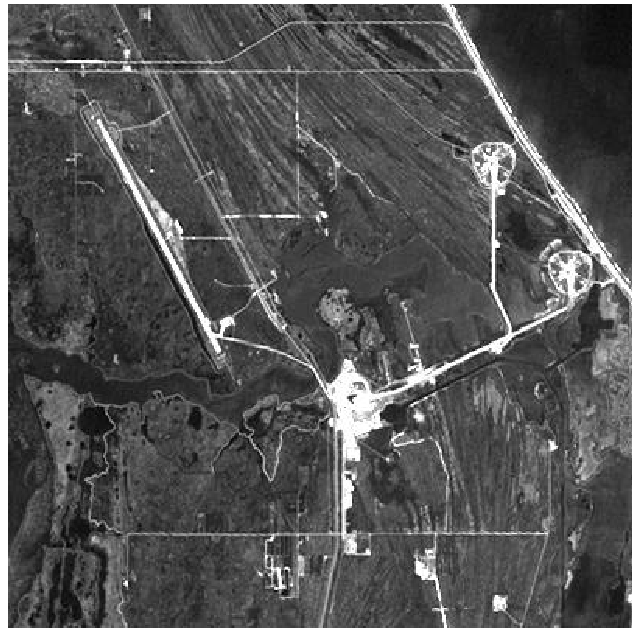

Fig. 3. Testing TM Image, Band 3.

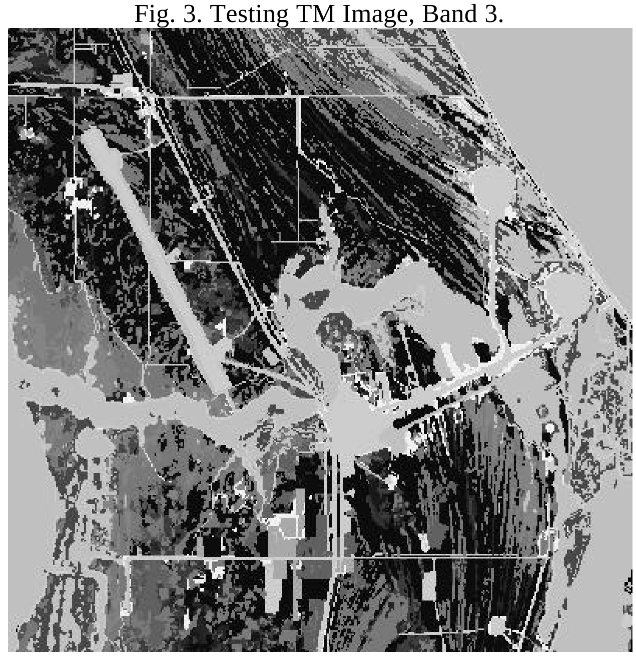

Fig. 4. Ground Truth Image.

Fig. 5. New Technique Classification Result.

Table 1. Classification accuracy with different methods.

Class Number	MLC %	MCA %	CA %	New Method
1	23	76	76	6
2	27	98	93	99
3	87	39	49	89
4	75	87	75	87
5	79	59	78	82
6	79	88	79	89
7	80	75	80	89
8	78	83	77	85
9	80	89	80	89
10	89	65	75	89
11	80	53	80	89
12	72	54	63	75
13	87	94	86	89
14	89	36	89	96
15	60	90	80	60
16	52	92	83	88
17	22	89	50	59
18	91	81	91	95
19	88	49	88	96
20	85	59	85	93