

Semi-Supervised Dimensionality Reduction of Hyperspectral Image Based on Sparse Multi-Manifold Learning

Hong Huang¹, Fulin Luo¹, Zezhong Ma², Hailiang Feng¹

¹Key Laboratory of Optoelectronic Technique and Systems, Ministry of Education, Chongqing University, Chongqing, China

²Chongqing Institute of Surveying and Planning for Land, Chongqing, China
Email: hhuang@cqu.edu.cn, luofl@cqu.edu.cn, prislab@gmail.com

Received October 2015

Abstract

In this paper, we proposed a new semi-supervised multi-manifold learning method, called semi-supervised sparse multi-manifold embedding (S³MME), for dimensionality reduction of hyperspectral image data. S³MME exploits both the labeled and unlabeled data to adaptively find neighbors of each sample from the same manifold by using an optimization program based on sparse representation, and naturally gives relative importance to the labeled ones through a graph-based methodology. Then it tries to extract discriminative features on each manifold such that the data points in the same manifold become closer. The effectiveness of the proposed multi-manifold learning algorithm is demonstrated and compared through experiments on a real hyperspectral images.

Keywords

Hyperspectral Image Classification; Dimensionality Reduction, Multiple Manifolds Structure, Sparse Representation, Semi-Supervised Learning

1. Introduction

Hyperspectral image (HSI) contains dozens or even hundreds of contiguous spectral bands, which has been widely used in land cover investigation [1]. However, the high dimensional characteristic of HSI will cause the curse of dimensionality [2]. Consequently, dimensionality reduction (DR) plays a critical role in HSI analysis, especially for the classification task when the number of available labeled training samples is limited.

In recent years, sparse representation (SR) has been successfully applied in HSI dimensionality reduction. SR aims at the sparse reconstructive weight which is associated with sample size. The representative methods include sparse preserving projection (SPP) [3], discriminative learning by sparse representation (DLSP) [4] and discriminant sparse neighborhood preserving embedding (DSNPE) [5].

However, these mentioned works implicitly assume that data points uniformly lie on a single manifold. In real applications, data points may lie on multiple manifolds. In view of this, Ehsan *et al.* [6] proposed an algorithm

called sparse manifold clustering and embedding (SMCE) for simultaneous clustering and DR of data lying in multiple non-linear manifolds. While SMCE is also suffered from the out-of-sample problem, and they do not use the class information provided by training samples, which restricts their discriminating capability. Lu *et al.* [7] introduced a discriminative multi-manifold analysis method by learning discriminative features from image patches, which can perform well when label information is sufficient.

In the real world, the labeled samples are often very difficult and expensive to obtain. The supervised methods cannot work well when lack of training examples, in contrast, unlabeled examples can be easily obtained [8]. In such situations, it can be beneficial to incorporate the information which is contained in unlabeled samples into a learning problem, *i.e.*, semi-supervised learning should be applied instead of supervised learning. At the same time, most of multi-manifold learning methods that have been applied to the processing of HSI rely either on supervised or unsupervised models, and only a few are focused on the semi-supervised setting.

To overcome the above drawbacks, we propose a new DR algorithm named semi-supervised sparse multi-manifold embedding (S^3MME) in this paper. S^3MME utilizes the merits of both sparsity property and multi-manifold learning to better characterize the discriminant property of the data. It exploits both the labeled and unlabeled pixels to adaptively find the local neighborhood of each data point by using an optimization program based on SR, and the selected neighbors are from the same manifold other than other manifolds. The weights associated to the chosen neighbors are automatically obtained simultaneously. It also exploits the wealth of labeled samples in HSI data, and naturally gives relative importance to the labeled ones following a semi-supervised approach. Then, an objective function pushes the homogeneous samples closer to each other is proposed, and the classification performance is further improved.

2. Semi-Supervised Sparse Multi-Manifold Embedding

Suppose that a HSI data set $X = [x_1, x_2, \dots, x_N]$, $x_i \in \mathfrak{R}^m$ sampled from different manifolds $\{\mathcal{M}_i\}_{i=1}^{N_c}$, the first n points are labeled and the rest $N-n$ points are unlabeled. Let $l_i \in \{1, \dots, N_c\}$ denote the class label of x_i . The goal of DR is to map $(\mathbf{X} \in \mathfrak{R}^m) \mapsto (\mathbf{Y} \in \mathfrak{R}^d)$ by $\mathbf{Y} = \mathbf{V}^T \mathbf{X}$, where $d \ll m$. For supervised methods, only labeled points are used for DR. While labeled and unlabeled samples are used for semi-supervised methods.

We assume that naturally occurring data have possibly much fewer degrees of freedom than what the ambient dimension would suggest. Thus, we consider the case where the data lies on or close to multiple low-dimensional manifolds reside in a high dimensional space. To model the manifold structure of data, a similarity graph should be constructed, where nodes represent the data points and edges represent the similarity between data points. Then, a key issue for the similarity graph is to decide which nodes should be connected and how.

To achieve optimal discriminant features, each point will be connected to the points from the same manifold with appropriate weights, while the data pairs from different manifolds are disconnected. At first, for labeled points, we select the data points which have the same class label. Then, we formulate an optimization algorithm as in SMCE to find the unlabeled neighbors from the same manifold. Based on SR techniques, it selects a few data points that are close to x_i and span a low-dimensional affine subspace passing near x_i . For labeled points, the points with the same class label and the unlabeled neighbors from the same manifold are used for the similarity graph. For unlabeled points, we try to find the neighbors which may come from the same manifold.

The sparse solutions $\{c_i\}_{i=1}^N$ can be obtained by SMCE [6], where both labeled and unlabeled points are used. This motivates that for each data point x_i to solve the following weighted sparse optimization program

$$\min \lambda \|Q_i c_i\|_1 + \frac{1}{2} \|X_i c_i\|_2^2 \quad (1)$$

where the ℓ_1 -norm promotes sparsity of the solution, the proximity inducing matrix Q_i is a positive-definite diagonal matrix, and X_i denotes the matrix of normalized vectors $\{x_j - x_i\}$ as follows:

$$X_i \triangleq \begin{bmatrix} \frac{x_1 - x_i}{\|x_1 - x_i\|_2} & \dots & \frac{x_n - x_i}{\|x_n - x_i\|_2} \end{bmatrix} \in \mathfrak{R}^{m \times (N-1)} \quad (2)$$

The elements of Q_i should be chosen such that the points that are closer to x_i have smaller weights, allowing the assignment of non-zero coefficients to them. The diagonal elements of Q_i can be defined as

$$Q_i \triangleq \text{diag}\left(\left[\frac{\|x_j - x_i\|_2}{\sum_{t \neq i} \|x_t - x_i\|_2}\right]_{j \neq i}\right) \in \mathfrak{R}^{(N-1) \times (N-1)} \quad (3)$$

The SR of each data point can be used for the construction of graph. Since the non-zero elements of $\{c_i\}$ are expected to come from the same manifold as of x_i . We first construct a sparse graph $\mathbf{G}_s(\mathbf{V}, \mathbf{E}, \mathbf{W}_s)$ with vertex set $\mathbf{V} = \{x_1, x_2, \dots, x_N\}$, edge set \mathbf{E} , and symmetric weight matrix \mathbf{W}_s . Then, we put an edge between nodes i and j if x_i and x_j are from the same class, or x_i or x_j is unlabeled but c_{ij} is a non-zero element.

Once the graph \mathbf{G}_s is constructed, the affinity weight $W_{s,i} \triangleq [w_{s,i1}, \dots, w_{s,iN}] \in \mathfrak{R}^N$ of x_i can be defined as

$$w_{s,ij} = \begin{cases} \beta \cdot \exp\left(-\frac{c_{ij}^2}{2t_{c_i}^2}\right) \cdot \exp\left(-\frac{d_{ij}^2}{2t_{d_i}^2}\right), & \text{if } c_{ij} \neq 0, x_i \text{ or } x_j \text{ is labeled and } \ell(x_i) = \ell(x_j) \\ \exp\left(-\frac{c_{ij}^2}{2t_{c_i}^2}\right) \cdot \exp\left(-\frac{d_{ij}^2}{2t_{d_i}^2}\right), & \text{if } c_{ij} \neq 0, x_i \text{ or } x_j \text{ is unlabeled} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where β is a trade-off parameter to adjust the contribution of labeled and unlabeled data, $t_{d_i} = \frac{1}{t} \sum_{c_{ij} \neq 0, j} d_{ij}$ and

$$t_{c_i} = \frac{1}{t} \sum_{c_{ij} \neq 0, j} c_{ij}, \quad t, \text{ is the number of non-zero elements in } c_i.$$

The obtained sparse graph built in this way has ideally several connected components, where points in the same manifold are connected to each other and there is no connection between two points from different manifolds. In other words, the weight matrix of this graph has the following form

$$\mathbf{W}_s \triangleq [w_{s,1} \cdots w_{s,N}] = \text{diag}([W_s[1], W_s[2], \dots, W_s[N_c]]) \quad (5)$$

where $W_s[\ell]$ is the weight matrix of data in \mathcal{M}_ℓ , which includes the weight of labeled and some unlabeled data.

The objective of S³MME is embodied as that it minimizes the sum of distances between data pairs which are expected to come from the same manifold. Then, a reasonable criterion for choosing an optimal projection vector with stronger intra-manifold compactness on manifold \mathcal{M}_ℓ is formulated as

$$\min_{\mathbf{V}_\ell} \sum_{ij} \left\| \mathbf{V}_\ell^T x_{\ell,i} - \mathbf{V}_\ell^T x_{\ell,j} \right\|^2 w_s[\ell, ij] = \min \text{tr}(\mathbf{V}_\ell^T X_\ell (\mathbf{D}_s[\ell]) - \mathbf{W}_s[\ell]) X_\ell^T \mathbf{V}_\ell) = \min \text{tr}(\mathbf{V}_\ell^T X_\ell \mathbf{L}_s[\ell]) X_\ell^T \mathbf{V}_\ell) \quad (6)$$

where x_i and x_j are related to the points in the same manifold \mathcal{M}_ℓ which are connected to each other.

$\mathbf{D}_s[\ell]$ is a diagonal matrix with $D_s[\ell, ii] = \sum_j w_s[\ell, ij]$, and $\mathbf{L}_s[\ell] = \mathbf{D}_s[\ell] - \mathbf{W}_s[\ell]$ is the Laplacian matrix.

To remove an arbitrary scaling factor in embedding space, we impose a constraint to vectors \mathbf{V}_ℓ on \mathcal{M}_ℓ , and the objective function on \mathcal{M}_ℓ can be recast as

$$\min \text{tr}(\mathbf{V}_\ell^T \mathbf{X}_\ell \mathbf{L}_s[\ell] \mathbf{X}_\ell^T \mathbf{V}_\ell), \quad \text{s.t. } \mathbf{V}_\ell^T X_\ell \mathbf{D}_s[\ell] X_\ell^T \mathbf{V}_\ell = I \quad (7)$$

Then we apply the Lagrange multipliers to Eq. 7, we can get

$$\mathbf{X}_\ell \mathbf{L}_s[\ell] \mathbf{X}_\ell^T \mathbf{v}_{\ell,i} = \lambda_i X_\ell \mathbf{D}_s[\ell] X_\ell^T \mathbf{v}_{\ell,i} \quad (8)$$

where $\mathbf{v}_{\ell,i}$ is the generalized eigenvector of $\mathbf{X}_\ell \mathbf{L}_s[\ell] \mathbf{X}_\ell^T$ and $X_\ell \mathbf{D}_s[\ell] X_\ell^T$.

Let the column vectors $\mathbf{v}_{\ell,0}, \mathbf{v}_{\ell,1}, \dots, \mathbf{v}_{\ell,d_\ell-1}$ be the solutions of Eq. (8) ordered according to their eigenvalues $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{\ell,d_\ell-1}$. Thus, the embedding is given as follows:

$$\mathbf{x}_{\ell,i} \mapsto \mathbf{y}_{\ell,i} = \mathbf{V}_\ell^T \mathbf{x}_{\ell,i}, \quad \mathbf{V}_\ell = [\mathbf{v}_{\ell,0}, \mathbf{v}_{\ell,1}, \dots, \mathbf{v}_{\ell,d_\ell-1}] \quad (9)$$

Thus, we can use different weight matrix $\{\mathbf{W}_s[\ell]\}_{\ell=1}^{N_c}$ of the i -th manifold as a similarity between points in the corresponding manifold, and obtain a low-dimensional embedding of the data points.

3. Experiments and Discussion

3.1. Experimental Design

The goal of the experiments is to investigate the effectiveness of the proposed algorithms for classification of PaviaU hyperspectral data set. In each experiment, the data set was divided into training and test sets, and we randomly split the training set into the labeled and unlabeled set. The number of labeled samples l is varied as $\{10, 20, 40, 80\}$ per class, while the number of unlabeled samples u is chosen in $\{100, 500, 1000, 2000, 3000\}$. For supervised DR methods, only the labeled set is used for training, while semi-supervised DR methods can utilize both labeled and unlabeled data. Then, all testing samples are projected into embedding space.

After that, reconstruction err classifier is used for multiple manifold classification [9], and nearest neighborhood (1-NN) is employed for other methods classification in all experiments. The classifier was evaluated against the test set, and we use overall classification accuracies (OAs) and the kappa coefficients (κ) to evaluate the classification results. We repeat the classification process 10 times in each condition.

We compare S³MME with several representative DR algorithms such as PCA, LDA, LPP, LFDA, SPP, DLSP, DSNPE, semi-supervised sub-manifold discriminant analysis (S³MPE) [10] and semi-supervised discriminant analysis (SDA) [11]. The parameter β is set as 40 with cross-validation in S³MME. Note that ε for SPP, DLSP and DSNPE is generally fixed across various instances of the problem, and we empirically set it to 0.1 in our experiments. For manifold learning methods, the number of nearest neighbors k is set to be 7. For all methods, the dimension of embedding features is set as 30.

3.2. Classification of PaviaU Data Set

The PaviaU dataset was acquired by the reflective optics system imaging spectrometer (ROSIS) sensor during flight campaigns in 2003 over the Pavia University, northern Italy. It consists of 610×340 pixels and 115 spectral reflectance bands in the wavelength range 430-860 nm with a high spatial resolution of 1.3 m per pixel. After removing noisy and water absorption bands, it reduced to 103 channels. **Figure 1(a)** shows a false color composite of the image, and **Figure 1(b)** shows the nine ground truth classes of interest.

In the first experiment, we evaluated the classification performance of S³MME. **Table 1** reports the average classification performance achieved by different DR algorithms, where the OAs and the κ coefficients are displayed for a varying number of labeled samples (10, 20, 40, 80 per class) and 3000 unlabeled samples. For illustrative purposes, **Figure 1** shows the classification maps obtained for different methods for the case of $l = 40$ per class and $u = 3000$ as in **Table 1**.

Table 1. OAs (in percent) and κ coefficients (in the brackets) with different numbers of labeled samples per class.

DR	$l = 10$	$l = 20$	$l = 40$	$l = 80$
Original	63.64 ± 4.35 (0.5496)	67.32 ± 3.04 (0.5924)	71.81 ± 1.34 (0.6439)	74.75 ± 0.96 (0.6780)
PCA	62.79 ± 3.98 (0.5428)	67.06 ± 2.25 (0.5907)	71.42 ± 1.11 (0.6395)	74.74 ± 0.99 (0.6780)
LDA	30.18 ± 4.39 (0.1737)	58.85 ± 1.72 (0.4885)	70.67 ± 2.04 (0.6284)	77.51 ± 1.46 (0.7101)
LPP	63.95 ± 3.97 (0.5569)	69.04 ± 2.62 (0.6129)	74.03 ± 1.53 (0.6712)	77.49 ± 1.71 (0.7121)
LFDA	40.59 ± 2.69 (0.2862)	65.82 ± 3.04 (0.5744)	67.78 ± 1.80 (0.5961)	68.10 ± 1.25 (0.6002)
SPP	64.45 ± 2.66 (0.5607)	68.16 ± 2.51 (0.6018)	71.89 ± 2.23 (0.6443)	75.01 ± 1.50 (0.6802)
DLSP	60.08 ± 4.71 (0.5433)	66.59 ± 3.01 (0.5840)	69.58 ± 2.66 (0.6177)	72.35 ± 2.02 (0.6496)
DSNPE	63.62 ± 4.36 (0.5494)	67.31 ± 3.05 (0.5923)	71.72 ± 1.37 (0.6428)	74.58 ± 0.94 (0.6761)
S3MPE	63.68 ± 4.45 (0.5502)	67.08 ± 3.21 (0.5897)	71.73 ± 1.33 (0.6434)	75.40 ± 1.12 (0.6861)
SDA	65.69 ± 4.07 (0.5748)	70.18 ± 2.62 (0.6269)	75.94 ± 1.07 (0.6942)	80.54 ± 1.77 (0.7491)
S3MME	72.24 ± 3.78 (0.6523)	81.60 ± 1.91 (0.7619)	84.39 ± 1.98 (0.7968)	86.73 ± 1.21 (0.8267)

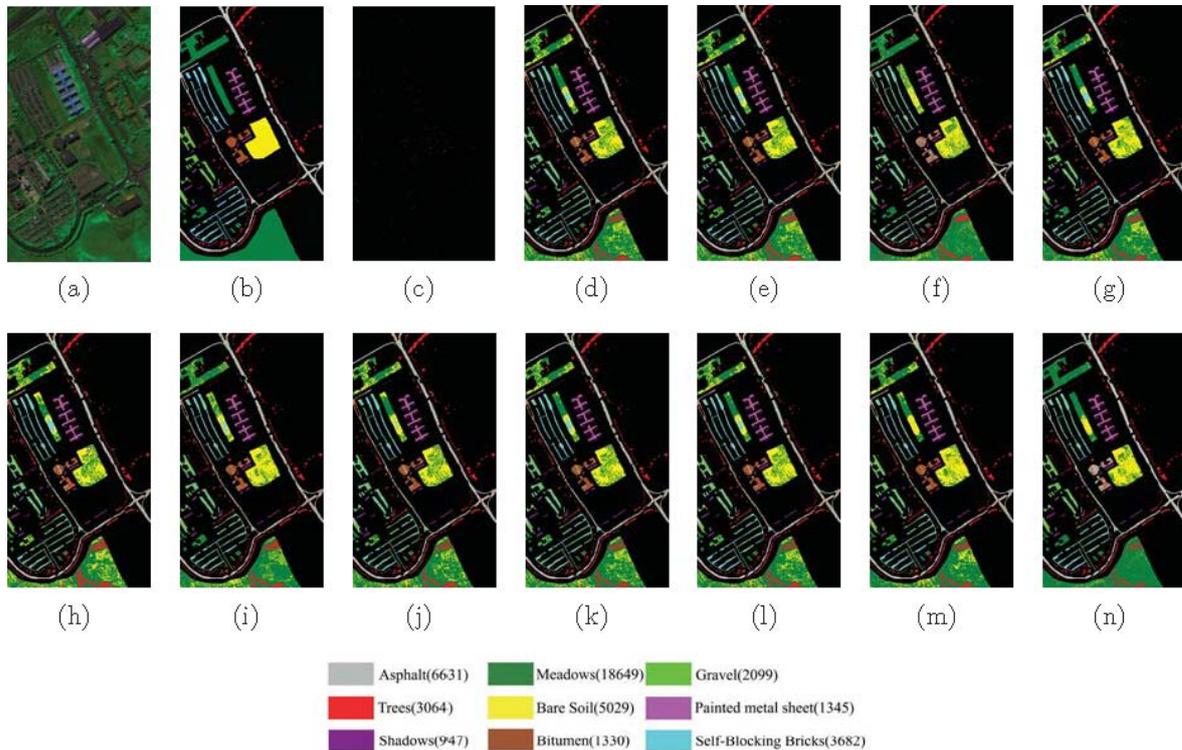


Figure 1. Classification maps with different algorithms ($l = 40$, $u = 3000$). (a) False-color composition (bands 5, 40 and 15 for RGB); (b) Ground truth; (c) Labeled data; (d) Original (0.7411); (e) PCA (0.7407); (f) LDA (0.7527); (g) LPP (0.7746); (h) LFDA (0.6969); (i) SPP (0.7589); (j) DLSP (0.7244); (k) DSNPE (0.7410); (l) S3MPE (0.7453); (m) SDA (0.7862); (n) S3MME (0.8491). Note that OAs are given in parentheses.

As can be seen from **Table 1**, the classification performance improves for all methods as more data samples are used for training. The supervised methods, *i.e.* LDA and LFDA, degrade the performance of classification accuracy when the labeled sample size of training set is small, mainly due to the overfitting or overtraining. Our S^3 MME method produces better classification results than other methods in all situations, and the improvement is particularly significant when low number of labeled samples is used. The reason is that S^3 MME exploits the wealth of labeled and unlabeled samples to adaptively select neighbors from the same manifold and discovers the multi-manifold structure in HSI data, which respects both sparsity property and semi-supervised learning to better characterize the discriminant property of data.

By observing the classification results shown in **Figure 1**, the numerical results are confirmed by visual inspection of the classification maps. The S^3 MME method produces more homogenous areas and better classification maps than other methods, especially in *Meadows*, *Trees*, and *Bare Soil*.

To investigate the influence of the numbers of unlabeled data on the performance of S^3 MME, we evaluate the classification accuracy using a small number of labeled samples ($l = 10$ per class) and different numbers of unlabeled samples ($u = 100, 500, 1000, 2000, 3000$), which are randomly selected for training. The classification results averaged over ten runs are shown in **Figure 2**.

An effective semi-supervised learning method can improve the performance when the number of the available unlabeled data increases. As expected, the OAs and κ coefficients of S^3 MME is significantly improved when the number of the available unlabeled data increases.

4. Experiments and Discussion

In this paper, we proposed a novel multi-manifold learning method for DR and classification of hyperspectral image. S^3 MME exploits both the labeled and unlabeled samples to selected neighbors from the same manifold based on sparse representation, and naturally gives relative importance to the labeled ones through a semi-supervised neighborhood graph. Then it tries to extract discriminative features on each manifold such that the

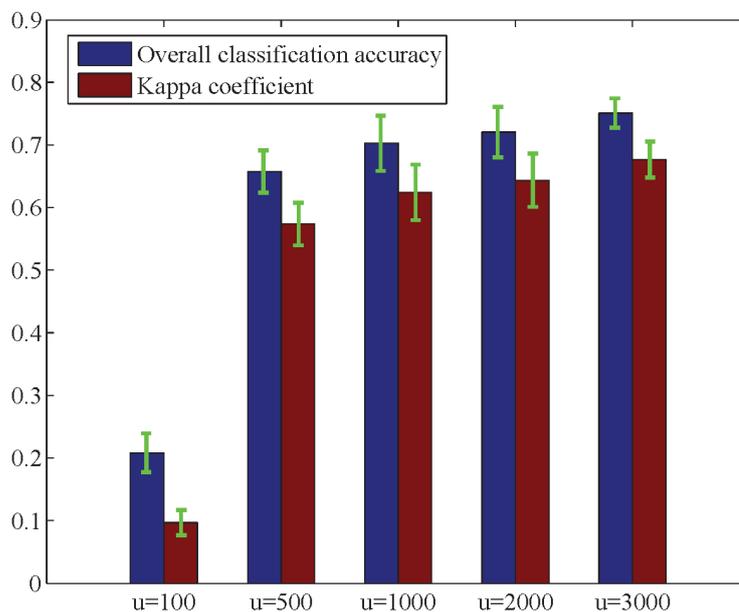


Figure 2. The classification results of S^3MME using different numbers of unlabeled samples.

data points in the same manifold become closer. The S^3MME method has been applied to a real hyperspectral image data set, and extensive experiments and comparisons have been conducted. Promising results have been obtained demonstrating the superiority of the proposed multi-manifold learning methods in hyperspectral image classification. The multi-manifold learning method proposed here are not utilize any spatial information in hyperspectral image. In our future work, we are going to consider the multi-manifold learning methods which make full use of both the spectral and spatial information provided by hyperspectral image.

Acknowledgements

This work is supported by National Science Foundation of China (41371338, 61101168), the Basic and Frontier Research Programs of Chongqing (cstc2013jcyjA40005), the Fundamental Research Funds for the Central Universities of China (106112013CDJZR125501, 1061120131204), and Chongqing University Postgraduates' Innovation Project (CYB15052). The authors would like to thank the anonymous reviewers for their constructive advice.

References

- [1] Chen, Y., Nasrabadi, N. and Tran, T. (2013) Hyperspectral Image Classification via Kernel Sparse Representation. *IEEE Transactions on Geoscience Remote Sensing*, **51**, 217-231. <http://dx.doi.org/10.1109/TGRS.2012.2201730>
- [2] Shao, Z. and Zhang, L. (2014) Sparse Dimensionality Reduction of Hyperspectral Image Based on Semi-Supervised Local Fisher Discriminant Analysis. *International Journal of Applied Earth Observation and Geoinformation*, **31**, 122-129. <http://dx.doi.org/10.1016/j.jag.2014.03.015>
- [3] Qiao, L., Chen, S. and Tan, X. (2010) Sparsity Preserving Projections with Applications to Face Recognition. *Pattern Recognition*, **43**, 331-341. <http://dx.doi.org/10.1016/j.patcog.2009.05.005>
- [4] Zang, F. and Zhang, J.S. (2011) Discriminative Learning by Sparse Representation for Classification. *Neurocomputing*, **74**, 2176-2183. <http://dx.doi.org/10.1016/j.neucom.2011.02.012>
- [5] Lu, G.F., Jin, Z. and Zou, J. (2012) Face Recognition Using Discriminant Sparsity Neighborhood Preserving Embedding. *Knowledge-Based Systems*, **31**, 119-127. <http://dx.doi.org/10.1016/j.knosys.2012.02.014>
- [6] Elhamifar, E. and Vidal, R. (2006) Sparse Manifold Clustering and Embedding. In: *Advances in Neural Information Processing Systems*, 609-616.
- [7] Lu, J.W., Peng, Y. and Wang, G. (2011) Discriminative Multi-Manifold Analysis for Face Recognition from a Single

-
- Training Sample per Person. *IEEE International Conference on Computer Vision*, 1943-1950. <http://dx.doi.org/10.1109/iccv.2011.6126464>
- [8] Shi, Q., Zhang, L.P. and Du, B. (2013) Semisupervised Discriminative Locally Enhanced Alignment for Hyperspectral Image Classification. *IEEE Transactions on Geoscience Remote Sensing*, **51**, 4800-4815. <http://dx.doi.org/10.1109/TGRS.2012.2230445>
- [9] Wang, L.Z., Huang, H. and Feng, H.L. (2012) Multi-Linear Local and Global Preserving Embedding and Its Application in Hyperspectral Remote Sensing Image Classification. *Journal of Computer-Aided Design & Computer Graphics*, **24**, 780-786.
- [10] Song, Y.Q., Nie, F.P. and Zhang, C.S. (2008) Semi-Supervised Sub-Manifold Discriminant Analysis. *Pattern Recognition*, **29**, 1806-1813. <http://dx.doi.org/10.1016/j.patrec.2008.05.024>
- [11] Cai, D., He, X.F. and Han, J.W. (2007) Semi-Supervised Discriminant Analysis. In: *Proceedings of the 11th IEEE International Conference on Computer Vision*, 609-616. <http://dx.doi.org/10.1109/iccv.2007.4408856>