International Journal of Applied Earth Observation and Geoinformation 31 (2014) 122–129

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Sparse dimensionality reduction of hyperspectral image based on semi-supervised local Fisher discriminant analysis

Zhenfeng Shao, Lei Zhang*

The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, PR China

A R T I C L E I N F O

Article history:
Received 22 January 2014
Accepted 20 March 2014
Available online 13 April 2014

Keywords:
Sparsity preserving projections (SPP)
Dimensionality reduction
Semi-supervised local Fisher discriminant analysis (SELF)

A B S T R A C T

This paper presents a novel sparse dimensionality reduction method of hyperspectral image based on semi-supervised local Fisher discriminant analysis (SELF). The proposed method is designed to be especially effective for dealing with the out-of-sample extrapolation to realize advantageous complementarities between SELF and sparsity preserving projections (SPP). Compared to SELF and SPP, the method proposed herein offers highly discriminative ability and produces an explicit nonlinear feature mapping for the out-of-sample extrapolation. This is due to the fact that the proposed method can get an explicit feature mapping for dimensionality reduction and improve the classification performance of classifiers by performing dimensionality reduction. Experimental analysis on the sparsity and efficacy of low dimensional outputs shows that, sparse dimensionality reduction based on SELF can yield good classification results and interpretability in the field of hyperspectral remote sensing.

1. Introduction

Since hyperspectral images can provide very high discrimination capabilities due to hundreds of spectral bands, hyperspectral images have been broadly applied to the areas of image classification (Demir and Erturk, 2007; Fauvel et al., 2008; Ma et al., 2010; Chen et al., 2008). However, high dimensional characteristic of hyperspectral images will cause increasing computation costs and memory demands, which is inefficient and undesirable for classification task. The problem caused by high number of bands and low number of labeled training samples, referred to as the Hughes phenomenon (Hughes, 1968; Serpico and Moser, 2007; Hsu, 2007) or the curse of dimensionality (Bellman, 1961; Chen, 2009), has been associated with remote sensing applications since hyperspectral data became available in the late 1980s (Landgrebe, 2003). Since the original data with redundant and noisy bands cannot be the most effective dimensions for representing the hyperspectral data, dimensionality reduction is a critical task for hyperspectral data classification.

In recent years, manifold learning as a perfect tool for data mining has been the main focus of dimensionality reduction. The representative methods include Isometric Feature Mapping (Isomap) (Tenenbaum et al., 2000), Laplacian eigenmaps (Belkin and Niyogi, 2003), Local tangent space alignment (LTSA) (Zhang and Zha, 2005) and Locally Linear Embedding (LLE) (Roweis and Saul, 2000), etc. However, these traditional manifold learning algorithms generally suffer from the “out-of-sample” (Trosset and Priebe, 2008) problem, that is, these algorithms do not provide a feature mapping (explicit or implicit) to map new data points that are not included in the training set into the learned manifold. In order to address this issue by constructing explicit feature mappings, sparsity preserving projections (SPP) (Qiao et al., 2010) aims at preserving the sparse reconstructive relations among samples in a low-dimensional space. Although it belongs to global methods, it owns local properties due to the sparse representation procedure. Since SPP is linear and defined everywhere, thus the “out-of-sample” problem is naturally solved. Besides it does not have to encounter model parameters such as the neighborhood size and can be easily extended to semi-supervised scenarios based on the existing dimensionality reduction framework.

However, SPP mainly focuses on the intra-class geometrical information and ignores the discriminative information hidden in the hyperspectral images. Local Fisher discriminant analysis (LFDA) (Sugiyama, 2007) localizes the evaluation of the within-class scatter, working well even when within-class multimodality or outliers exist, and was shown to compare favorably with other supervised dimensionality reduction methods through experiments. But the performance of LFDA as a supervised dimensionality reduction method tends to be degraded when only a small number of labeled samples are available. Now semi-supervised dimensionality reduction as a new issue in semi-supervised learning from

* Corresponding author at: No. 129, Luoyu Road, Wuhan, PR China. Tel.: +86 15271828522.
E-mail address: xysyleilei@126.com (L. Zhang).

http://dx.doi.org/10.1016/j.jag.2014.03.015

both unlabeled samples and labeled samples is appealing. Semi-supervised local Fisher discriminant analysis (SELF) (Sugiyama et al., 2010) is a recently proposed semi-supervised manifold learning method to exploit both geometrical and discriminant information simultaneously for dimensionality reduction, which preserves the global structure of unlabeled samples in addition to separating labeled samples in different classes from each other. It has demonstrated better performance than some other manifold learning methods such as locality preserving projection (LPP) (He et al., 2005) and LFDA for classification.

Since SELF produces a low dimensional subspace and each basis of the subspace is a linear combination of all the original features used for high dimensional sample representation, the classification result cannot be interpreted precisely. The sparse representation reduces the space cost and offers explicit interpretations to new coordinates. Moreover, sparse representations of samples in the same class tend to share the same support set, and thus the subsequent classification can be improved. Therefore, in this paper, we propose a sparse dimensionality reduction method based on semi-supervised local Fisher discriminant analysis to realize advantageous complementarities between SELF and SPP.

The remainder of the paper is organized as follows. Section 2 first presents a brief review of the related dimensionality reduction methods and then introduces the rationale and formulation of our method. Experimental results and discussions are demonstrated in Section 3. Finally, Section 4 draws the conclusion of the work.

2. Methodology

In this section, we provide a brief review of Fisher discriminant analysis (FDA) (Belhumeur et al., 1997), local Fisher discriminant analysis, and Semi-supervised local Fisher discriminant analysis, which are relevant to subsequent method we propose.

2.1. Fisher discriminant analysis

Fisher discriminant analysis (FDA) is a popular supervised dimensionality reduction technique. Let $x_i \in \mathbb{R}^d (i = 1, 2, \ldots, n)$ be d-dimensional samples and $y_i \in \{1, 2, \ldots, c\}$ be corresponding class labels, where n is the number of samples and c is the number of classes. Let n_k be the number of samples in class $k \in \{1, 2, \ldots, c\}$:

$$\sum_{k=1}^{c} n_k = n \tag{1}$$

Let X be the matrix of all samples:

$$X \equiv (\, x_1 | x_2 | \cdots | x_n \,) \tag{2}$$

Let $z_i \in \mathbb{R}^r (1 \le r \le d)$ be low-dimensional representations of x_i, where r is the dimension of the embedding space. Let $S^{(b)}$ and $S^{(w)}$ be the between-class scatter matrix and the within-class scatter matrix:

$$S^{(b)} := \sum_{k=1}^{c} n_k (\mu_k - \mu)(\mu_k - \mu)^T \tag{3}$$

$$S^{(w)} := \sum_{k=1}^{c} \sum_{i:y_i=k} (x_i - \mu_k)(x_i - \mu_k)^T \tag{4}$$

where $\sum_{i:y_i=k}$ indicates the summation over i such that $y_i = k$ and μ is the mean of all samples, μ_k is the mean of samples in class k:

$$\mu_k = \frac{1}{n_k} \sum_{i:y_i=k} x_i \tag{5}$$

The FDA transformation matrix $T^{(\mathrm{FDA})}$ is defined as

$$T^{(\mathrm{FDA})} = \underset{T \in \mathbb{R}^{d \times r}}{\arg\max} [tr(T^T S^{(b)} T (T^T S^{(w)} T)^{-1})] \tag{6}$$

The total scatter matrix $S^{(t)}$ is defined as

$$S^{(t)} = S^{(b)} + S^{(w)} \tag{7}$$

2.2. Local Fisher discriminant analysis

Local Fisher discriminant analysis is a supervised dimensionality reduction method, which overcomes the weakness of the original FDA against within-class multimodality or outliers.

Let $S^{(lb)}$ and $S^{(lw)}$ be the local between-class scatter matrix and the local within-class scatter matrix:

$$S^{(lb)} := \frac{1}{2} \sum_{i,j=1}^{n} W_{i,j}^{lb} (x_i - x_j)(x_i - x_j)^T \tag{8}$$

$$S^{(lw)} := \frac{1}{2} \sum_{i,j=1}^{n} W_{i,j}^{lw} (x_i - x_j)(x_i - x_j)^T \tag{9}$$

where T denotes the transpose of a matrix or vector, and W^{lb} and W^{lw} are the $n \times n$ matrices with

$$W_{i,j}^{lb} := \begin{cases} A_{i,j}(1/n - 1/n_{y_i}) & \text{if } y_i = y_j \\ 1/n & \text{if } y_i \ne y_j \end{cases} \tag{10}$$

$$W_{i,j}^{lw} := \begin{cases} A_{i,j}/n_{y_i} & \text{if } y_i = y_j \\ 0 & \text{if } y_i \ne y_j \end{cases} \tag{11}$$

where $A_{i,j}$ is the affinity value between x_i and x_j based on the local scaling heuristic (Zelnik-Manor and Perona, 2004):

$$A_{ij} = \begin{cases} \exp\left(-\dfrac{d^2(x_i, x_j)}{\sigma^i \sigma^j}\right), & \text{if } x_i \in N(x_j) \text{ or } x_j \in N(x_i), \\ 0, & \text{otherwise}, \end{cases} \tag{12}$$

where $d^2(x_i, x_j)$ is the distance from x_i to x_j, σ is the local scaling parameter and $N(x_i)$ is the set of k nearest neighbors of x_i.

The LFDA transformation matrix $T^{(\mathrm{LFDA})}$ is defined as

$$T^{(\mathrm{LFDA})} := \underset{T \in \mathbb{R}^{d \times r}}{\arg\max} [\,] \left[tr(T^T S^{(lb)} T (T^T S^{(lw)} T)^{-1}) \right] \tag{13}$$

Eq. (13) can be maximized directly by calculating the generalized eigenvectors of the following generalized eigen equation

$$S^{(lb)} T = \lambda S^{(lw)} T \tag{14}$$

LFDA seeks a transformation matrix such that the local between-class scatter in the embedding space is maximized and the local within-class scatter in the embedding space is minimized.

2.3. Semi-supervised local Fisher discriminant analysis

SELF is a new dimensionality reduction method for semi-supervised learning scenarios. SELF smoothly bridges the LFDA and PCA so that the global structures of all points as well as local structures defined by a small number of labeled data can be controlled.

SELF tries to find a projection which maximizes the regularized local between-class scatter and minimizes the regularized local within-class scatter in the embedding space. So, a reasonable criterion for choosing a good map is to optimize the following two objective functions

$$\begin{aligned} &\max \; TS^{(rlb)} T^T \\ &\min \; TS^{(rlw)} T^T \end{aligned} \tag{15}$$

where $S^{(rlb)}$ and $S^{(rlw)}$ are the regularized local between-class scatter matrix and the regularized local within-class scatter matrix respectively:

$$S^{(rlb)} = (1-\beta)S^{(lb)} + \beta S^{t}$$
$$S^{(rlw)} = (1-\beta)S^{(lw)} + \beta I \tag{16}$$

where I is the identity matrix, S^t is the total scatter matrix, and $\beta \in [0, 1]$ is a tradeoff parameter.

2.4. Sparse dimensionality reduction based on SELF

SPP firstly seeks a sparse coefficient vector s_i for each x_i which satisfies $x_i = Xs_i$ through the following modified L_1-norm optimization problem:

$$\min \|s_i\|_1 \quad s.t. \quad \|y - Xs_i\| < \varepsilon \quad 1 = e^T s_i \tag{17}$$

where ε is used to control the reconstructive error and $s_i = [s_{i,1}, \ldots, s_{i,i-1}, 0, s_{i,i+1}, \ldots, s_{i,n}]^T$ is a n-dimensional vector in which the ith element is equal to zero. That is to say, the x_i is removed from X. And the elements $s_{i,j}(i \neq j)$ denote the contribution of each x_j to reconstruct x_i. e is a vector of all ones. Then the optimal weight vector $\tilde{s}_i$ obtained from Eq. (17) is used to the following objective function:

$$\min \sum_{i=1}^{n} \|w^T x_i - w^T X s_i\|^2 = \min w^T X(I - S - S^T + SS^T)X^T w \quad s.t. \quad w^T X X^T w = 1 \tag{18}$$

where $S = [\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n]$ and $w^T X X^T w = 1$ can be regarded as a constraint to guarantee stable solution. Eq. (18) can be written in a matrix form as:

$$\min \frac{w^T X(I - S - S^T + SS^T)X^T w}{w^T X X^T w} \tag{19}$$

And, the optimal of SPP are the eigenvectors of the following generalized eigenvalue problem:

$$w^T X(I - S - S^T + SS^T)X^T w = \lambda w^T X X^T w \tag{20}$$

The minimization problem of SPP can further be transformed to an equivalent maximization issue:

$$\max \frac{w^T X(S + S^T - SS^T)X^T w}{w^T X X^T w} \tag{21}$$

We propose the following formula to solve the following generalized eigenvalue issue:

$$S^{(\mu)}\alpha = \lambda S^{(\mu)}\alpha \tag{22}$$

where $S^{(\mu)} = S + S^T - SS^T$. In order to combine the SPP and SELF, we establish the new objective function as follows:

$$F = \max\ TS^{(rlb)}T^T S^{(\mu)} \tag{23}$$

Subject to the constraints

$$TS^{(rlw)}T^T = 1 \quad \text{and} \quad w^T X X^T w = 1 \tag{24}$$

Thus, the maximum of the objective function F is calculated using the Lagrange multiplier method, the Lagrange function is obtained as follows:

$$L_{(F)} = TS^{(rlb)}T^T S^{(\mu)} - \lambda_1(TS^{(rlw)}T^T - 1) - \lambda_2(w^T X X^T w - 1) \tag{25}$$

Then the following generalized eigenvalue issue can be solved:

$$S^{(mrlb)}\alpha = \lambda S^{(mrlw)}\alpha \tag{26}$$

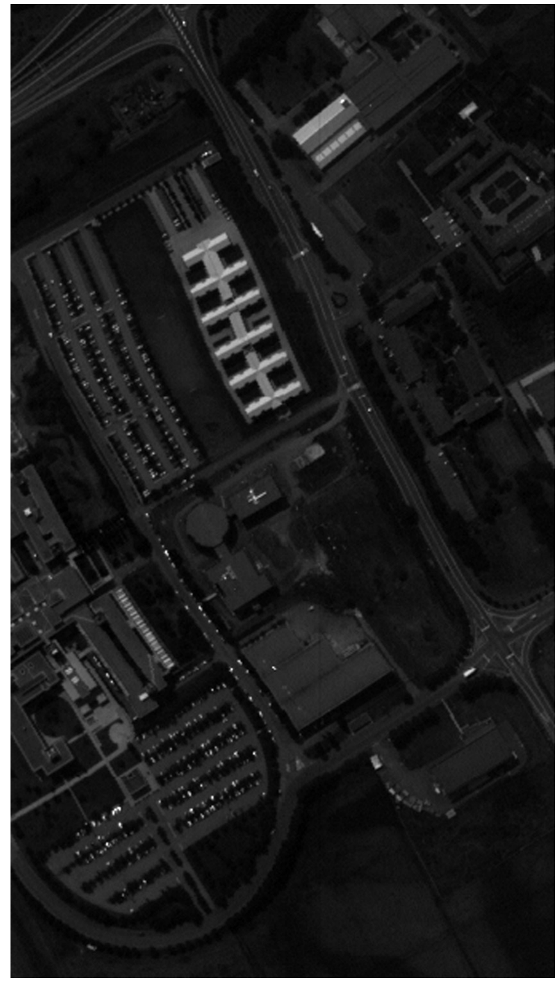

Fig. 1. Sample band of Pavia University.

where $S^{(mrlb)}$ and $S^{(mrlw)}$ are the modified local between-class scatter matrix and the modified local within-class scatter matrix respectively:

$$S^{(mrlb)} = S^{(rlb)}S^{(\mu)} + \delta w^T w \tag{27}$$

$$S^{(mrlw)} = S^{(rlw)} + \delta w^T w \tag{28}$$

where $\delta \in [0, 1]$ is a trade-off parameter. When $\delta = 0$, the algorithm tends to SELF, and when $\delta = 1$, the algorithm tends to SPP. In general, the new algorithm with $0 < \delta < 1$ inherits the characteristics of both SELF and SPP.

3. Experimental results and discussions

3.1. Classification of Pavia University image and Indiana Pines image

3.1.1. Dataset

In this experiment, we use Pavia University image and Indiana image to validate the validity of our algorithm.

Pavia University data was collected by the ROSIS sensor during a flight campaign over the Pavia district in north Italy. 103 spectral bands were used for data acquisition in this dataset comprising of 340×610 pixel images with a geometric resolution of 1.3 m. A sample image has been portrayed in Fig. 1, with the ground truth map in Fig. 2.

AVIRIS hyperspectral image of Indiana Pines consists of 145×145 pixels by 220 bands of radiance data with about

Table 1
Classification accuracy of Pavia University based on different methods.

Class	SPP		SELF		Our method	
	Production accuracy	User accuracy	Production accuracy	User accuracy	Production accuracy	User accuracy
Asphalt	40	52.17	73.33	75.86	86.67	100
Meadows	91.3	87.5	86.96	100	95.65	100
Gravel	85.71	60	85.71	60	100	82.35
Trees	95.24	100	95.24	100	100	100
Painted metal sheets	100	100	100	100	100	100
Bare Soil	84.21	88.89	100	90.48	100	95
Bitumen	65	52	95	95	100	90.91
Self-Blocking Bricks	70.27	78.79	67.57	73.53	94.59	97.22
Shadows	100	96.77	100	100	100	100
Overall accuracy	79.05%		87.14%		96.67%	
Kappa coefficient	0.7618		0.8541		0.9621	

two-thirds agriculture, and one-third forest or other natural perennial vegetation. We have reduced the number of bands to 200 by removing bands covering the region of water absorption. A sample image has been portrayed in Fig. 3, with the ground truth map in Fig. 4.

3.1.2. Experimental results and discussion

In this paper, we determine the target dimensionality in the experiments by means of the maximum likelihood intrinsic dimensionality estimator (Levina and Bickel, 2004). The intrinsic

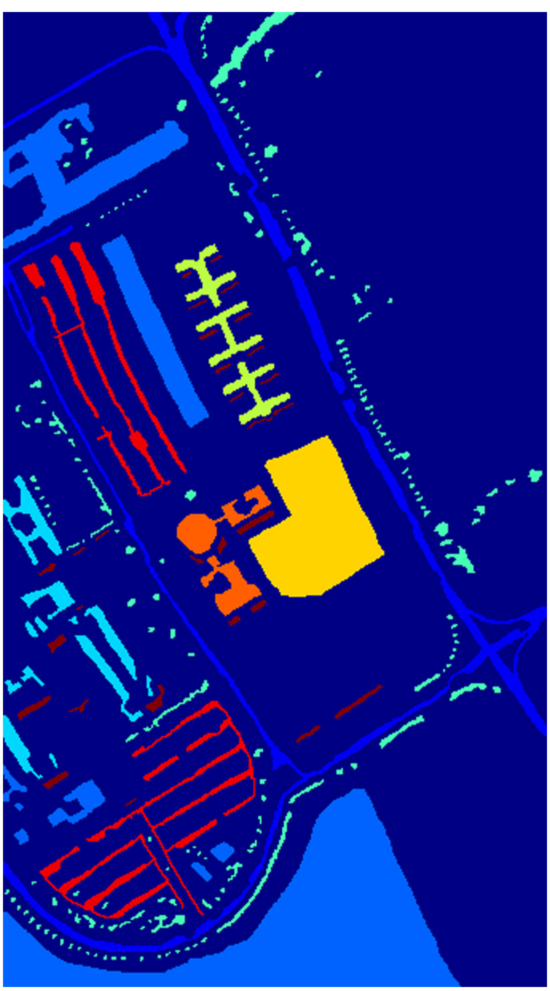

Fig. 2. Ground truth of Pavia University.

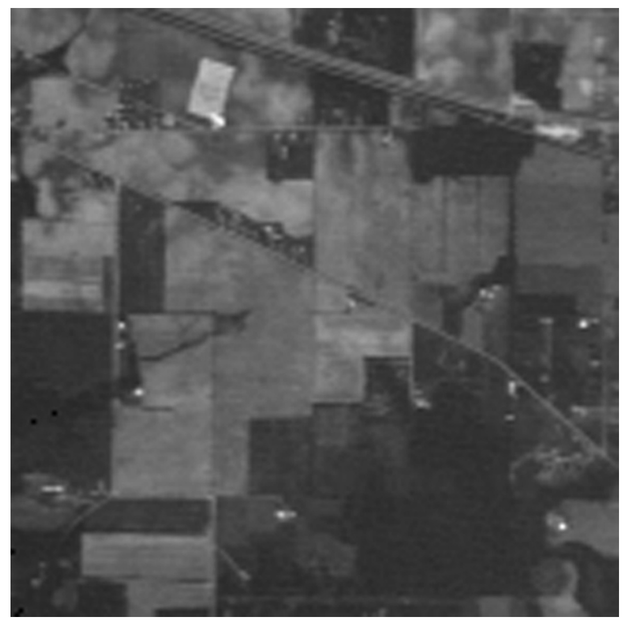

Fig. 3. Sample band of Indiana Pines.

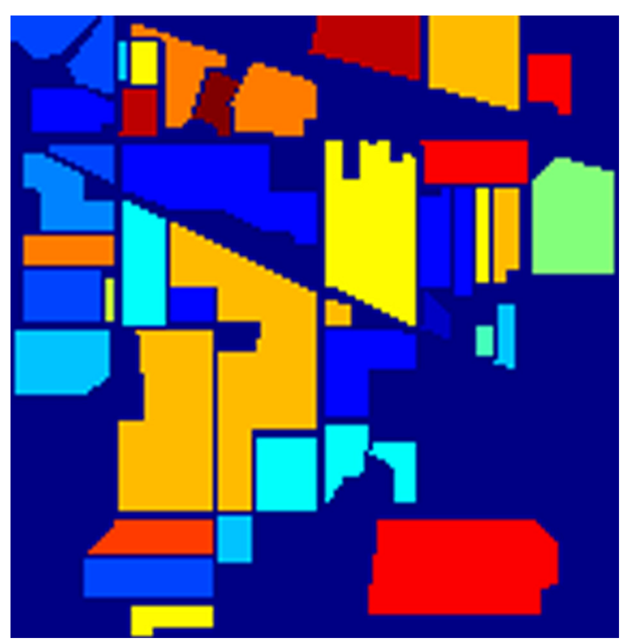

Fig. 4. Ground truth of Indiana Pines.

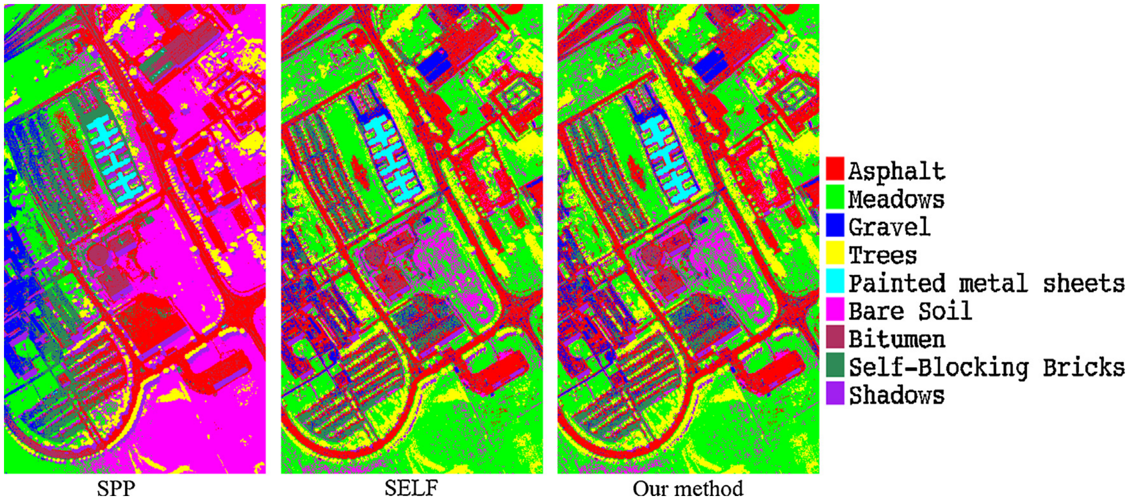

Fig. 5. Classification maps of Pavia University based on different methods.

Table 2
Classification accuracy of Indiana Pines based on different methods.

Class	SPP		SELF		Our method	
	Production accuracy	User accuracy	Production accuracy	User accuracy	Production accuracy	User accuracy
Alfalfa	100	100	100	100	100	100
Corn-notill	27.59	88.89	48.28	66.67	75.86	88
Corn-mintill	50	58.33	57.14	66.67	100	87.5
Corn	25	5.56	100	80	100	100
Grass-pasture	45.45	83.33	100	78.57	90.91	90.91
Grass-trees	100	81.82	94.44	94.44	100	94.74
Grass-pasture-mowed	100	100	100	100	100	100
Hay-windrowed	100	100	100	100	100	100
Oats	70	100	100	100	100	100
Soybean-notill	33.33	54.55	72.22	65	100	69.23
Soybean-mintill	70.97	55	70.97	66.67	64.52	83.33
Soybean-clean	55.56	35.71	77.78	63.64	100	100
Wheat	100	66.67	100	100	100	100
Woods	100	60	100	100	100	100
Buildings-Grass-Trees-Drives	*66.67*	*80*	83.33	83.33	*100*	*100*
Stone-Steel-Towers	100	100	100	100	100	100
Overall accuracy	63.78%		78.92%		89.73%	
Kappa coefficient	0.6046		0.7681		0.8871	

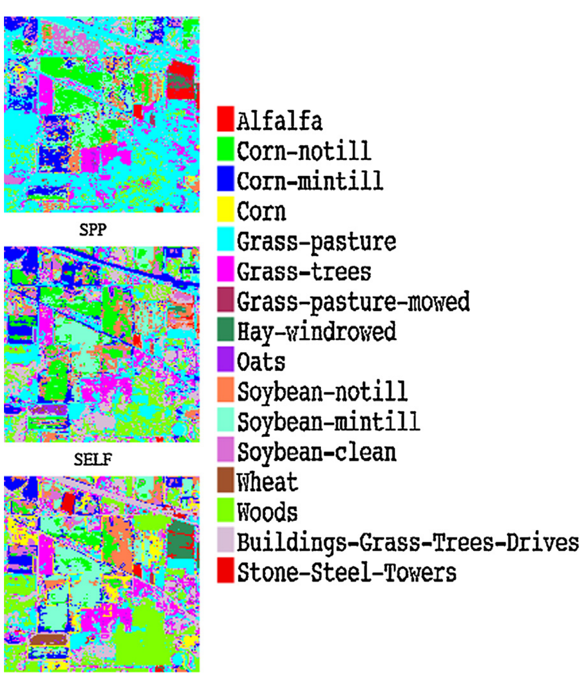

Fig. 6. Classification maps of Indiana Pines based on different methods.

Fig. 7. Sample band of the Real-World Hyperspectral Image.

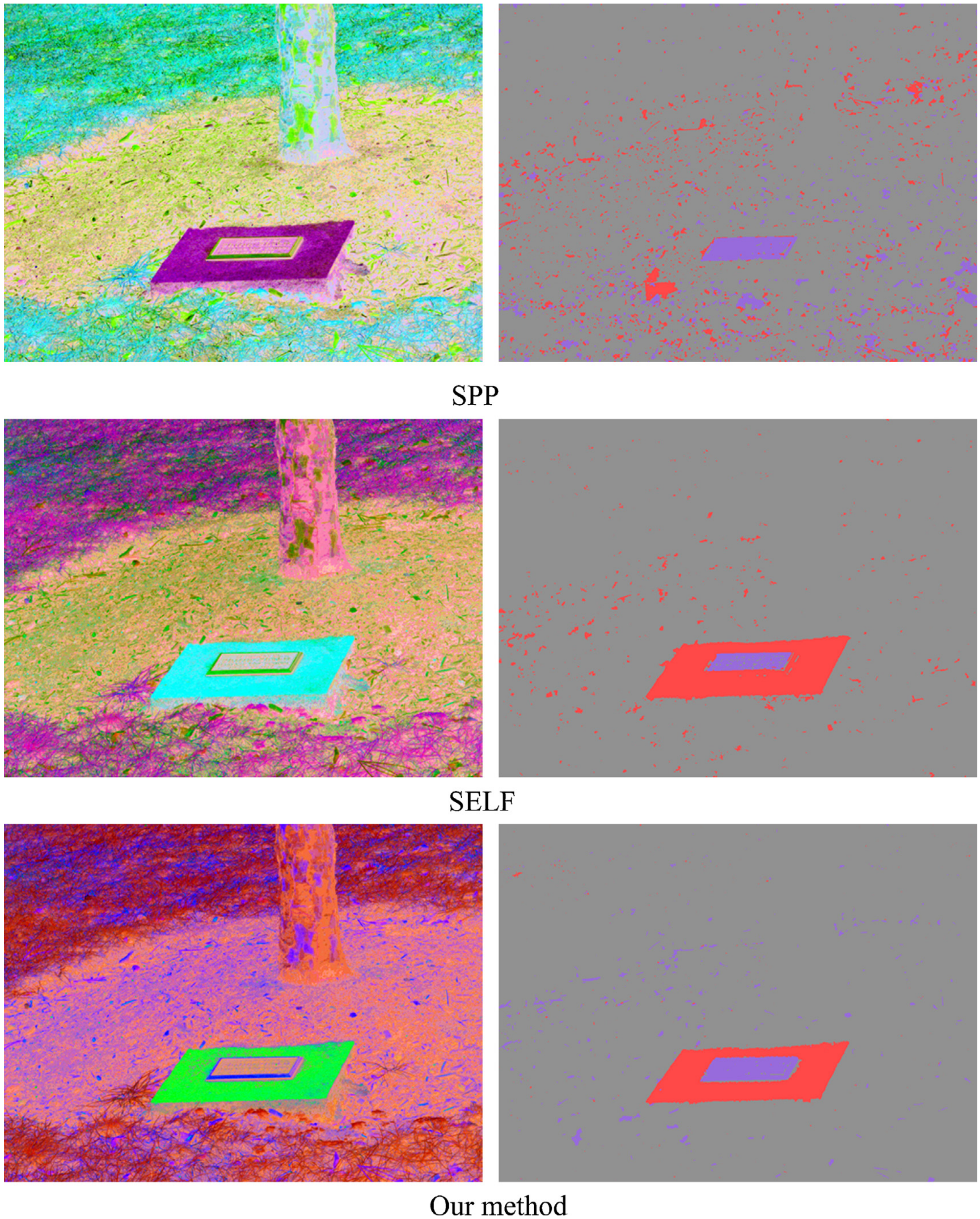

Fig. 8. Feature maps of Real-World Hyperspectral Image after dimensionality reduction.

dimensionality of data is the minimum number of parameters needed to account for the observed properties of the data (Fukunaga, 1990). Then, the intrinsic dimensionality of the Pavia University scene and Indian Pines are calculated as 11 and 10 respectively.

In order to investigate the performances of our method, we adopt minimum distance classifier to verify the classification results after dimensionality reduction. The classification results based on our method with SPP and SELF are compared respectively. Table 1 shows the classification accuracy of Pavia University scene and Fig. 5 shows the corresponding classification maps. And Table 2 illustrates the classification accuracy of Indiana Pines and Fig. 6 displays the corresponding classification maps.

With figures and maps above we can make the following conclusions:

1) SPP performs the worst. The reason is that SPP is an unsupervised linear dimensionality reduction method and the separability cannot be expressed in its linearly projected feature space due to small spectral difference. However, there are several significant gaps among different class accuracy in the classifier based on SPP. For example, in Pavia University some production accuracy of Painted metal sheets and user accuracy of Trees are higher than classifier based on SELF. Classifier based on SPP has high classification accuracy for some special classes such as Meadows and Self-Blocking Bricks, which is mainly because of the sparsity for these special classes.

2) SELF as a semi-supervised dimensionality reduction improves the classification ability by utilizing both labeled samples and unlabeled samples. In most cases, classifier based on SELF can obtain higher classification accuracy than that based on SPP. Since classifier based on SELF has low classification accuracy for some special classes, SPP and SELF have different advantages in dimensionality reduction. The above table also implies that SPP and SELF can compensate for each other's weaknesses.
3) Our method achieves best classification results. The overall accuracy is higher than other methods. Our method can own both advantages of SPP and SELF, which makes classification results effective and robust. This result is mainly because that our method can not only use both the geometrical and discriminant information simultaneously for dimensionality reduction, but also preserve the sparse reconstructive relations among samples in a low-dimensional space.

In assessing the dimensionality reduction approaches we note that the results for the Indiana Pines image are different from those for the University of Pavia image. For the Indiana Pines image, the classification accuracy of our method is improved about 25.95 and 10.81% respectively, which is higher than that of the University of Pavia image. This can be explained that the Indiana Pines image contains more spectral bands. In this case, it appears to be more appropriate to do classification by performing dimensionality reduction. The above obtained results show that our method can improve the classification accuracies of hyperspectral images significantly.

3.2. *Classification of Real-World Hyperspectral Image*

3.2.1. *Dataset*

In addition to the results noted above, Real-World Hyperspectral Image is adopted (Chakrabarti and Zickler, 2011) to test the validity of our algorithm.

The Real-World Hyperspectral Image database adopted is a database of fifty hyperspectral images of indoor and outdoor scenes under daylight illumination, and an additional twenty-five images under artificial and mixed illumination. Each image has a spatial resolution of 1392×1040 with thirty-one spectral measurements at each pixel. A sample image choosed as experimental data has been portrayed in Fig. 7. In this experiment, we compare the feature extraction results based on each dimensionality reduction method.

3.2.2. *Experimental results and discussion*

In this experiment, the intrinsic dimensionality of Real-World Hyperspectral Image is calculated as 6.

In order to investigate the performances of our method, we compare the feature extraction based on our method with SPP and SELF respectively. The left of Fig. 8 shows the corresponding maps after dimensionality reduction. And the right of Fig. 8 illustrates the feature extraction maps based on the left maps.

In order to extract features, this experiment is the combined process of segmenting the Real-World hyperspectral image into regions of pixels, computing their spectral and spatial attributes for each region to create objects, and classifying the objects with supervised classification based on attributes. In this experiment, K nearest neighbor is adopted as the supervised classification method. Experimental results show that the method proposed in this paper can eliminate image noise effectively. Since the separability cannot be expressed in the linearly projected feature space due to small spectral difference, SPP cannot distinguish object feature from the background effectively and it needs to add discriminant information to extract object features from background. Although several feature points are choosed as labeled samples and SELF can make features characteristic, some background areas are extracted as object feature too. Thus SELF needs improvement to eliminate image noise. The method proposed in this paper can use both sparsity and discriminant information to extract features and obtain better result, which shows that our method is not only suitable for classification but also for feature extraction.

4. Conclusion

In this paper, a sparse dimensionality reduction method based on semi-supervised local Fisher discriminant analysis is proposed to realize advantageous complementarities between SELF and SPP, which can get an explicit feature mapping for the out-of-sample extrapolation and improve the classification performance of classifiers by performing dimensionality reduction. And this method also applies to feature extraction. The advantages of the proposed method are that it has highly discriminative ability and produces an explicit nonlinear feature mapping for the out-of-sample extrapolation. Meanwhile, as a semi-supervised method, it utilizes both labeled and unlabeled points simultaneously to help improve the classification ability of hyperspectral image. Experiments prove the method proposed outperforms SPP and SELF. However, due to the characteristics of the maximum likelihood intrinsic dimensionality estimator, it may select target dimensionalities that are suboptimal in the sense which do not minimize the generalization error of the trained classifiers, which will become a study emphasis in our further research.

Acknowledgements

This work is sponsored by National Natural Science Foundation of China (No. 61172174), National Science & Technology Specific Projects (No. 2012YQ16018505 & No. 2013BAH42F03), Program for New Century Excellent Talents in University (No. NCET-12-0426) and the Basic Research Program of Hubei Province (2013CFA024).

References

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1997. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. 19 (7), 711–720.
Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. 15 (6), 1373–1396.
Bellman, R., 1961. Adaptive Control Processes: A Guided Tour. Princeton University Press, Princeton.
Chakrabarti, A., Zickler, T., 2011. Statistics of Real-World Hyperspectral Images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Chen, L., 2009. Curse of Dimensionality. Encyclopedia of Database Systems. Springer, US, pp. 545–546.
Chen, J., Wang, R., Wang, C., 2008. A multiresolution spectral angle-based hyperspectral classification method. Int. J. Remote Sens. 29 (11), 3159–3169.
Demir, B., Erturk, S., 2007. Hyperspectral image classification using relevance vector machines. IEEE Geosci. Remote Sens. 4 (4), 586–590.
Fauvel, M., Benediktsson, J.A., Chanussot, J., et al., 2008. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Trans. Geosci. Remote 46 (11), 3804–3814.
Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. Access Online via Elsevier.
He, X., Yan, S., Hu, Y., et al., 2005. Face recognition using laplacianfaces. IEEE Trans. Pattern Anal. 27 (3), 328–340.
Hsu, P.H., 2007. Feature extraction of hyperspectral images using wavelet and matching pursuit. ISPRS J. Photogramm. 62 (2), 78–92.
Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inform. Theory 14 (1), 55–63.
Landgrebe, D.A., 2003. Signal theory methods in multispectral remote sensing. In: Frontmatter. John Wiley & Sons, Inc.
Levina, E., Bickel, P.J., 2004. Maximum likelihood estimation of intrinsic dimension. In: Advances in Neural Information Processing Systems, pp. 777–784.
Ma, L., Crawford, M.M., Tian, J., 2010. Local manifold learning-based k-nearest-neighbor for hyperspectral image classification. IEEE Trans. Geosci. Remote 48 (11), 4099–4109.
Qiao, L., Chen, S., Tan, X., 2010. Sparsity preserving projections with applications to face recognition. Pattern Recogn. 43 (1), 331–341.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290 (5500), 2323–2326.
Serpico, S.B., Moser, G., 2007. Extraction of spectral channels from hyperspectral images for classification purposes. IEEE Trans. Geosci. Remote 45 (2), 484–495.
Sugiyama, M., 2007. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. J. Mach. Learn. Res. 8, 1027–1061.
Sugiyama, M., Idé, T., Nakajima, S., et al., 2010. Semi-supervised local Fisher discriminant analysis for dimensionality reduction. Mach. Learn. 78 (1–2), 35–61.
Tenenbaum, J.B., de Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290 (5500), 2319–2323.
Trosset, M.W., Priebe, C.E., 2008. The out-of-sample problem for classical multidimensional scaling. Comput. Stat. Data Anal. 52 (10), 4635–4642.
Zelnik-Manor, L., Perona, P., 2004. Self-tuning spectral clustering. In: Advances in Neural Information Processing Systems, pp. 1601–1608.
Zhang, Z.Y., Zha, H.Y., 2005. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. SIAM J. Sci. Comput. 26 (1), 313–338.