

Letter

Multiview Locally Linear Embedding for Spectral-Spatial Dimensionality Reduction of Hyperspectral Imagery

Haochen Ji and Zongyu Zuo, *Senior Member, IEEE*

Dear Editor,

Dimensionality reduction (DR) plays a prominent role in the processing of hyperspectral imagery. Considering the high dimensionality of multiple features, this letter presents a new unsupervised DR method named multiview locally linear embedding (MLLE), which captures the local linearity and global nonlinearity of the data sufficiently. We formulate MLLE as an optimization problem, where the diversity and complementarity of multiple features is fully exploited. An effective alternating optimization scheme is derived, and a linear model based on ridge regression is extended to alleviate the high correlation among single-view features. Experimental results on the Indian Pines and Pavia University datasets demonstrate the superiority of our proposed MLLE.

Introduction: Hyperspectral imaging, as an emerging technology in the field of remote sensing, makes it possible to collect data about spectral channels and spatial pattern of ground objects simultaneously [1]. Any pixel of hyperspectral imagery (HSI) can be represented by different views [2], e.g., spectral, texture, and shape, providing detailed and complementary information on a variety of applications especially in classification. Nevertheless, the high dimension of multiple features also introduces several difficulties, such as the computational burden and the Hughes phenomenon [3]. Therefore, it is necessary to reduce the dimensionality of HSI, which learns a common low-dimensional embedding of different views.

Over the past decades, various unsupervised DR methods, which can be categorized into feature selection and feature extraction (FE), have been proposed for HSI. This letter will focus on unsupervised FE. Instead of conventional linear technique [4], manifold learning-based methods aim to explore the nonlinear structure of HSI, including Isomap [5], locally linear embedding (LLE) [6], Laplacian eigenmaps (LE) [7], etc. In [8], a local manifold learning is proposed for hyperspectral DR, which obtains a discriminative low-dimensional embedding. To enhance the aggregation of the data, the work [9] models the scatter information and the dual graph structure of HSI. Additionally, Laplacian regularized collaborative representation projection (LRCRP) [10] attempts to preserve the local manifold structure fully, by adding Laplacian regularization and local enhancement into the collaborative representation projection. However, these methods are merely designed for spectral signatures, and concatenating features from different views as a new vector serves no practical purpose.

To conquer the aforementioned difficulties, a multiple features combining (MFC) framework [2] is presented, which learns a unified low-dimensional representation successfully. Afterwards, [11] investigates an unsupervised approach which incorporates multiple features into orthogonal nonnegative matrix factorization (ONMF) for effective DR. Rather than utilizing the spatial feature directly, a novel FE approach based on superpixelwise principal component analysis (SuperPCA) is proposed in [12], which performs PCA on

Corresponding author: Zongyu Zuo.

Citation: H. C. Ji and Z. Y. Zuo, "Multiview locally linear embedding for spectral-spatial dimensionality reduction of hyperspectral imagery," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 6, pp. 1091–1094, Jun. 2022.

The authors are with the Seventh Research Division, Beihang University (BUAA), Beijing 100191, China (e-mail: jhc03@buaa.edu.cn; zzybobby@buaa.edu.cn).

Digital Object Identifier 10.1109/JAS.2022.105638

various homogeneous regions.

Among all these methods, the MFC framework grabs our attention. Firstly, MFC constructs the weight matrix of each single view separately based on LE. Then, the objective functions of different views are incorporated as a unified one. By optimizing a set of combination weights, a unified low-dimensional embedding is obtained. Although the framework has proven to be effective in [2], [13], [14], the heat kernels in the weight matrices of all views are artificially designed, leading to relatively weak representation of local manifold without exploiting the natural property of local neighbors. Besides, when coping with the out-of-samples problem, the linear version of MFC neglects the multicollinearity of features in single view, particularly in the spectral domain. Therefore, in this letter, we propose a new unsupervised DR method named MLLE. Specifically, MLLE explores and models the manifold structure of different observation spaces assuming local linearity. To fully exploit the complementary information of various views, we use the reciprocal of the combination weights to determine the contribution of each view. Moreover, we have further extended to a new linear model based on the ridge regression. Finally, experimental study verifies the effectiveness of the proposed method. Compared to MFC, the classification performance of MLLE is significantly improved, even comparable to those of state-of-the-art techniques.

Proposed method:

1) MLLE: Given a hyperspectral image with N pixels, it can be described by V different views $\{\mathbf{X}^{(v)} = [\mathbf{x}_1^{(v)}, \dots, \mathbf{x}_N^{(v)}] \in \mathbb{R}^{D_v \times N}, v = 1, \dots, V\}$, where D_v is the dimension of features in the v th view. MLLE aims at learning a common low-dimensional embedding $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbb{R}^{d \times N}$, in which $d \ll \sum_{v=1}^V D_v$.

Considering the v th single view, we firstly find k nearest neighbors of each pixel using pairwise Euclidean distance. Suppose $\mathbf{x}_{ij}^{(v)}$ denotes the j th neighbor of pixel $\mathbf{x}_i^{(v)}$ in the v th view, so the matrix including k nearest neighbors (k NN) of $\mathbf{x}_i^{(v)}$ can be denoted as $\mathbf{X}_i^{(v)} = [\mathbf{x}_{i1}^{(v)}, \dots, \mathbf{x}_{ik}^{(v)}]$.

Afterwards, to explore the local linearity among neighboring pixels, the errors of linear reconstruction should be minimized, which can be formulated as a following optimization problem:

$$\begin{aligned} \arg \min_{\mathbf{w}_i^{(v)}} \sum_{i=1}^N \left\| \mathbf{x}_i^{(v)} - \sum_{j=1}^k w_{ij}^{(v)} \mathbf{x}_{ij}^{(v)} \right\|_2^2 \\ \text{s.t.} \quad \sum_{j=1}^k w_{ij}^{(v)} = 1, \quad i = 1, \dots, N \end{aligned} \quad (1)$$

where $\mathbf{w}_i^{(v)} = [w_{i1}^{(v)}, \dots, w_{ik}^{(v)}]^T$ is the weight vector including the reconstruction weights between the i th pixel and its neighbors. Let $\mathbf{G}_i^{(v)} = (\mathbf{x}_i^{(v)} \mathbf{1}^T - \mathbf{X}_i^{(v)})^T (\mathbf{x}_i^{(v)} \mathbf{1}^T - \mathbf{X}_i^{(v)})$, where $\mathbf{1}$ is a column vector whose elements are 1. According to the Lagrange multiplier method, the optimal solution to (1) is denoted as

$$\mathbf{w}_i^{(v)} = \frac{(\mathbf{G}_i^{(v)})^{-1} \mathbf{1}}{\mathbf{1}^T (\mathbf{G}_i^{(v)})^{-1} \mathbf{1}}. \quad (2)$$

By solving (1) of each pixel in single view independently, we get the coefficient matrices $\mathbf{C}^{(v)}$ whose entries meet

$$c_{ij}^{(v)} = \begin{cases} w_{ij}^{(v)}, & \text{if } \mathbf{x}_j^{(v)} \in kNN(\mathbf{x}_i^{(v)}) \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

LLE says that in a low-dimensional space, the linear combination between each original feature and its neighbors should be preserved.

Therefore, the goal of linear embedding in single view can be denoted as

$$\arg \min_{\mathbf{Y}} \sum_{i=1}^N \left\| \mathbf{y}_i - \sum_{j=1}^N c_{ij}^{(v)} \mathbf{y}_j \right\|_2^2 = \text{tr}(\mathbf{Y} \mathbf{L}^{(v)} \mathbf{Y}^T)$$

s.t. $\mathbf{Y} \mathbf{Y}^T = \mathbf{I}$ (4)

where $\text{tr}(\cdot)$ indicates the trace of matrix, \mathbf{I} is the identity matrix, and $\mathbf{L}^{(v)} := (\mathbf{I} - \mathbf{C}^{(v)})(\mathbf{I} - \mathbf{C}^{(v)})^T$ encoding the objective function of v th view. Equation (4) embeds the raw features of HSI into a low-dimensional space which can be effective on preservation of locally linear structure.

In order to combine multiple features as a unified representation, a simple and effective method is to give nonnegative combination weights $\alpha = [\alpha_1, \dots, \alpha_V]^T$ to objective functions of different views. To avoid the weights being zero, MFC makes a relaxation by setting α_v^r to replace α_v , with $r > 1$, which introduces an additional parameter r . Assuming that $\text{tr}(\mathbf{Y} \mathbf{L}^{(u)} \mathbf{Y}^T)$ is minimal over different views. If $r \rightarrow 1$, α_u is close to 1. As r becomes larger, different weights will get closer. Let $\delta_v = \alpha_v^r / \sum_{v=1}^V \alpha_v^r$ denoting the normalized weights, we have $\delta_u \gg \delta_v$ (where $v \neq u$) as long as the difference of weights exists. That is to say, introducing $r > 1$ may attenuate the contributions to the low-dimensional embedding \mathbf{Y} of the other views implicitly. So in this letter, we pick $1/\alpha_v$ (i.e., $r = -1$) to replace α_v , which avoids the challenge of adjusting r . And most importantly, this could fully mine the diversity and complementarity of multiple features. In this case, the calculation of low-dimensional embedding can be restated as

$$\arg \min_{\mathbf{Y}, \alpha} \sum_{v=1}^V \frac{1}{\alpha_v} \text{tr}(\mathbf{Y} \mathbf{L}^{(v)} \mathbf{Y}^T)$$

s.t. $\mathbf{Y} \mathbf{Y}^T = \mathbf{I}, \quad \sum_{v=1}^V \alpha_v = 1, \quad \alpha_v > 0.$ (5)

Obviously, the smaller α_v is, the more important role the v th view plays in the low-dimensional representation. Note that (5) is a nonlinearly constrained nonconvex optimization problem, so it is almost impossible to find the global optimal solution. Therefore, in the next section, we will adopt the alternating optimization strategy to solve it.

2) Alternating optimization: In order to solve for \mathbf{Y} and α simultaneously, we develop an alternating iteration optimization approach to attaining a local minimum, which is detailed as follows.

a) Fix α and update \mathbf{Y} . The optimization can be reduced to

$$\arg \min_{\mathbf{Y}} \text{tr}(\mathbf{Y} \mathbf{L} \mathbf{Y}^T)$$

s.t. $\mathbf{Y} \mathbf{Y}^T = \mathbf{I}$ (6)

where $\mathbf{L} = \sum_{v=1}^V \frac{1}{\alpha_v} \mathbf{L}^{(v)}$, which is a symmetric matrix. According to [15], the global optimal \mathbf{Y} is given as the eigenvectors associated with the smallest d eigenvalues of \mathbf{L} .

b) Fix \mathbf{Y} and update α . Then, the optimization becomes

$$\arg \min_{\alpha} \sum_{v=1}^V \frac{1}{\alpha_v} \text{tr}(\mathbf{Y} \mathbf{L}^{(v)} \mathbf{Y}^T)$$

s.t. $\sum_{v=1}^V \alpha_v = 1, \quad \alpha_v > 0.$ (7)

Note that $\mathbf{L}^{(v)}$ is symmetric positive semi-definite, so the global optimal solution to (7) can be given as

$$\alpha_v = \frac{\sqrt{\text{tr}(\mathbf{Y} \mathbf{L}^{(v)} \mathbf{Y}^T)}}{\sum_{v=1}^V \sqrt{\text{tr}(\mathbf{Y} \mathbf{L}^{(v)} \mathbf{Y}^T)}}. \quad (8)$$

3) Linearization: Using the alternating optimization above, we map the hyperspectral data into an underlying lower space. However, the cost of storing and manipulating the matrix \mathbf{L} makes it infeasible for large scale samples. For example, considering the Pavia University dataset containing more than 2×10^5 pixels, both the computing and eigenvalue decomposition of \mathbf{L} are intractable. This poses a tricky problem in practical application.

In [2], a linear approximation trained by a subset of samples has been proposed to construct an explicit projection matrix, which reduces the burden of storage and computation. However, in this linear regression model, the high correlation (i.e., multicollinearity) of features from the same view is neglected, especially in the spectral perspective. This phenomenon easily results in poor generalization of the projection matrix. To mitigate the multicollinearity among multiple features, in this letter, a ridge regression model is designed. Therefore, when randomly selecting a subset of pixels $\{\mathbf{X}_s^{(v)} \in \mathbb{R}^{D_v \times n}, v = 1, \dots, V, n \ll N\}$ and its low-dimensional features $\mathbf{Y}_s \in \mathbb{R}^{d \times n}$, an unconstrained optimization with respect to the projection matrix \mathbf{U} can be summarized as

$$\arg \min_{\mathbf{U}} \left(\|\mathbf{Y}_s - \mathbf{U}^T \mathbf{X}_s\|_F^2 + \beta \|\mathbf{U}\|_F^2 \right) \quad (9)$$

where $\mathbf{X}_s = [\mathbf{X}_s^{(1)T}, \dots, \mathbf{X}_s^{(V)T}]^T$ and $\beta > 0$ is a regularization parameter. The optimal solution to (9) can be easily solved as

$$\mathbf{U} = (\mathbf{X}_s \mathbf{X}_s^T + \beta \mathbf{I})^{-1} \mathbf{X}_s \mathbf{Y}_s^T. \quad (10)$$

4) Computational complexity analysis: According to the above, the computational complexity of MLLE consists of three parts. To be specific, the calculation of the reconstruction weights of each view attempts to solve n set of linear equations in size of $k \times k$ requiring $O(nk^3)$. The computational complexity of the alternating optimization is $O(n^3) \times T$, where T is the number of iterations. With respect to linearization, the computational complexity is $O(n^3)$. Therefore, the total time complexity of MLLE is $O(nk^3 + n^3)$.

Experiments: To evaluate the performance of MLLE, the pixel-oriented classification experiments are conducted on two HSI datasets, compared with some state-of-the-art unsupervised methods.

1) Datasets: Collected by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor over North-western Indiana in 1992, the Indian Pines (IP) dataset provides detailed information through 220 spectral channels ranging from 0.4 μm to 2.5 μm . By removing 20 bands with high water absorption, the size of data available is $145 \times 145 \times 200$. Specifically, the first dataset contains 16 classes of interests with a total of 10 249 labeled pixels. The second dataset was called the Pavia University (PU) dataset, which was captured by reflective optics system imaging spectrometer (ROSIS) sensor over the University of Pavia in 2002. PU has 103 bands and 610×340 pixels, including nine different groundtruth classes.

2) Experimental design: In the experiment, spectral, texture, and shape, which can profile different feature attributes, are introduced as input of the proposed method. Specifically, the spectral feature of a pixel can be characterized by its reflectance value of all bands. To exploit the texture details, the 2-D Gabor filters [16] with 12 orientations and 5 scales are implemented on the first principal component of IP dataset (only 8 orientations for PU). Finally, we use the pixel shape index (PSI) method [2] to extract shape feature with 20 orientations.

There are three parameters needed to be determined. The number of neighbors k and the regularization parameter β are selected from $[50, 60, \dots, 120]$ and $55 \times [10^{-5}, 10^{-4}, \dots, 10^5]$, respectively. After cross-validation, we pick $k = 100$ and $\beta = 5$ for IP as well as $k = 100$ and $\beta = 0.5$ for PU. Besides, the number of the extracted features d should be no more than 50 in light of the purpose of DR. With respect to the projection matrix, we randomly sample 5000 pixels to train it. The 1-nearest neighbor (1NN) is used as classifier, where the number of training and test samples of each class from two datasets are showed in Tables 1 and 2. Moreover, we choose the training data

Table 1. Number of Samples as well as Classification Accuracy of Various Methods Using INN on IP

No.	Class name	Train	Test	OSF/200	CAF/280	RLMR/50	LRCRP/30	MFC/15	MFONMF/32	SuperPCA/30	MLLE/50
1	Alfalfa	10	36	56.94	98.61	95.56	61.39	91.30	95.65	100.00	97.50
2	Corn-notill	50	1378	40.13	74.77	73.30	43.67	65.34	86.22	83.22	90.01
3	Corn-mintill	50	780	50.60	85.31	72.64	50.03	62.29	83.48	92.21	94.03
4	Cron	50	187	65.94	98.66	81.55	67.59	79.32	95.78	98.40	98.13
5	Grass-pasture	50	433	84.36	92.45	94.57	84.64	77.23	96.27	97.14	96.77
6	Grass-trees	50	680	84.51	95.06	99.12	85.46	88.08	98.08	99.34	99.54
7	Grass-pasture-mowed	10	18	89.44	94.44	93.33	87.78	82.14	96.43	97.22	100.00
8	Hay-windrowed	50	428	93.46	97.87	99.79	91.66	94.98	100.00	99.58	100.00
9	Oats	10	10	62.00	96.00	100.00	64.00	90.00	100.00	100.00	100.00
10	Soybean-notill	50	922	66.48	79.32	89.13	66.23	74.18	80.04	93.11	94.75
11	Soybean-mintill	50	2405	48.91	65.36	71.00	50.90	61.18	73.75	93.13	88.05
12	Soybean-clean	50	543	48.21	93.44	83.55	49.72	59.70	86.00	91.57	96.50
13	Wheat	50	155	94.52	99.74	99.55	92.77	96.59	100.00	99.42	99.48
14	Woods	50	1215	78.07	97.47	90.46	76.97	91.46	94.44	98.49	98.91
15	Building-GTD	50	1336	45.09	99.46	81.64	47.14	70.98	94.56	99.23	99.49
16	Stone-steel-towers	10	83	87.11	99.16	94.94	88.55	96.77	95.70	97.83	98.07
OA (%)				60.61	82.98	82.12	61.59	73.95	87.23	93.92	93.98
κ				0.5577	0.8074	0.7967	0.5682	0.7053	0.8538	0.9304	0.9312

Table 2. Number of Samples as well as Classification Accuracy of Various Methods Using INN on PU

No.	Class name	Train	Test	OSF/103	CAF/163	RLMR/29	LRCRP/30	MFC/19	MFONMF/16	SuperPCA/30	MLLE/15
1	Asphalt	10	6621	60.31	56.89	63.57	64.61	40.69	57.41	70.95	65.81
2	Meadows	10	18639	57.86	57.71	61.72	58.72	46.47	75.91	78.73	81.97
3	Gravel	10	2089	55.75	45.06	50.05	58.40	72.08	83.56	83.61	65.87
4	Tree	10	3054	86.64	95.07	85.33	89.05	88.97	87.27	66.41	92.17
5	Painted metal sheets	10	1335	98.93	98.08	99.47	99.22	98.66	99.63	92.04	99.84
6	Bare soil	10	5019	57.14	42.24	63.97	58.71	72.90	60.99	84.13	78.12
7	Bitumen	10	1320	87.87	66.85	83.20	84.11	63.46	69.85	95.45	80.06
8	Self-blocking bricks	10	3672	62.42	63.34	63.71	65.78	56.84	74.69	82.23	57.69
9	Shadows	10	937	99.80	99.74	99.97	99.59	78.46	91.76	99.90	95.04
OA (%)				63.64	60.82	66.25	65.35	56.75	73.28	79.22	77.65
κ				0.5490	0.5188	0.5792	0.5694	0.4799	0.6597	0.7337	0.7123

randomly and average the classification results over ten runs.

To validate the effectiveness of the proposed MLLE, we compare the classification performance with original spectral feature (OSF), the concatenation of all features (CAF), two state-of-the-art spectral DR techniques (a robust local manifold representation (RLMR) [8] and LRCRP [10]), and three state-of-the-art spectral-spatial DR methods (MFC [2], multiple-features ONMF (MFONMF) [11] and SuperPCA [12]).

3) Experimental results: The accuracy of each class, overall accuracy (OA), and kappa coefficient (κ) of different methods under the optimal parameters for IP are illustrated in Table 1 clearly, where the numbers in the first line are the final dimensions of corresponding DR methods. From this table, we arrive at some conclusions. Firstly, for the two spectral-based DR methods, the improvement of classification performance is limited without considering spatial information. And unfortunately, MFC framework cannot exploit the multiple features effectively on IP dataset, whose accuracy is significantly lower than MFONMF, SuperPCA and MLLE. Finally, the proposed MLLE outperforms the other methods, with highest OA and κ as well as highest accuracy of ten categories.

Also, the observations, which reveal the classification accuracy of

different methods using INN on PU dataset, are illustrated in Table 2. As we can see, SuperPCA performs best, and the proposed MLLE can achieve the optimized goal of using the fewest dimensions to yield the second best accuracy, followed by MFONMF, RLMR, LRCRP, OSF, CAF and MFC.

Fig. 1(a) shows how the objective function value of MLLE changes as the iteration number increases, which proves a fast convergence speed. We now observe α and δ of each feature with respect to r from Table 3. Obviously, when $r = -1$, the complementary nature of different features can be explored in an effective manner in ways that $r > 1$ cannot match. Additionally, we compare the OA of the proposed scheme (MLLE-Ridge) to those of MLLE with full samples (MLLE-Full) and linear regression (MLLE-Linear). The OA curves versus the dimension d on IP dataset are displayed in Fig. 1 (b). It is shown that the proposed scheme has obvious predominant in accuracy, which lies in the mitigation of the highly correlated features.

Conclusion: In this letter, we propose a new unsupervised DR framework for HSI based on MLLE. Not only does the framework apply LLE to preserve the locally linear structure of each single view, but it also learns a unified and sufficiently discriminative

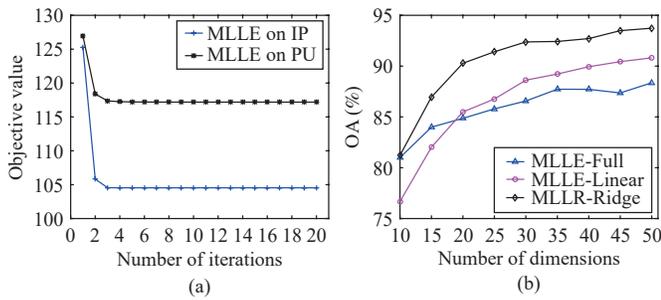


Fig. 1. Performance validation. (a) Convergence of MLE on two datasets; (b) Classification results of full samples MLE and two linear versions using INN on IP.

Table 3. Weights and Performance Comparison of MLE With Different r on PU ($d = 15$)

	$r = 10$		$r = 50$		$r = -1$	
	α	δ	α	δ	α	δ
Spectral	0.7830	≈ 1	0.4182	≈ 1	0.1457	0.5585
Texture	0.1075	2×10^{-9}	0.2904	1×10^{-8}	0.2690	0.3025
Shape	0.1095	2×10^{-9}	0.2914	1×10^{-8}	0.5853	0.1390
OA	62.13		60.66		77.65	

representation over multiple views. As an extension of MFC, the proposed MLE alleviates the multicollinearity in the out-of-sample problem. Experimental study on two benchmark datasets demonstrates the effectiveness of the proposed MLE.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (62073019).

References

- [1] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by CNN denoiser," *IEEE Trans. Neural Networks and Learning Systems*, vol. 32, no. 3, pp. 1124–1135, 2021.
- [2] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 879–893, 2012.
- [3] Z. Cai and W. Zhu, "Feature selection for multi-label classification using neighborhood preservation," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 1, pp. 320–330, 2018.
- [4] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [5] Y. Wang, Z. Zhang, and Y. Lin, "Multi-cluster feature selection based on isometric mapping," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 3, pp. 570–572, 2022.
- [6] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [7] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing Systems*, vol. 14, pp. 585–591, 2001.
- [8] D. Hong, N. Yokoya, and X. X. Zhu, "Learning a robust local manifold representation for hyperspectral dimensionality reduction," *IEEE J. Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 2960–2975, 2017.
- [9] G. Shi, H. Huang, and L. Wang, "Unsupervised dimensionality reduction for hyperspectral imagery via local geometric structure feature learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1425–1429, 2020.
- [10] X. Jiang, L. Xiong, Q. Yan, Y. Zhang, X. Liu, and Z. Cai, "Unsupervised dimensionality reduction for hyperspectral imagery via laplacian regularized collaborative representation projection," *IEEE Geoscience and Remote Sensing Letters*, p. 6007805, 2022.
- [11] J. Wen, J. E. Fowler, M. He, Y.-Q. Zhao, C. Deng, and V. Menon, "Orthogonal nonnegative matrix factorization combining multiple features for spectral-spatial dimensionality reduction of hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 4272–4286, 2016.
- [12] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4581–4593, 2018.
- [13] T. Xia, D. C. Tao, T. Mei, and Y. D. Zhang, "Multiview spectral embedding," *IEEE Trans. Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 6, pp. 1438–1446, 2010.
- [14] L. Zhang, L. P. Zhang, D. C. Tao, and X. Huang, "A modified stochastic neighbor embedding for multi-feature dimension reduction of remote sensing images," *ISPRS J. Photogrammetry and Remote Sensing*, vol. 83, pp. 30–39, 2013.
- [15] R. Bhatia, *Matrix Analysis*. Berlin, Germany: Springer Science & Business Media, 2013, vol. 169, pp. 1–347.
- [16] M. Haghghat, S. Zonouz, and M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," *Expert Systems With Applications*, vol. 42, no. 21, pp. 7905–7916, 2015.