A Hyper spectral Images Classification Method Based on Maximum Scatter Discriminant Analysis

Huo-Yuan LI¹ and Yong-Feng QI²

¹College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou, 730070, China

²College of Computer Science & Engineering, Northwest Normal University, Lanzhou, 730070, China

Abstract. To overcome “small sample size problem” problem faced by some hyper spectral classification methods, the Maximum Scatter Discriminant criterion is used to analyzed hyperspectral data. Maximum Scatter Discriminant analysis searches for the project axes by maximizing the difference of between-class scatter and within-class scatter matrices, which avoid to calculate the inverse of matrices. Experiment results on Indian Pines HSI data set show that the proposed method outperforms the other methods in terms of recognition accuracy. The proposed method is an effective and feasible method for hyper spectral data classification.

1 Introduction

With the constant development of imaging spectroscopy, Hyper spectral Images (HSI) collected by imaging spectrometers provides detailed spectral information about objects in hundreds of spectral bands in the electromagnetic, which increases the possibility of more accurately discriminating materials of interest[1]. However, the information content in a hyper spectral image is highly correlated with neighboring bands and suffers from the curse of dimensionality[2]. Dimensionality reduction is a common pre-processing step in the classification of hyper spectral images for various applications such as land cover mapping, and any spectral feature in the sample space can be represented in a lower-dimensional subspace without losing a significant amount of information. Principal Component Analysis (PCA)[3] and Linear Discriminant Analysis (LDA)[4] is two popular technique for finding such a lower dimensional subspace. PCA performs dimensionality reduction by projecting the original data onto a low dimensional linear subspace which is spanned by the leading eigenvectors of the covariance matrix of all the sample points. PCA projections are optimal for reconstruction from a low dimensional basis, but they may not be optimal from a discrimination standpoint due to not taking the class information of the input data into account. LDA is a supervised learning algorithm, which preserves discriminant information by maximizing the ratio of between-class scatter and within-class scatter. However,

LDA often suffers from the so-called “small sample size problem” since the number of samples is generally smaller than the dimension of the sample space in face recognition task. To overcome the “small sample size problem”, Song et al. [5] proposed a method, which adopts the difference of both between-class scatter and within-class scatter as discriminant criterion, was named as Maximum Scatter Discriminant (MSD) analysis. The goal of MSD is to search for the project axes by maximizing the difference of between-class scatter and within-class scatter matrices rather than the ratio of between-class against within-class scatter matrices.

The rest of the paper is organized as follows: the maximum scatter difference criterion is introduced in Section 2. In Section 3, a series of experiments and results are described. Finally, conclusions are summarized in Section 4.

2 Maximum Scatter Difference (MSD) Criterion

Suppose there are c known pattern classes. P is the total number of training samples, and N_i is the number of training samples in class i . In class i , the j th training sample, which is an n -dimension vector, is denoted by x_i^j . The

within-class scatter matrix S_w^x and between-class scatter matrix S_b^x can be denoted as:

$$S_w^x = \frac{1}{P} \sum_{i=1}^c \sum_{j=1}^{N_i} (x_i^j - m_i)(x_i^j - m_i)^T \quad (1)$$

$$S_b^x = \frac{1}{P} \sum_{i=1}^c N_i (m_i - m_0)(m_i - m_0)^T \quad (2)$$

Where m_i is the mean of i th class samples, and m_0 is the mean of all training samples.

Unlike classical LDA, which search for the project axes on which the ratio of the between-class and the within-class scatter matrices of the projected samples reaches its maximum, the goal of maximum scatter difference criterion is find the project axes on which the difference of the between-class and the within-class scatter matrices is maximized. The criterion can be written as following:

$$J(a) = \max_w \arg A^T (S_b^x - S_w^x) A \quad (3)$$

Obviously, the projection axes are the eigenvectors $a_1, a_2, \dots$ corresponding to the first k largest eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$ of $(S_b^x - S_w^x)$, where $\lambda_1 \geq \lambda_2 \geq \dots$

In fact, in hyper spectral images recognition task, the dimension of hyper spectral data is very high, which will consume much memory and time. To overcome this problem, we first project the hyper spectral data set to a PCA subspace to reduce the dimension, while using PCA as preprocessing can reduce the noise. Thus, a detailed description of the proposed method is stated as follows:

1. PCA projection: the training samples are projected into a PCA subspace which retains a certain level energy by throwing away the smallest principle components. The transformation matrix of PCA is denoted by A_{PCA} .
2. Constructing the within-class scatter matrix S_w^x and between-class scatter matrix S_b^x .
3. MSD projection: The optimal projection matrix A^* is acquired by maximizing (3)
4. OUT transformation matrix: $A = A_{PCA} A^*$

3 Feature Extraction and Classification

Supposed projection matrix A has been obtained, projecting a given spectral vector x onto A , yielding an n -dimensional column vector c :

$$c = A^T x \quad (4)$$

We called this vector c as the feature vector of x . For hyper spectral data recognition, projecting each training vector $x_1, x_2, \dots$ onto A , we obtain the corresponding feature vectors $c_1, c_2, \dots$. Given a test vector x , first use Eq. (4) to get its feature vector c , then a nearest neighbor classifier is used for classification. Here the distance between c and c_i is defined by

$$d(c, c_i) = \|c - c_i\| = \sqrt{\sum_{j=1}^n (c(j) - c_i(j))^2} \quad (5)$$

Where $c(j)$ denotes the j -th element of vector c

4. Experiments

In this section, we conduct a set of experiments on Indian Pines HSI data set^[6] to further evaluate the effectiveness of the proposed method and the proposed method is also compare with PCA and LDA. The Indian Pines HSI data set is a scene of the Northwest Indiana gathered by the AVIRIS sensor in 1992. It consists of 145×145 pixels and 220 spectral bands within the range of 375–2500 nm. Several spectral bands with noise and water absorption phenomena are removed from the data set, leaving a total of 200 radiance channels to be used in the experiments. This HSI is available at <https://engineering.purdue.edu/~biehl/>. Sixteen ground truth classes of interest are considered in the data set. The HSI in false color and its corresponding ground truth are shown in Figure 1(a) and (b) respectively.

(a)

Figure 1. The Indian Pines Hsi Data Set. (a) the Hsi in False Color. (b) Corresponding Ground Truth.

10% pixels are randomly selected from each class to construct the training data set, the remaining images being used as the test images. Every algorithm is repeated 10 times, the recognition accuracies are plotted for the proposed method, PCA and LDA with varying dimension of feature vectors in Fig.2. It can be observed from Fig.2 that the proposed approach has better recognition rate and is more robust. Compared with LDA, the proposed method avoid to calculate the inverse of scatter matrices, which reduces the algorithm complexity

Figure 2. Performance Comparison of Recognition Rates Using LDA, MSD and PCA with Varying Dimension on Indian Pines HSI Data Set.

5. Conclusions

In this paper, an efficient hyper spectral images classification method which adopts maximum scatter discriminant analysis, the major advantage of the proposed is to overcome the “small sample size problem” faced by LDA. Experimental results on Indian Pines HSI data set reveal that the proposed method has better recognition accuracy and is more robust.

Acknowledgements

This article is partly supported by Program for Gansu Province Natural Science Fund (145RJZA183)

References

1. H. Huang, F.Luo, J. Liu and Y. Yang. Dimensionality reduction of hyperspectral images based on sparse discriminant manifold embedding. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 106:42-54
2. L.O. Jimenez, D.A. Landgrebe. Supervised classification in highdimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems Man & Cybernetics Part C*, 1998, 28(1):39-54.
3. N. D. Le Hanh, M. S Kim, and D. H.Kim. Comparison of Singular Value Decomposition and Principal Component Analysis applied to Hyper spectral Imaging of biofilm. *IEEE Photonics Conference*, 2012:6 - 7.
4. T. Bandos, L. Bruzzone and G. Camps-Valls. Classification of Hyper spectral Images With Regularized Linear Discriminant Analysis. *IEEE Transactions on Geoscience & Remote Sensing*, 2009, 47(3):862-873.
5. F. X. Song, K. Cheng, J.Y. Yang, et al., Maximum scatter difference, large margin linear projection and support vector machines, *Acta Automat. Sin.* 2004, 30 (6):890-896. (in Chinese).
6. W. Sun, A. Halevy, J. J. Benedetto, et al. UL_isomap based nonlinear dimensionality reduction for hyperspectral imagery classification. *Isprs Journal of Photogrammetry & Remote Sensing*, 2014, 89(2):25-36.