Spatial and Spectral Nonparametric Linear Feature Extraction Method for Hyperspectral Image Classification

Jinn-Min Yang*, Shih-Hsuan Wei

Department of Mathematics Education, National Taichung University of Education, Taichung, Taiwan.

Received 02 May 2016; received in revised form 21 June 2016; accepted 21 June 2016

Abstract

Feature extraction (FE) or dimensionality reduction (DR) plays quite an important role in the field of pattern recognition. Feature extraction aims to reduce the dimensionality of the high-dimensional dataset to enhance the classification accuracy and foster the classification speed, particularly when the training sample size is small, namely the small sample size (SSS) problem. Remotely sensed hyperspectral images (HSIs) are often with hundreds of measured features (bands) which potentially provides more accurate and detailed information for classification, but it generally needs more samples to estimate parameters to achieve a satisfactory result. The cost of collecting ground-truth of remotely sensed hyperspectral scene can be considerably difficult and expensive. Therefore, FE techniques have been an important part for hyperspectral image classification. Unlike lots of feature extraction methods are based only on the spectral (band) information of the training samples, some feature extraction methods integrating both spatial and spectral information of training samples show more effective results in recent years. Spatial contexture information has been proven to be useful to improve the HSI data representation and to increase classification accuracy. In this paper, we propose a spatial and spectral nonparametric linear feature extraction method for hyperspectral image classification. The spatial and spectral information is extracted for each training sample and used to design the within-class and between-class scatter matrices for constructing the feature extraction model. The experimental results on one benchmark hyperspectral image demonstrate that the proposed method obtains stable and satisfactory results than some existing spectral-based feature extraction.

Keywords: feature extraction, hyperspectral image, dimensionality reduction, classification, small sample size problem

1. Introduction

Remotely sensed hyperspectral images (HSIs) are often with hundreds of measured features (spectral bands) potentially provides more accurate and detailed information for classification and widely used in environmental mapping, geological research, and mineral identification in recent years. The cost of collecting ground-truth of remotely sensed hyperspectral image can be considerably difficult and expensive. Therefore, FE techniques have been an important part for hyperspectral image classification.

In general feature extraction (FE) or dimensionality reduction (DR) plays quite an important role in the field of pattern recognition. Linear discriminant analysis (LDA) [1] is one of the most well-known linear feature extraction methods and has been successfully applied to many fields. The purpose of LDA is to find a linear transformation matrix that can be used to project data from a high-dimensional space into a low-dimensional subspace to mitigate the so-called curse of dimensionality [2], [3] or the Hughes phenomenon [4], [5]. The Hughes phenomenon describes that the ratio of the number of training samples and the number of features must be maintained at or above some minimum value to achieve statistical confidence [5]. Otherwise, the classification accuracy will decline with an increase in the dimensionality of data to some extent. However, it is not necessary to have sufficient training samples to keep the ratio in a high-dimensional classification task. Therefore, by feature extraction, the ratio can be relatively enlarged and the curse of dimensionality can therefore be improved, this will result in an enhancement of classification accuracy. Meanwhile, the computational time can be reduced as well.

* Corresponding author,
Email: jinnminyang@mail.ntcu.edu.tw

Basically, LDA has three inherent deficiencies in dealing with classification problems. First, LDA is only well-suited for normally distributed data [1]. If the distributions are significantly non-normal, the use of LDA cannot be expected to accurately indicate which features should be extracted to preserve complex structures needed for classification. Second, since the rank of between-class scatter matrix is the number of classes minus one [1], the number of features can be extracted at most remains the same. Third, the singularity problem arises when dealing with high-dimensional and small sample size (SSS) data [1, 3, 6-7]. Nonparametric linear discriminant analysis such as nonparametric discriminant analysis (NDA) [1], nonparametric weighted feature extraction (NWFE) [6] and cosine-based feature extraction (CNFE) [7] provide solutions for circumventing the previously mentioned problems.

The aforementioned FE methods are spectral-based algorithms; in other words, they measure similarity in the spectral space. Using only spectral information to classification tasks is insufficient. Spatial contexture information has been proven to be useful to improve the classification of HSI data in recent years [8]-[9]. In the paper, a nonparametric feature extraction method, integrating both spectral and spatial information, is proposed.

The rest of this paper is organized as follows. The proposed method and its experiment are described in Section 2. The experimental results and discussion are provided in Section 3. Finally, Section 4 gives some conclusions of the paper.

2. Method

The goal of FE is to find a transformation matrix A which maximizes the class separability $J = \mathrm{tr}(\mathbf{S}_w^{-1}\mathbf{S}_b)$ in the transformed space, where $\mathbf{S}_w$ and $\mathbf{S}_b$ denote the within-class and between-class scatter matrices, respectively. That is

$$\mathbf{A} = \underset{\mathbf{A}}{\mathrm{argmax}}\mathrm{tr}((\mathbf{A}^T\mathbf{S}_w\mathbf{A})^{-1}\mathbf{A}^T\mathbf{S}_b\mathbf{A}) \tag{1}$$

The maximization of (1) is equivalent to solving the generalized eigenvalue decomposition problem

$$\mathbf{S}_i\mathbf{v}_i = \lambda_i\mathbf{S}_w\mathbf{v}_i, \quad i = 1, \dots, p$$

where p denotes the dimensionality of the transformed space, $(\lambda_i, \mathbf{v}_i)$ represent the eigen-pair of $\mathbf{S}_w^{-1}\mathbf{S}_b$, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Thus, the transformation matrix $\mathrm{A} = [\mathrm{v}_1, \dots, \mathrm{v}_p]$ can be obtained. The proposed spatial and spectral feature extraction method includes two parts, one idea is to incorporate the spatial information into the within-class and scatter matrix design, and the other idea is to incorporate another scatter matrix to regularize the within-class scatter matrix.

2.1. The Nonparametric Linear Discriminate Analysis

The within-class matrix and between-class scatter matrix of the nonparametric linear discriminant feature extraction (NLDA) are defined as follows, respectively.

$$\mathrm{S}_w = \sum_{i=1}^{L} P_i \sum_{\ell=1}^{N_i} (\mathbf{x}_\ell^{(i)} - \mathrm{M}_i(\mathbf{x}_\ell^{(i)}))(\mathbf{x}_\ell^{(i)} - \mathrm{M}_i(x_\ell^{(i)}))^T \tag{2}$$

$$\mathbf{S}_b = \sum_{i=1}^{L} P_i \sum_{\substack{j=1,\\ j\neq i}}^{L} \sum_{\ell=1}^{N_i} (\mathbf{x}_\ell^{(i)} - \mathrm{M}_j(x_\ell^{(i)}))(\mathbf{x}_\ell^{(i)} - \mathrm{M}_j(x_\ell^{(i)}))^T \tag{3}$$

where $\mathbf{M}_i(\mathbf{x}_\ell^{(i)})$ and $\mathbf{M}_j(\mathbf{x}_\ell^{(i)})$ denote the local mean of training sample $\mathbf{x}_\ell^{(i)}$ corresponding to the ith class and jth class, respectively. The local mean of $\mathbf{x}_\ell^{(i)}$ is computed by its k-nearest neighbors (kNNs) in the same class or in the different classes as shown in (4).

$$\mathbf{M}_j(\mathbf{x}_\ell^{(i)}) = \frac{1}{k}\sum_{s=1}^{k}\mathbf{x}_{sNN}^{(i)} \tag{4}$$

2.2. The Spatial and Spectral Nonparametric Linear Discriminate Analysis

The within-class matrix and between-class scatter matrix of the spatial and spectral nonparametric linear discriminant feature extraction (SSNLDA) are defined as follows, respectively.

$$S_w^{SS} = \sum_{i=1}^{L} P_I \sum_{\ell=1}^{N_i}\left(x_\ell^{(i)} - M_i\left(x_\ell^{(i)}\right)\right)\cdot\left(x_\ell^{(i)} - M_i\left(x_\ell^{(i)}\right)\right)^T \tag{5}$$

$$S_b^{SS} = \sum_{i=1}^{L} P_i \sum_{\substack{j=1\\ j\neq i}}^{L}\sum_{\ell=1}^{N_i}\left(x_\ell^{(i)} - M_j\left(x_\ell^{(i)}\right)\right)\left(x_\ell^{(i)} - M_j\left(x_\ell^{(i)}\right)\right)^T \tag{6}$$

where $\mathbf{M}_i(\mathbf{x}_\ell^{(i)})$ and $\mathbf{M}_j(\mathbf{x}_\ell^{(i)})$ denote the local mean of training sample $\mathbf{x}_\ell^{(i)}$ corresponding to the ith class and jth class, respectively. The local mean of $\mathbf{x}_\ell^{(i)}$ is computed by utilizing the spectral weighted local mean $\mathbf{M}_j^t(\mathbf{x}_\ell^{(i)})$ and spatial weighted local mean $\mathbf{M}_j^s(\mathbf{x}_\ell^{(i)})$, as shown in (7).

$$\mathbf{M}_j(\mathbf{x}_\ell^{(i)}) = \gamma \mathbf{M}_j^t(\mathbf{x}_\ell^{(i)}) + (1-\gamma)\mathbf{M}_j^s(\mathbf{x}_\ell^{(i)}) \tag{7}$$

where

$$\mathbf{M}_j^t(\mathbf{x}_\ell^{(i)}) = \frac{1}{k}\sum_{s=1}^{k} w_{is}^t \mathbf{x}_{sNN}^{(j)} \tag{8}$$

with

$$w_{is}^t = \frac{\left\|\mathbf{x}_\ell^{(i)} - \mathbf{x}_{sNN}^{(j)}\right\|}{\sum_{s=1}^{k}\left\|\mathbf{x}_\ell^{(i)} - \mathbf{x}_{sNN}^{(j)}\right\|} \tag{9}$$

and

$$\mathbf{M}_j^s(\mathbf{x}_\ell^{(i)}) = \frac{1}{k}\sum_{s=1}^{k} w_{is}^s \mathbf{x}_{sNN}^{(j)} \tag{10}$$

with

$$w_{is}^s = \frac{\left\|c(\mathbf{x}_\ell^{(i)}) - c(\mathbf{x}_{sNN}^{(j)})\right\|}{\sum_{s=1}^{k}\left\|c(\mathbf{x}_\ell^{(i)}) - c(\mathbf{x}_{sNN}^{(j)})\right\|} \tag{11}$$

and $c(\mathbf{x}_\ell^{(i)})$ denotes the coordinate of training sample $\mathbf{x}_\ell^{(i)}$, the parameter $0 \le \gamma \le 1$.

Another part of is the distance scatter matrix introduce in [8]. Based on a $z \times z$ window, a training sample $\mathbf{x}_\ell^{(i)}$ and its pixel neighbors form a local patch $N(\mathbf{x}_\ell^{(i)})$, where the odd number z is the width of the neighborhood window. The scatter matrix $\mathbf{H}^s = \sum_{\ell=1}^{N} \mathbf{H}_\ell$ and

$$\mathbf{H}_\ell = \sum_{k=1}^{s} \frac{w_k}{\sum_{j=1}^{s} w_j}(\mathbf{x}_\ell^{(i)} - \mathbf{x}_k)(\mathbf{x}_\ell^{(i)} - \mathbf{x}_k)^T \tag{12}$$

where $s = z^2 - 1$, $\mathbf{x}_k \in N(\mathbf{x}_\ell^{(i)})$ and $w_k = exp\{-r_0||\mathbf{x}_\ell^{(i)} - \mathbf{x}_k||\}$. Parameter r_0 reflects the degree of filtering.

Regularization is employed to improve the singularity problem in SSNLDA. The within-class scatter matrix is replaced by

$$\mathbf{S}_w^{SS} = (1-\beta)[\alpha \mathbf{S}_w^{SS} + (1-\alpha)\mathrm{diag}(\mathbf{S}_w^{SS})] + \beta \mathbf{H} \tag{13}$$

where $\mathrm{diag}(\mathbf{A})$ denotes the diagonal parts of a matrix $\mathbf{A}$ and $0 \le \alpha, \beta \le 1$.

2.3. Dataset

The Indian Pines image, mounted from an aircraft flown at 65000-ft altitude and operated by the NASA/Jet Propulsion Laboratory, with the size of 145 × 145 pixels has 220 spectral bands measuring approximately 20 m across the ground. We have also reduced the number of bands to 200 by removing bands covering the region of water absorption: 104-108, 150-163, and 220. There are 16 classes in the data set. The total number of samples is 10249, ranging from 20 to 2455 in each class. In figure 1, the left and right images depict the false color composition of three sample bands 50, 27 and 17 and its ground truth of the Indian Pines dataset, respectively.

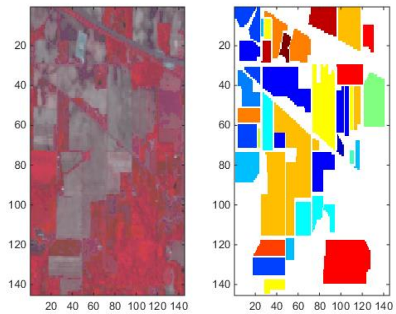


Fig. 1 The left figure depicts Indian Pines image of band 50, 27 and 17; the right one shows its ground truth.

2.4. Experiment Design

Three different cases, each class with 5 (case I), 10 (case II), and 20 (case III) training samples are investigated to discover the effect on the sizes of training samples in the experiments. The remaining samples are employed as the test samples. The cases I and II are the so-called ill-posed and poorly posed classification problems [7], respectively. They are challenging cases in the field of pattern recognition. In each case, the training and testing datasets are randomly selected. We will repeat each case for 10 times and report the averaged overall accuracy (OA) and standard deviation.

Two other linear feature extraction methods, CNFE and NWFE, are utilized to compare the classification performance with the proposed SSNLDA. The 1-nearest neighbor (1NN) classifier is employed. In SSNLDA, we adopt a 5×5 window to form a local patch and the values of α, β and ζ are set as 0.5.

Table 1 Graph representations

class	# pixels	class name
1	46	Alfalfa
2	1428	Corn-notill
3	830	Corn-mintill
4	237	Corn
5	483	Grass-pasture
6	730	Grass-trees
7	28	Grass-pasture-mowed
8	478	Hay-windrowed
9	20	Oats
10	972	Soybean-notill
11	2455	Soybean-mintill
12	593	Soybean-clean
13	205	Wheat
14	1265	Woods
15	386	Buildings-Grass-Trees-Drives
16	93	Stone-Steel-Towers

3. Results and Discussion

Table 2 lists the best classification accuracies of the three cases of the Indian Pines dataset. As we can see from the table, 1NN classifier with SSNLDA features can achieve better results than with CNFE and NWFE features. SSNLDA provides about 6% improvement as compared with the other two methods. Meanwhile, the standard deviations is smaller as well. Fig. 2 demonstrates the variations of OAs with the reduced dimensions where 5, 10 and 20 training samples are utilized. The proposed SSNLDA significantly outperform the other two methods. Fig. 3 shows the classification map of the Indian Pines scene using 1NN classifier for 20 training samples case. The dimensionality of the reduced space is 30. As shown in Fig. 3, the 1NN classifier with SSNLDA feature can get better results.

Table 2 Classification accuracies (in percent) in Indian Pines scene

Case	FE	OA±std (#features)
$N_i = 5$	NWFE	63.03±4.68(28)
	CNFE	65.10±4.91(15)
	SSNLDA	71.05±3.28(28)
$N_i = 10$	NWFE	72.82±2.28(29)
	CNFE	75.67±1.77(29)
	SSNLDA	81.79±1.81(30)
$N_i = 20$	NWFE	80.15±1.76(28)
	CNFE	81.93±2.17(28)
	SSNLDA	88.08±1.21(30)

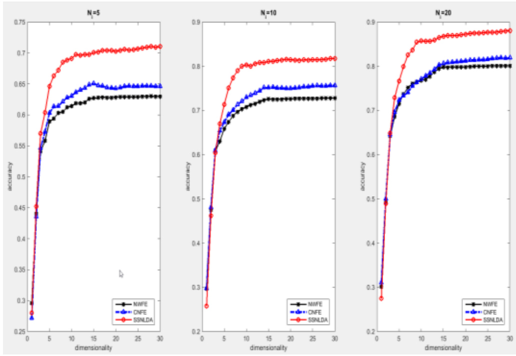
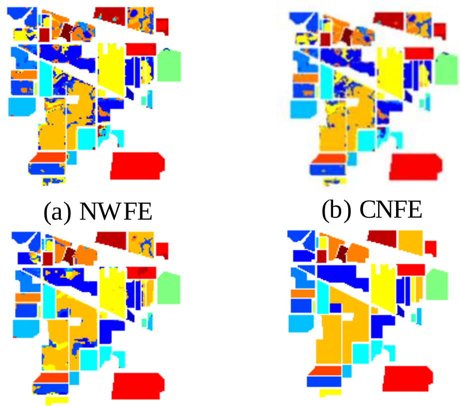


Fig. 2 Classification results on the Indian Pines dataset for the three feature extraction method.



(a) NWFE (b) CNFE

(c) SSNLDA (d) Ground-truth

Fig. 3 Classification maps of the Indian Pines scene using 1NN classifier for 20 training samples case.

4. Conclusions

In this paper, a spectral and spatial information-based nonparametric feature extraction SSNLDA is proposed. From the above results, we find SSNLDA can achieve more stable and effective results. In most of cases, 1NN classifier with SSNLDA features can obtain better results than other spectral-based FE, CNFE and NLDA, particularly when the training sample size is quite small.

Acknowledgement

This study is supported by Ministry of Science and Technology, R.O.C., under the contract number of MOST 104-2221-E-142- 005.

References

[1] K. Fukunaga, Introduction to statistical pattern recognition, 2nd ed., New York: Academic Press, 1990.

[2] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern classification, 2nd ed., New York: John Wiley & Sons, 2001.
[3] S. J. Raudys and A. K. Jain, "Small sample size effects in statistical pattern recognition: recommendations for practitioners," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 13 no. 3, pp. 252-264, 1991.
[4] D. A. Landgrebe, Signal theory methods in Multispectral Remote Sensing, New Jersey: John Wiley and Sons, 2003.
[5] P. K. Varshney and M. K Arora, Advanced image processing techniques for remotely sensed hyperspectral data, New York: Springer, 2004.
[6] B. C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," IEEE Transaction on Geoscience and Remote Sensing, vol. 42, no. 5, pp. 1096-1105, 2004.
[7] J. M. Yang, P. T. Yu, and B. C. Kuo, "A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data," IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 3, pp. 1279-1293, 2010.
[8] Y. Zhou, J. Peng, and C. L. Philip Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 2, pp. 1082-1095, 2015.
[9] H. Pu, Z. Chen, B. Wang, and G. M. Jiang, "A novel spatial–spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery," IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 11, pp. 7008-7022, 2014.
[10] Hyperspectral remote sensing scenes, [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes. [Accessed 24 May 2016]

