

# Sparse Principal Component Analysis and Adaptive Multigraph Learning for Hyperspectral Band Selection

Wenxian Zhang<sup>1b</sup>, Member, IEEE, Aihong Yuan<sup>1b</sup>, Jinglei Tang<sup>1b</sup>, and Xuelong Li<sup>1b</sup>, Fellow, IEEE

**Abstract**—Band selection (BS) is an effective dimensionality reduction technique for hyperspectral images. Although many relevant methods have been proposed, they often only focus on the bandwise information and the correlation between the bands, and few of them pay attention to the manifold preservation in low-dimensional space, which may lead to the intrinsic structure of the data being damaged. In this article, we propose a novel method called sparse principal component analysis and adaptive multigraph learning (SPCA-AMGL) to address this issue. First, it applies SPCA to the BS task to learn the projection weight matrix, which utilizes the orthogonal constraint to remove redundant bands and uses the  $L_{2,1}$  norm to impose a sparse regularization on the weight matrix to ensure the selection of effective bands. Then, to select the bands with manifold preserving capability, AMGL is proposed to capture the local neighbor structure of data by combining the benefit of multiple graphs, which can not only adaptively learn the graph structure but also obtain the analytical solution of the multigraph coefficients. Finally, an alternate iterative algorithm is designed to optimize the proposed method. Abundant experiments on three hyperspectral datasets prove the reliability and superiority of the proposed method.

**Index Terms**—Adaptive multigraph learning (AMGL), local manifold preserving, sparse principal component analysis (SPCA), unsupervised band selection (BS).

## I. INTRODUCTION

WITH the rapid development of remote sensing technology, the current spectral imaging technology can image ground objects with an extremely high spectral resolution to obtain hyperspectral images (HSIs), which usually have hundreds of bands and can provide fine information reflecting the

characteristics of ground objects [1]. Therefore, it is widely used in the geological survey [2], vegetation detection [3], environmental monitoring [4], and many other fields. However, this also causes a problem that HSI has a huge amount of data compared with the traditional three-channel image, which brings great difficulties to the storage, transmission, and processing of images. Meanwhile, dense spectral sampling results in a high correlation between the adjacent bands, which makes the HSI contain many redundant bands. In addition, the existence of noise bands, irrelevant bands, and the curse of dimensionality phenomenon also increases the difficulty of HSI analysis. To solve the above issues, it is an effective means to apply dimensionality reduction technologies to HSI analysis for reducing the amount of data while preserving the discriminative spectral information [5].

The current dimensionality reduction technologies mainly include two types: feature extraction [6], [7] and feature selection [8], [9], [10] [also called band selection (BS) in HSI processing]. Feature extraction obtains low-dimensional data by combining or projecting the original features to generate new features. However, it changes the spectral values of the data, which results in the loss of the statistical property and physical meaning contained in the original bands, making it difficult for some scenes that need these original spectral values. Different from feature extraction, the idea of BS is to select a discriminative and representative band subset from the original band space, while it can well preserve the physical meaning of the bands [11]. Therefore, BS is more suitable for HSI dimensionality reduction in most cases.

From the perspective of label information usage, BS methods can be classified as supervised [12], [13], semisupervised [14], [15], and unsupervised methods [16], [17], [18]. In general, the label information can effectively promote the selection of high-quality bands and achieve better performance in supervised and semisupervised band methods. Nevertheless, it is very expensive work to obtain the label information of data, which hinders the application of these two methods in reality. On the contrary, unsupervised BS methods mine information from the data itself without using additional label information, which makes it of great application value. Therefore, unsupervised BS is the focus of this article.

Currently, the unsupervised BS methods mainly include ranking-based [19], clustering-based [20], and sparsity-based

Manuscript received 30 April 2023; revised 15 August 2023 and 27 October 2023; accepted 8 November 2023. Date of publication 21 November 2023; date of current version 14 December 2023. This work was supported in part by the Innovation Support Plan of Shaanxi Province under Grant 2021TD-31, and in part by the National Key Research and Development Program of China under Grant 2021YFD1600704. (Wenxian Zhang and Aihong Yuan are co-first authors.) (Corresponding author: Jinglei Tang.)

Wenxian Zhang, Aihong Yuan, and Jinglei Tang are with the College of Information Engineering, Northwest A&F University, YangLing 712100, China (e-mail: zwx\_cy@126.com; ahyuan@nwafu.edu.cn; tangjinglei@nwsuaf.edu.cn).

Xuelong Li is with the School of Artificial Intelligence, Optics and ElectroNics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China, and also with the Key Laboratory of Intelligent Interaction and Applications, Ministry of Industry and Information Technology, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: li@nwpu.edu.cn).

Code is available online at <https://github.com/ZWX0823/SPCA-AMGL>.

Digital Object Identifier 10.1109/JSTARS.2023.3335286

methods [21]. Specifically, ranking-based methods generally sort the bands under a certain criterion and then select the top-ranked bands. For example, maximum-variance principal component analysis [19] chooses bands according to their corresponding variance. However, the selected bands are usually highly corrected and may not be satisfactory because of the influence of noise bands. Clustering-based methods utilize clustering algorithms to divide the band space and select the most representative bands from each cluster. Ward's linkage and divergence information [22] and optimal clustering framework (OCF) [20] are two typical examples. The selected bands can effectively avoid redundancy, but the overall usefulness cannot be guaranteed, and these methods highly depend on the clustering algorithms. Sparsity-based methods try to learn a sparse representation of HSI to obtain the target subset. The typical methods include sparse representation-based band selection [21] and correntropy-based sparse spectral clustering [23]. Nevertheless, the sparse representation is sensitive to the quality of HSI data, which greatly influences the performance and stability of these methods. Recently, there has been a surge of interest in deep learning-based approaches for BS. This is primarily due to their remarkable ability to learn intricate features from data. In [24], a deep neural network-based method (BS-Nets) is introduced for BS, which transforms the task of BS into a spectral reconstruction task. In [25], a dual-GCN based on band attention and sparse constraint (BSD-GCN) is proposed for the BS. It uses the local and nonlocal correlations among spectral bands by constructing two types of graphs and aggregating the band information with the topological structure to select a better band subset. Despite the advancements in deep learning, traditional methods still hold significance, especially in scenarios with limited data or computational resources.

Although the above methods have achieved great success in HSI dimensionality reduction tasks, they only utilize the spectral information of HSIs while neglecting the spatial information, which can help to select bands with distinguishability. In [26], the spatial information of HSIs is preserved by adopting superpixel segmentation to segment their first principal component into diverse homogeneous regions. To promote BS, Shang et al. [27] utilized hypergraph to characterize the local structure of bands. By combining feature extraction and feature selection, the method in [28] learns a spectral-spatial feature representation through graph Laplacian. In recent years, some studies have shown that using both spectral and spatial information from HSIs simultaneously can help improve the classification accuracy of HSIs.

In spectral-spatial methods, there are three key aspects that need to be addressed. First, how to ensure that the selected bands have sufficient information and discriminability. For instance, while the original PCA can capture the most informative features, it is sensitive to outliers. Hence, the presence of noise and outliers in HSIs may significantly affect its performance. Second, how to fully utilize the spatial information in HSIs. Most methods rely on constructing similarity graphs to describe the geometric relationships between pixels or bands. However, these methods often employ a single graph construction approach, such as using the heat kernel graph. Different graph construction

methods can capture different structural information, so the graphs constructed by these methods may not comprehensively represent the spatial information in HSIs. Third, how to effectively integrate spectral and spatial information. Most methods adopt a two-step process, first constructing a similarity graph, and then using it for BS. However, in these two-step methods, the similarity graph relies on the raw data and is fixed, which fails to fully promote the interaction and collaboration between these two tasks (BS and graph construction).

To address the aforementioned issues, we propose a sparse principal component analysis and adaptive multigraph learning (SPCA-AMGL) method for hyperspectral BS in an unsupervised manner. This method assumes that there is a low-dimensional orthogonal subspace, and the original HSI can be mapped to this subspace through a projection matrix, which reveals the importance of each original band in constructing low-dimensional features. Its goal is to find such a projection matrix that can optimally measure the importance of each band. In the proposed method, an SPCA term is used to project the original high-dimensional data into a orthogonal low-dimensional subspace and ensure that the new features have high information content and low correlation. A local manifold preserving constraint utilizes the spatial information of pixel nearest neighbor structure to constrain the projection matrix, assigning greater weights to bands with manifold preserving ability, making them easier to select. In order to learn a better quality similarity graph, this method utilizes an AMGL strategy that combines the advantages of multigraph learning and considers the complementary information of different similarity graphs. It learns a more complete and robust similarity graph from multiple similarity graphs, and embeds this multigraph learning process into BS, achieving an adaptive learning approach. Then, an alternate iterative algorithm is given to optimize the SPCA-AMGL, which iteratively updates the projection matrix, the similarity graph, and the corresponding fusion weights.

In detail, the main contributions of this article are summarized as follows.

- 1) This article proposes a novel unsupervised method for hyperspectral BS. Its goal is to find an optimal projection matrix that enables the projected new features to have a large amount of information and low correlation, while maintaining the local geometric structure between pixels. Then, based on the obtained projection matrix, selecting a certain number of bands that contribute significantly to the construction of the low-dimensional subspace.
- 2) An SPCA term with  $L_{2,1}$  norm regularization is proposed to learn the projection matrix, which can effectively selects bands with high information content and low redundancy, and reduce the impact of outliers. The advantages of manifold learning are integrated into the model to help describe the nearest neighbor structure between pixels. To give a comprehensive description of the neighborhood graph between pixels, multiple graphs from different perspectives are utilized. This can overcome the inadequacy of using a single graph and help construct a more complete description of the interpixel structure.

- 3) In this article, we integrate the BS and multigraph learning into a unified framework, where they can promote each other and further improve their performance during optimization process. Furthermore, an alternate iterative algorithm is given to obtain the optimal solution of the proposed method.

## II. PRELIMINARY

A brief review of PCA, locality preserving projections (LPP), and AMGL, which are related to our proposed method, are presented in this section.

### A. Principal Component Analysis

Assume that  $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{b \times n}$  is an HSI matrix, and  $X$  has been centralized, where  $b$  is the number of bands and  $n$  is the number of pixels. From the perspective of PCA, the aim is to seek out an orthogonal projection matrix  $W \in \mathbb{R}^{b \times d}$  with  $d \ll b$ , making the reconstruction error minimized as follows:

$$\begin{aligned} \min_W \quad & \|X - WW^T X\|_F^2 \\ \text{s.t.} \quad & W^T W = I \end{aligned} \quad (1)$$

where  $W^T$  is the transposed matrix of  $W$  and  $I$  is an identity matrix of size  $d \times d$ .  $W^T$  is first used to project the  $X$  into a low-dimensional space, and then  $W$  is used to recover the  $X$ . After learning an optimal projection matrix  $W$ , the low-dimensional data can be denoted as  $Y = W^T X \in \mathbb{R}^{d \times n}$ , and its each column is a principal component [29]. As a projection weight matrix, the row vector  $W_i$  reveals the importance of the  $i$ th band in constructing low-dimensional space.

### B. Locality Preserving Projections

Manifold learning is an effective dimensionality reduction technology in machine learning, which is designed to investigate the intrinsic structure of data. Representative methods, include isometric mapping [30], locally linear embedding [31], and Laplacian eigenmaps (LE) [32]. Among them, LE constructs a Laplacian graph to preserve the local neighbor information of data. However, since these manifold learning methods are nonlinear projections, they cannot derive an explicit projection matrix to achieve the transformation of the out-of-samples [33]. To address this problem, LPP [34] is proposed based on LE, which is a linear approximation of the LE and can generate an explicit projection matrix to directly produce a low-dimensional representation of data. Given the data pairwise similarity matrix  $S \in \mathbb{R}^{n \times n}$  containing the edge weights on a graph with  $n$  nodes, the formula of LPP is expressed as follows:

$$\begin{aligned} \min \quad & \sum_{i=1}^n \sum_{j=1}^n \|W^T x_i - W^T x_j\|_2^2 S_{ij} \\ \text{s.t.} \quad & W^T X X^T W = I \end{aligned} \quad (2)$$

where  $S_{ij}$  is the  $(i, j)$ th element of  $S$ , and  $I$  is an identity matrix of size  $d \times d$ . The purpose of the constraint here is to eliminate the influence of an arbitrary scale factor.

It can be seen that if  $S_{ij}$  is large, then the value of  $\|W^T x_i - W^T x_j\|_2^2$  is likely to be small. Therefore, by minimizing the function in (2), the low-dimensional representation of data can preserve the local nearest neighbor relationships in the high-dimensional space [35].

### C. Adaptive Multiple Graph Learning

Graph representation-based methods have been proved to be effective for representing the structure of data. It has been widely used in the related methods of feature selection. For example, Lu et al. [35] used a heat kernel graph to describe the weights between adjacent data points for preserving the local manifold structure, and Yuan et al. [36] adopted an adaptive graph constraint to learn the relationships among data. Compared with these monotonous graph construction patterns, in [37], a more robust AMGL algorithm is proposed to learn an adaptive graph from multiple initial graphs. Given multiple initial graphs  $S^{(1)}, S^{(2)}, \dots, S^{(K)} \in \mathbb{R}^{n \times n}$  constructed by different metric methods for representing the similarity of pairwise data points. The aim is to simultaneously learn a consensus graph  $S \in \mathbb{R}^{n \times n}$  from these initial graphs during the feature selection process. For each initial graph  $S^{(k)}$ , it should satisfy the following constraints. 1)  $S_{ij}^{(k)} \geq 0$ ; and 2)  $\sum_{j=1}^n S_{ij}^{(k)} = 1$ . Through these constraints, each row (i.e.,  $S_i^{(k)}$ ) can be regarded as a clear probability distribution. Therefore, the Kullback–Leibler divergence [38] is used for consensus measuring, i.e., minimize  $\sum_k \alpha_k^2 \sum_i KL(S_i^{(k)}, S_i)$ , where  $\alpha_k$  is the weight of the  $k$ th initial graph. The function is defined as follows:

$$\begin{aligned} \min_{S, \alpha} \quad & \sum_{k=1}^K \alpha_k^2 \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(k)} \log \frac{S_{ij}^{(k)}}{S_{ij}} \\ \text{s.t.} \quad & \sum_{j=1}^n S_{ij} = 1, 0 \leq S_{ij} \leq 1; \sum_{k=1}^K \alpha_k = 1, \alpha_k \geq 0 \end{aligned} \quad (3)$$

where the first constraint ensures that each row of  $S$  is also a probability distribution and the second one guarantees that the sum of all graph weights is 1.

## III. PROPOSED MODEL

The details of SPCA-AMGL are introduced in this section, including the construction of the model and the corresponding optimization algorithm.

### A. Model Construction

1) *Learning projection matrix*: Given the problem defined in (1), we want to learn an optimal  $W$  that can project the original bands space to a best approximation of low-dimensional manifold [15]. The  $W_{ij}$  is the projection weight of the  $i$ th band to reconstruct the  $j$ th low-dimensional band. Therefore, the importance of the  $b$  bands can be measured according to  $\{\|W_1\|_2, \|W_2\|_2, \dots, \|W_b\|_2\}$ , from which the most important bands can be selected [39]. Further, to discard redundant and irrelevant bands, each low-dimensional band ought to be reconstructed from the most effective original bands [40]. To this end,

the  $L_{2,1}$  norm is integrated into PCA for guaranteeing the row sparse of  $W$ . This forms the formula of sparse PCA as follows:

$$\begin{aligned} \min_W \quad & \|X - WW^T X\|_F^2 + \lambda_1 \|W\|_{2,1} \\ \text{s.t.} \quad & W^T W = I \end{aligned} \quad (4)$$

where  $\lambda_1$  is a balance parameter to adjust the sparse regularization, and  $\|W\|_{2,1}$  is defined as

$$\|W\|_{2,1} = \sum_{i=1}^b \|W_i\|_2 = \sum_{i=1}^b \left( \sum_{j=1}^d W_{ij}^2 \right)^{1/2}. \quad (5)$$

2) *Preserving local manifold*: In fact, adjacent pixels are relevant in real images and are more likely to come from the same class [28]. Related studies have shown that the nearest neighbor relationship is important for HSI classification [41], [42]. Based on this fact, we want to make the bands that are capable of keeping the local manifold structure of HSI more likely to be selected. Specifically, if the pixels  $x_i$  and  $x_j$  in the original space are similar, then their corresponding pixels in the low-dimensional space  $W^T x_i$  and  $W^T x_j$  should also be similar [43]. Based on the aforementioned idea, a local manifold preserving constraint is introduced into our model to achieve this goal, which can be expressed as follows:

$$\min_W \quad \frac{\lambda_2}{2} \sum_{i=1}^n \sum_{j=1}^n \|W^T x_i - W^T x_j\|_2^2 S_{ij} \quad (6)$$

where  $\lambda_2$  is a balance parameter, where  $S_{ij}$  represents the similarity between  $x_i$  and  $x_j$ , and the similarity matrix is  $S$ .

According to the spectral graph theory and manifold learning theory, constructing a nearest neighbor graph is an excellent way to represent the local manifold structure of data. Hence, our method utilizes the  $k$ -nearest neighbor (KNN) graph to calculate the  $S$ , which is defined as

$$S_{ij} = \begin{cases} f(x_i, x_j), & \text{if } x_j \in N(x_i) \text{ or } x_i \in N(x_j) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where  $N(x_i)$  denotes the set of KNN of  $x_i$ , and  $f(x_i, x_j)$  is a function for calculating the similarity between  $x_i$  and  $x_j$ .

3) *Learning similarity graph using AMGL*: The quality of graph will greatly affect the effect of BS. In the previous work, most of the constructed graphs are fixed, because they are calculated directly according to the given original data. The graphs obtained by this construction way are sensitive to redundant bands and outliers to a great extent. To solve this problem, we need to reduce the dependence of the model on the initial graphs. An effective method is to learn an adaptive graph in the process of BS [44]. Besides, there are many ways to build graphs according to different metric methods. Graphs with different definitions can depict a variety of structure information, and in most cases, these graphs can complement each other to form a more complete description of the data structure [37]. Therefore, an AMGL strategy is utilized in this article to learn an adaptive graph from multiple initial graphs that are constructed in various ways. It combines the advantages of multiple graphs and uses the adaptive learning method to reduce the dependence of the

model on the initial graphs. Its formula is defined as follows:

$$\begin{aligned} \min_{S, \alpha} \quad & \lambda_3 \sum_{k=1}^K \alpha_k^2 \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(k)} \log \frac{S_{ij}^{(k)}}{S_{ij}} \\ \text{s.t.} \quad & \sum_{j=1}^n S_{ij} = 1, 0 \leq S_{ij} \leq 1; \quad \sum_{k=1}^K \alpha_k = 1, \alpha_k \geq 0 \end{aligned} \quad (8)$$

where  $\lambda_3$  is a regularization parameter, and others have been introduced before.

Note that, the graphs constructed in our method are all self-connected, i.e.,  $S_{ii}^{(k)}$  has the largest value in the  $i$ th row. Considering that the scales of graphs constructed in different ways may be different, we need to normalize all graphs first. Through the transformation of  $S_{ij}^{(k)} / \sum_{j=1}^n S_{ij}^{(k)} \rightarrow S_{ij}^{(k)}$ , each row vector of  $S^{(k)}$  satisfies  $\sum_{j=1}^n S_{ij}^{(k)} = 1$ . After the normalization, the elements in each row (i.e.,  $S_i^{(k)}$ ) have a probability distribution. The larger the  $S_{ij}^{(k)}$ , the greater the similarity between the  $i$ th pixel and the  $j$ th pixel in the  $k$ th graph.

More importantly, the multigraph patterns are flexible, which can be constructed through multiple similarity metrics of the same kind but with different parameter values or multiple similarity metrics of different kinds. For example, in the AMGL model, multiple graphs based on the Gaussian kernel function can be constructed, which have different values of the scale parameter  $\sigma$ .

Finally, by integrating (4), (6), and (8), the objective function of SPCA-AMGL is obtained as follows:

$$\begin{aligned} \min_{W, S, \alpha} \quad & \|X - WW^T X\|_F^2 + \lambda_1 \|W\|_{2,1} \\ & + \frac{\lambda_2}{2} \sum_{i=1}^n \sum_{j=1}^n \|W^T x_i - W^T x_j\|_2^2 S_{ij} \\ & + \lambda_3 \sum_{k=1}^K \alpha_k^2 \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(k)} \log \frac{S_{ij}^{(k)}}{S_{ij}} \\ \text{s.t.} \quad & W^T W = I \\ & \sum_{j=1}^n S_{ij} = 1, 0 \leq S_{ij} \leq 1; \quad \sum_{k=1}^K \alpha_k = 1, \alpha_k \geq 0. \end{aligned} \quad (9)$$

## B. Model Optimization

As we can see from (9), there are three variables that need to be optimized:  $W$ ,  $S$ , and  $\alpha$ . This article uses an alternate iterative update optimization algorithm to optimize these variables. Its main idea is to optimize one variable while fixing the other variables.

1) *Optimize  $W$  by fixing  $S$  and  $\alpha$* : When fixing the  $S$  and  $\alpha$ , we can get the following formula:

$$\begin{aligned}
\min_W \quad & \|X - WW^T X\|_F^2 + \lambda_1 \|W\|_{2,1} \\
& + \frac{\lambda_2}{2} \sum_{i=1}^n \sum_{j=1}^n \|W^T x_i - W^T x_j\|_2^2 S_{ij} \\
\text{s.t.} \quad & W^T W = I.
\end{aligned} \tag{10}$$

Here, we need to make some transformations to (10) for facilitating its optimization. Specifically, (10) is equivalent to the following trace minimization problem:

$$\begin{aligned}
\min_W \quad & -\text{Tr}(W^T X X^T W) + \lambda_1 \text{Tr}(W^T Q W) \\
& + \lambda_2 \text{Tr}(W^T X L X^T W) \\
\text{s.t.} \quad & W^T W = I
\end{aligned} \tag{11}$$

where  $Q \in \mathbb{R}^{b \times b}$  is a diagonal matrix, and its diagonal element  $Q_{ii}$  is defined as

$$Q_{ii} = \frac{1}{\sqrt{\sum_{j=1}^d W_{ij}^2 + \theta}}, \quad (i = 1, 2, \dots, b). \tag{12}$$

The  $\theta$  is a positive constant with a sufficiently small value to ensure that the denominator makes sense during the actual coding process. And  $L \in \mathbb{R}^{n \times n}$  is the Laplacian matrix, which is calculated by  $L = D - (S + S^T)/2$ . The  $D$  is a degree matrix and its diagonal element  $D_{ii} = \sum_{j=1}^n (S_{ij} + S_{ji})/2$ .

Noting  $A = -X X^T + \lambda_1 Q + \lambda_2 X L X^T$ , the (11) can be rewritten as

$$\min_W \text{Tr}(W^T A W), \quad \text{s.t. } W^T W = I. \tag{13}$$

The eigenvectors of the first  $d$  minimum eigenvalues of  $A$  constitutes the optimal solution of  $W$ . Here, an orthogonal iteration-based method is proposed to solve the problem (13), and the algorithm is presented in Algorithm 1.

2) *Optimize  $S$  by fixing  $W$  and  $\alpha$ :* When the  $W$  and  $\alpha$  are fixed, the (9) is equivalent to the following formula:

$$\begin{aligned}
\min_S \quad & \frac{\lambda_2}{2} \sum_{i=1}^n \sum_{j=1}^n \|W^T x_i - W^T x_j\|_2^2 S_{ij} \\
& - \lambda_3 \sum_{k=1}^K \alpha_k^2 \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(k)} \log S_{ij} \\
\text{s.t.} \quad & \sum_{j=1}^n S_{ij} = 1, 0 \leq S_{ij} \leq 1.
\end{aligned} \tag{14}$$

Since each row of  $S$  can be analyzed independently, here we take the  $i$ th row as an example. We can see that if  $S_i$  satisfies  $\sum_{j=1}^n S_{ij} = 1$  and  $S_{ij} \geq 0$ , then  $S_{ij}$  must also satisfy  $S_{ij} \leq 1$ . So we can remove the constraint  $S_{ij} \leq 1$  safely. Using the Lagrange multiplier method, we can get the following Lagrange function:

$$\begin{aligned}
\mathcal{L} = \quad & \frac{\lambda_2}{2} \sum_{j=1}^n B_{ij} S_{ij} - \lambda_3 \sum_{j=1}^n C_{ij} \log S_{ij} \\
& + \lambda \left( \sum_{j=1}^n S_{ij} - 1 \right) - \sum_{j=1}^n \mu_j S_{ij}
\end{aligned} \tag{15}$$

---

**Algorithm 1:** Algorithm for Solving (13).

---

**Input:** The HSI matrix  $X \in \mathbb{R}^{b \times n}$ , the initial projection matrix  $W \in \mathbb{R}^{b \times d}$ , the similarity graph matrix  $S \in \mathbb{R}^{n \times n}$ , and the parameters  $\lambda_1, \lambda_2$ .

**Output:** The projection matrix  $W \in \mathbb{R}^{b \times d}$ .

1: Calculate  $Q$  by

$$Q_{ii} = \frac{1}{\sqrt{\sum_{j=1}^d W_{ij}^2 + \theta}}.$$

2: Calculate  $L = D - (S + S^T)/2$ .

3: Calculate  $A = -X X^T + \lambda_1 Q + \lambda_2 X L X^T$ .

4: Calculate  $\lambda_A$  that is the greatest eigenvalue of  $A$  calculated by power iteration method [45].

5: Calculate  $\tilde{A} = \lambda_A I - A$ .

6: **repeat**

7: Calculate  $M = \tilde{A} W$ .

8: Calculate  $U S V^T = M$  via the compact SVD method where  $U \in \mathbb{R}^{b \times d}$ ,  $S \in \mathbb{R}^{d \times d}$ , and  $V \in \mathbb{R}^{d \times d}$ .

9: Update  $W = U V^T$ .

10: **until** convergence.

---

where  $\lambda$  and  $\mu_j$  ( $j = 1, 2, \dots, n$ ) are Lagrange multipliers.  $B_{ij} = \|W^T x_i - W^T x_j\|_2^2$  and  $C_{ij} = \sum_{k=1}^K \alpha_k^2 S_{ij}^{(k)}$ .

Since problem (14) is convex and has a lower boundary, a solution that satisfies the Karush–Kuhn–Tucker (KKT) conditions is the global optimal solution [37]. Therefore, we can get the following KKT conditions:

$$\begin{cases} \frac{\lambda_2}{2} B_{ij} - \lambda_3 \frac{C_{ij}}{S_{ij}} + \lambda - \mu_j = 0 \\ \sum_{j=1}^n S_{ij} = 1 \\ S_{ij} \geq 0 \\ \mu_j S_{ij} = 0 \\ \mu_j \geq 0. \end{cases} \tag{16}$$

For all  $j$  that satisfy  $C_{ij} \neq 0$ , according to (16), we can get  $S_{ij} = \frac{\lambda_3 C_{ij}}{\frac{\lambda_2}{2} B_{ij} + \lambda - \mu_j}$  and  $S_{ij} \neq 0$ , so  $\mu_j = 0$ . Then,  $S_{ij} = \frac{\lambda_3 C_{ij}}{\frac{\lambda_2}{2} B_{ij} + \lambda}$ . For all  $j$  that satisfy  $C_{ij} = 0$ , we can get  $\frac{\lambda_2}{2} B_{ij} + \lambda - \mu_j = 0$  and  $\mu_j = \frac{\lambda_2}{2} B_{ij} + \lambda \geq 0$ .

Obviously, the smallest value in  $B_i$  is  $B_{ii}$ , which is equal to 0, and the corresponding  $C_{ii} \neq 0$ , so  $\mu_i = 0$ . Therefore, we can first calculate the value of  $\lambda$  by solving  $\sum_{j: C_{ij} \neq 0} S_{ij} = \sum_{j: C_{ij} \neq 0} \frac{\lambda_3 C_{ij}}{\frac{\lambda_2}{2} B_{ij} + \lambda} = 1$  in the range  $(-\frac{\lambda_2}{2} B_{ii}, +\infty)$ , i.e.,  $(0, +\infty)$ . Here, we define a function  $f(\lambda) = \sum_{j: C_{ij} \neq 0} \frac{\lambda_3 C_{ij}}{\frac{\lambda_2}{2} B_{ij} + \lambda}$ . It can be seen that  $\lim_{\lambda \rightarrow 0^+} f(\lambda) \rightarrow +\infty$  and  $\lim_{\lambda \rightarrow +\infty} f(\lambda) = 0$ . Since  $f(\lambda)$  decreases monotonically in  $(0, +\infty)$ , the equation  $f(\lambda) = 1$  must have a unique solution in  $(0, +\infty)$ . After solving the value of  $\lambda$ , we set  $S_{ij} = \frac{\lambda_3 C_{ij}}{\frac{\lambda_2}{2} B_{ij} + \lambda}$  and  $\mu_j = 0$  for all  $j$  that make  $C_{ij} \neq 0$ . And for all other  $j$ , we can set  $S_{ij} = 0$  and  $\mu_j = \frac{\lambda_2}{2} B_{ij} + \lambda$ . The detail algorithm for solving (16) is presented in Algorithm 2.

**Algorithm 2:** Algorithm for Solving (16).

**Input:** The HSI matrix  $X \in \mathbb{R}^{b \times n}$ , the projection matrix  $W \in \mathbb{R}^{b \times d}$ , the initial similarity matrices  $S^{(1)}, S^{(2)}, \dots, S^{(K)} \in \mathbb{R}^{n \times n}$ , and the parameters  $\lambda_2, \lambda_3$ , and  $\alpha$ .

**Output:** Similarity graph matrix  $S \in \mathbb{R}^{n \times n}$ .

- 1: Calculate  $B$  and  $C$  according to their definition.
- 2: **for**  $i = 1$  to  $n$  **do**
- 3: Calculate  $\lambda$  via solving  $f(\lambda) = \sum_{j: C_{ij} \neq 0} \frac{\lambda_3 C_{ij}}{\frac{\lambda_2}{2} B_{ij} + \lambda} = 1$ .
- 4: **for**  $j = 1$  to  $n$  **do**
- 5: **if**  $C_{ij} \neq 0$  **then**
- 6: Set  $S_{ij} = \frac{\lambda_3 C_{ij}}{\frac{\lambda_2}{2} B_{ij} + \lambda}$ .
- 7: **else**
- 8: Set  $S_{ij} = 0$ .
- 9: **end if**
- 10: **end for**
- 11: **end for**

3) *Optimize  $\alpha$  by fixing  $W$  and  $S$ :* When the  $W$  and  $S$  are fixed, the problem (9) can be rewritten as

$$\min_{\alpha} \sum_{k=1}^K \alpha_k^2 c_k, \quad \text{s.t.} \quad \sum_{k=1}^K \alpha_k = 1, \quad \alpha_k \geq 0 \quad (17)$$

where  $c_k = \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(k)} \log \frac{S_{ij}^{(k)}}{S_{ij}}$ .

According to the Cauchy–Schwarz Inequality, we can get the solution of problem (17) as follows:

$$\alpha_k = \frac{c_k^{-1}}{\sum_{i=1}^K c_i^{-1}}. \quad (18)$$

It can be seen from the (18) that  $\alpha_k$  can update adaptively during the optimization process. Moreover,  $\alpha_k$  and  $c_k$  show a negative correlation, i.e., the smaller  $c_k$  is, the larger  $\alpha_k$  is. According to the definition of  $c_k$ , it represents the divergence between  $S^{(k)}$  and  $S$ . Thus, the small  $c_k$  indicates that  $S^{(k)}$  is closer to  $S$ , that is,  $S^{(k)}$  has a greater quality [37]. In this respect, the learned  $\alpha_k$  does play the role of weight.

We iteratively optimize  $W$ ,  $S$ , and  $\alpha$  until they converge. After obtaining the optimal  $W$ , we sort all bands according to  $\|W_i\|_2$  in descending order and select the top  $d$  ranked bands to form the target band subset. The whole BS process is summarized in Algorithm 3.

### C. Computational Complexity Analysis

Computational complexity is a critical metric for evaluating an algorithm. In this section, the computational complexity of SPCA-AMGL is analyzed. According to Algorithm 3, the overall optimization process involves the update of two matrices and one vector (in three steps). In each iteration, updating  $W$  needs  $bd + 2nb^2 + n^2b + b^2 + b^2d + d^3 + bd^2$  flmlt (a floating-point multiplication), updating  $S$  needs  $bdn^2 + Kn^2 + n^2$  flmlt, updating  $\alpha$  needs  $Kn^2$  flmlt. In most cases,  $K < d$ ,  $d < b$ ,

**Algorithm 3:** Band Selection via SPCA-AMGL.

**Input:** The HSI matrix  $X \in \mathbb{R}^{b \times n}$ , the initial similarity matrices  $S^{(1)}, S^{(2)}, \dots, S^{(K)} \in \mathbb{R}^{n \times n}$ , the parameters  $\lambda_1, \lambda_2, \lambda_3$ , and the number of selected bands  $d$ .

**Output:**  $d$  selected bands.

- 1: Initialize a random projection matrix  $W \in \mathbb{R}^{b \times d}$  that satisfies  $W^T W = I$ .
- 2: Initialize  $S = \sum_{k=1}^K S^{(k)} / K$  and  $\alpha_k = 1 / K$ .
- 3: **repeat**
- 4: Update  $W$  via Alg. 1.
- 5: Update  $S$  via Alg. 2.
- 6: Update  $\alpha_k$  via Eq. (18).
- 7: **until**  $W$  convergence.
- 8: Sort all bands according to  $\|W_i\|_2$  in descending order and select the top  $d$  ranked bands.

and  $b < n$ . Thus, the computational complexity of the proposed method is  $O(nb^2 + n^2b)$ .

## IV. EXPERIMENTS

This section shows detailed experiments to verify the effectiveness of SPCA-AMGL.

### A. Experimental Setup

1) *HSI datasets:* This article uses three popular public HSI datasets for experiment and analysis, which are Indian Pines, Pavia University, and WHU-Hi-HanChuan, respectively.

- a) The Indian Pines dataset was developed in collaboration with Purdue University, West Lafayette, IN, USA and NASA, Washington, DC, USA, which was collected by AVIRIS sensor over the Indian Pines test site in Indiana, USA, in 1992. The original size of this dataset is  $145 \times 145 \times 224$ , of which  $145 \times 145$  represents the pixel size, 224 represents the number of bands. The entire image collection scene is relatively rich, including farmland, trees, roads, railways, and houses. In the experiment, this article focuses on the information of 16 types of ground truth and excludes 24 invalid bands: [104–108], [150–163], 220.
- b) The Pavia University dataset was collected and provided by the remote sensing laboratory of Pavia University in Italy, which was acquired by the ROSIS sensor over Pavia, Northern Italy. It contains  $610 \times 610$  pixels and 103 consecutive spectral bands. The dataset covers various types of land objects, including different types of vegetation, buildings, roads, etc. In this article, we focus on the information of nine types of ground truth.
- c) The WHU-Hi-HanChuan dataset was collected in Hanchuan City, Hubei Province, China, in June 2016, with a 17-mm focal length Headwall Nano-Hyperspec imaging sensor equipped on a Leica Aibot X6 UAV V1 platform. The study area is a rural–urban fringe zone with buildings, water, and cultivated land, which contains seven crop species: strawberry, cowpea, soybean, sorghum, water spinach, watermelon, and greens. The size of this dataset

TABLE I  
CLASSIFICATION PERFORMANCE OF DIFFERENT BS METHODS USING 50 BANDS ON INDIAN PINES DATASET

No.	OCF	ISSC	ONR	ASPS	FNGBS	MGSR	SPCA-AMGL	All bands
1	68.54±6.59	58.78±11.63	71.22±10.56	66.10±12.93	64.63± 13.85	69.02 ±15.96	<b>77.15±6.41</b>	56.59 ±11.64
2	73.37±2.35	79.03±1.62	79.53±1.83	<b>80.13±2.41</b>	77.84±2.63	67.07±2.67	75.85±2.17	79.25±1.54
3	62.95±2.11	70.18±2.46	68.27±3.62	70.95±4.07	<b>71.50±2.55</b>	57.86±2.28	62.17±2.53	71.41± 2.23
4	50.47±7.63	51.08±7.08	62.77±5.48	68.97±7.09	66.20±6.28	51.46±6.21	<b>69.75±7.21</b>	50.70±4.42
5	87.63±3.05	90.58±1.96	88.39 ±3.20	90.30±2.36	<b>90.62± 2.48</b>	86.37± 3.87	89.22±3.17	90.90± 1.17
6	92.62±2.93	95.20±1.55	94.44± 3.22	95.59±1.54	94.46± 1.71	91.51± 2.81	<b>95.66±1.50</b>	95.46± 1.63
7	70.40± 22.43	74.67±10.15	82.40± 13.65	72.00±17.71	<b>82.40± 3.20</b>	68.00± 21.61	74.40±11.59	72.00± 13.15
8	95.07± 2.36	96.39±1.56	<b>98.23± 1.24</b>	97.51±2.09	97.35± 0.94	96.16± 1.96	95.91±2.04	96.07± 1.78
9	40.56± 15.92	41.67±16.74	32.78± 21.72	61.11±15.32	58.89± 22.8	21.67± 10.08	<b>68.15±14.66</b>	38.89± 20.34
10	67.52± 4.16	71.80±2.56	<b>74.53± 3.80</b>	72.89±3.13	71.42± 3.58	62.33± 2.97	64.25±2.63	69.95± 3.16
11	78.21± 1.37	78.01±1.33	77.23± 2.41	77.06±1.59	75.79± 1.29	68.89± 0.95	<b>79.03±1.94</b>	78.83± 1.80
12	72.51± 4.57	68.18±3.66	71.63± 3.09	<b>76.76±1.90</b>	76.05± 2.06	46.07± 5.32	74.67±3.93	66.5± 5.44
13	93.68± 2.45	96.41±2.00	94.27± 3.51	97.19±1.82	97.08± 1.73	92.92± 3.91	<b>98.54±1.09</b>	95.51± 1.41
14	93.81± 1.82	91.84±1.91	90.70± 1.84	91.87±1.91	91.54± 2.31	91.94± 2.52	<b>94.28±1.63</b>	92.43± 1.16
15	52.48± 6.31	56.13±4.40	54.87± 4.22	<b>58.33±5.57</b>	55.62± 4.16	42.88± 2.46	54.41±4.26	53.60± 4.73
16	89.88± 4.10	90.04±3.66	85.48± 5.95	88.33±5.10	86.07 ±5.53	83.10± 7.10	<b>90.99±4.20</b>	90.36± 3.82
OA (%)	77.71± 0.51	79.69± 0.67	79.98± 0.68	80.39± 0.51	80.18± 0.71	71.77± 0.72	<b>81.05± 0.66</b>	79.58± 0.77
AA (%)	74.36± 5.63	75.62± 0.31	76.67± 0.58	<b>79.93± 0.59</b>	78.59± 0.59	68.58± 0.42	79.15± 0.63	74.90± 0.46
Kappa	0.741± 0.025	0.760± 0.018	0.774± 0.013	0.782± 0.015	0.775± 0.012	0.682± 0.013	<b>0.800± 0.012</b>	0.762± 0.016

The best results are highlighted in bold.

TABLE II  
CLASSIFICATION PERFORMANCE OF DIFFERENT BS METHODS USING 50 BANDS ON PAVIA UNIVERSITY DATASET

No.	OCF	ISSC	ONR	ASPS	FNGBS	MGSR	SPCA-AMGL	All bands
1	89.48 0.95	89.96± 0.95	93.19± 0.58	92.26± 0.66	92.26± 0.59	90.97± 0.87	<b>93.26±0.67</b>	89.57± 0.82
2	95.70± 0.41	95.94± 0.34	97.65± 0.21	97.18± 0.22	97.37± 0.28	96.54± 0.31	<b>97.91±0.38</b>	95.77± 0.19
3	71.29± 1.56	72.52± 1.69	77.74± 1.98	77.46± 1.7	<b>79.95± 1.79</b>	74.57± 1.99	76.95±1.38	71.73± 2.89
4	93.10± 0.87	93.31± 1.02	94.90± 0.86	93.97± 0.76	94.93± 0.86	93.01± 0.64	<b>95.99±1.07</b>	93.15± 1.20
5	99.60± 0.29	99.63± 0.26	99.57± 0.20	99.30± 0.28	99.33± 0.20	99.46± 0.37	<b>99.87±0.21</b>	99.63± 0.26
6	77.54± 0.75	79.29± 1.12	<b>90.20± 0.70</b>	89.09± 0.92	89.89± 0.98	83.66± 0.86	86.70±1.19	78.97± 1.50
7	81.00± 1.07	80.20± 1.60	83.15± 1.37	84.41± 1.87	<b>85.26± 2.14</b>	78.86±2.20	84.98±2.42	81.39 ±1.80
8	80.93± 1.90	80.35± 1.62	86.60± 1.45	<b>87.00± 1.16</b>	86.89± 1.24	84.22± 1.10	86.36±1.19	80.94± 1.30
9	97.89± 0.46	97.95± 0.82	98.95± 0.06	98.85± 0.14	98.94± 0.11	98.84± 0.17	<b>98.99±0.83</b>	98.02± 0.86
OA (%)	89.75± 0.21	89.21± 0.56	91.56± 0.42	92.28± 0.55	92.56± 0.64	91.42± 0.42	<b>93.94± 0.54</b>	90.10± 0.22
AA (%)	87.39± 0.32	87.68± 0.45	91.33± 0.62	91.06± 0.34	<b>91.65± 0.55</b>	88.90± 0.43	91.22± 0.49	87.69± 0.65
Kappa	0.872± 0.014	0.875± 0.023	<b>0.891± 0.037</b>	0.876± 0.042	0.882± 0.021	0.864± 0.068	0.878± 0.031	0.885± 0.018

The best results are highlighted in bold.

is  $1217 \times 303$  pixels. There are 274 bands from 400 to 1000 nm, and the spatial resolution of the UAV-borne hyperspectral imagery is about 0.109 m. Its image contains 16 types of ground truth [46], [47].

2) *Comparison methods*: To more intuitively evaluate the performance of SPCA-AMGL, six representative BS methods are used for comparison: 1) OCF [20], 2) ISSC [48], 3) ONR [49], 4) ASPS [50], 5) FNGBS [51], and 6) MGSR [52]. In addition, the experimental results obtained by using all bands are used as the baselines.

3) *Classifiers*: To evaluate the quality of the bands selected by the above methods, we adopt two widely used hyperspectral classifiers for classification experiments, which are support vector machine (SVM), and KNN. For each classifier, 10% of samples in each category are selected as training set, and the remaining 90% are used as test set. To make the experimental results more credible, the average of ten experimental data is taken as the experimental result.

The kernel function in SVM is set as ‘‘RBF’’, and the parameters  $C$  and  $\gamma$  are set as 10 000 and  $1/(\text{number of bands} \times \text{variance of data})$ , respectively. Regarding KNN, the parameter  $k$  is initialized to 5.

4) *Accuracy measures*: In this article, three popular quantitative indices, i.e., overall accuracy (OA), average accuracy

(AA), and kappa coefficient (Kappa), are selected to calculate classification results.

5) *Parameter setting*: For the SPCA-AMGL, there are four main parameters that need to be adjusted:  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ , and  $k$  in the initial KNN graphs. We use the grid search strategy to determine the values of these parameters. First, we make a rough adjustment in the range of  $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3\}$ , and then adjust the values of parameters more finely in a small step. Finally, in this article, we set the parameters  $\lambda_1 = 1$ ,  $\lambda_2 = 300$ , and  $\lambda_3 = 0.3$ , respectively, and the value of  $k$  is set to 6. The specific contents of parameter analysis will be described in Section IV-D.

For the comparison methods, all parameters are set according to their references or determined via experiment. For example, the  $\beta$  in ISSC is set as 0.0001. For MGSR, the regularized parameter  $\lambda$  and noise probability  $p$  are set as 0.5 and 1 on the Indian Pines dataset, 0.6 and 0.1 on the Pavia University dataset, and 0.1 and 0.01 on the WHU-Hi-HanChuan dataset, respectively.

## B. Comparison of Classification Performance

In this section, we assess the classification accuracy of various band subsets obtained through distinct BS methods. The

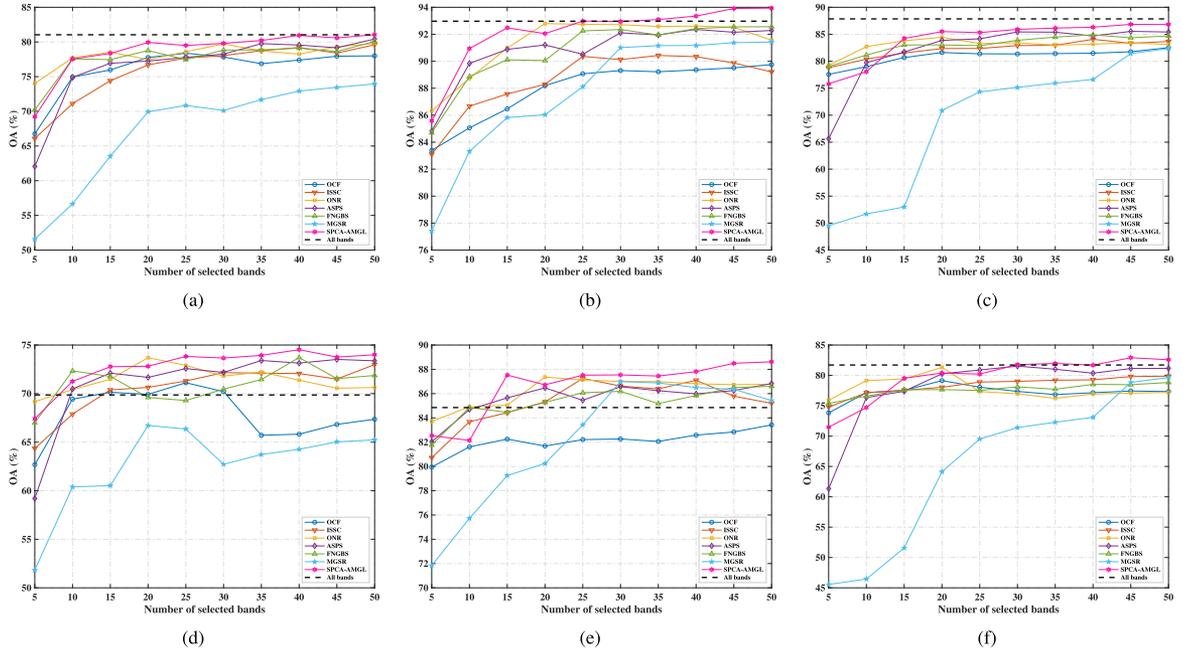


Fig. 1. Classification performance of different methods under varying band subset sizes. (a)–(c) Classification results of using SVM on Indian Pine, Pavia University, and WHU-Hi-HanChuan datasets, respectively. (d)–(f) Classification results of using KNN on Indian Pine, Pavia University, and WHU-Hi-HanChuan datasets, respectively.

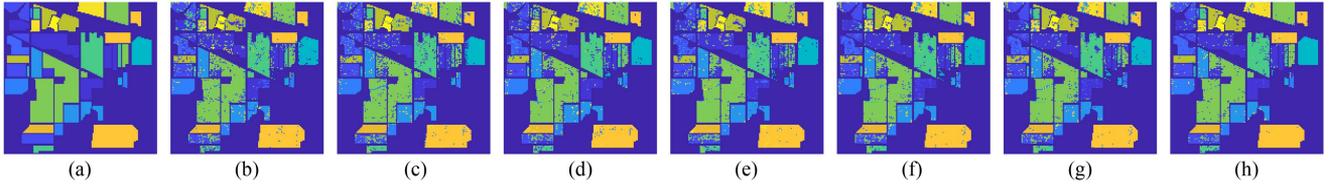


Fig. 2. Classification maps of Indian Pine dataset obtained by different methods. (a) Ground truth. (b) OCF. (c) ISSC. (d) ONR. (e) ASPS. (f) FNGBS. (g) MGSR. (h) SPCA-AMGL.

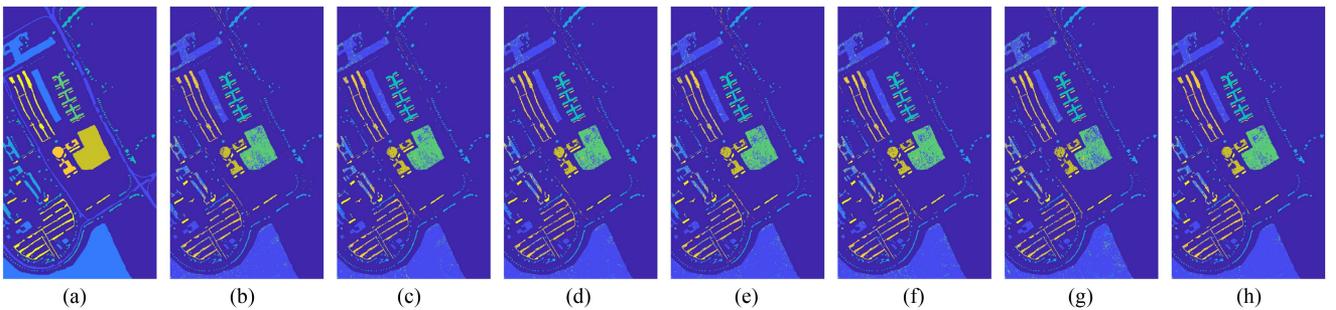


Fig. 3. Classification maps of Pavia University dataset obtained by different methods. (a) Ground truth. (b) OCF. (c) ISSC. (d) ONR. (e) ASPS. (f) FNGBS. (g) MGSR. (h) SPCA-AMGL.

OA achieved by these methods is presented in Fig. 1(a)–(f), where the band subset sizes range from 5 to 50 with a step of 5. In addition, we provide a comprehensive performance evaluation in Tables I–III. For Indian Pine, Pavia University, and WHU-Hi-HanChuan datasets, the band subset sizes are set to 50, respectively. This table includes class accuracy, OA, AA, and Kappa. Moreover, we conduct a visual analysis of the classification maps generated by the SVM classifier, utilizing different BS

techniques on the experimental datasets. Figs. 2–4 display these results. Notably, the proposed SPCA-AMGL method demonstrates a consistent superiority over the alternative BS methods.

First, as shown in Fig. 1, SPCA-AMGL achieves the best OA across all datasets, both in the case of SVM and KNN classifiers. Notably, even with a relatively small band subset size, such as the number of bands is 15, SPCA-AMGL still has a better OA compared with other BS methods. This indicates

TABLE III  
CLASSIFICATION PERFORMANCE OF DIFFERENT BS METHODS USING 50 BANDS ON WHU-HI-HANCHUAN DATASET

No.	OCF	ISSC	ONR	ASPS	FNGBS	MGSR	SPCA-AMGL	All bands
1	92.75± 0.88	93.15± 0.64	92.47± 0.58	<b>94.04± 0.83</b>	92.65± 0.25	91.76± 0.61	93.78± 0.27	92.95± 0.32
2	73.30± 0.74	<b>77.75± 0.37</b>	76.07± 0.64	76.62± 0.66	76.80± 0.34	69.03± 1.24	76.03± 0.32	86.23± 0.28
3	73.84± 0.76	<b>77.19± 0.55</b>	72.00± 0.94	76.03± 0.75	73.96± 0.64	69.30± 0.64	76.80± 0.55	81.33± 0.49
4	91.80± 0.62	92.59± 0.68	88.72± 0.66	92.11± 0.51	90.84± 0.45	79.70± 0.57	<b>93.19± 0.34</b>	95.16± 0.61
5	39.16± 0.75	38.14± 0.73	35.27± 0.83	43.79± 0.85	<b>45.09± 0.52</b>	31.94± 0.64	43.15± 0.58	46.29± 0.86
6	37.67± 0.64	40.19± 0.99	38.77± 2.10	39.58± 0.45	39.11± 0.77	23.89± 2.32	<b>40.81± 0.42</b>	54.36± 0.56
7	72.20± 0.58	74.72± 0.84	69.69± 0.85	76.45± 0.63	75.38± 0.56	68.54± 1.44	<b>77.28± 0.31</b>	75.51± 0.45
8	67.71± 0.61	73.38± 0.62	66.37± 0.64	73.39± 0.84	71.05± 0.84	65.83± 0.85	<b>73.64± 0.77</b>	76.87± 0.61
9	61.24± 0.43	67.15± 0.77	60.43± 0.34	65.20± 0.73	63.90± 0.64	55.05± 1.10	<b>69.46± 0.65</b>	77.72± 0.34
10	91.49± 0.55	91.69± 0.89	89.52± 0.45	91.31± 0.74	92.03± 0.77	93.11± 0.58	<b>93.27± 0.31</b>	94.09± 0.75
11	87.65± 0.76	88.87± 0.62	87.26± 0.61	89.67± 0.91	89.27± 0.59	88.54± 0.75	<b>89.84± 0.42</b>	89.18± 0.58
12	46.81± 0.78	46.45± 0.79	42.43± 0.52	<b>47.14± 0.88</b>	46.30± 0.85	45.81± 0.46	44.61± 0.43	57.80± 0.42
13	54.16± 0.69	<b>57.53± 0.88</b>	50.51± 0.57	54.76± 0.64	55.31± 0.81	51.17± 0.58	56.57± 0.25	62.56± 0.65
14	78.91± 0.87	79.98± 0.65	76.62± 0.68	81.13± 0.54	77.74± 0.43	77.46± 0.93	<b>81.87± 0.16</b>	82.51± 0.55
15	70.74± 0.61	70.25± 0.83	64.09± 0.73	<b>78.37± 0.23</b>	77.88± 0.61	74.65± 0.85	74.17± 0.52	75.14± 0.42
16	98.22± 0.83	98.34± 0.33	98.29± 0.49	98.42± 0.49	98.49± 0.63	98.58± 0.77	<b>98.76± 0.38</b>	99.08± 0.34
OA (%)	83.52± 0.52	85.66± 0.34	83.09± 0.58	85.42± 0.41	84.70± 0.23	82.35± 0.36	<b>86.81± 0.26</b>	87.85± 0.33
AA (%)	71.10± 0.79	72.96± 0.77	69.28± 0.83	73.63± 0.67	72.90± 0.88	67.79± 0.98	<b>73.90± 0.78</b>	77.92± 0.65
Kappa	0.809± 0.011	0.832± 0.007	0.805± 0.012	0.828± 0.009	0.823± 0.013	0.789± 0.015	<b>0.837± 0.012</b>	0.860± 0.005

The best results are highlighted in bold.

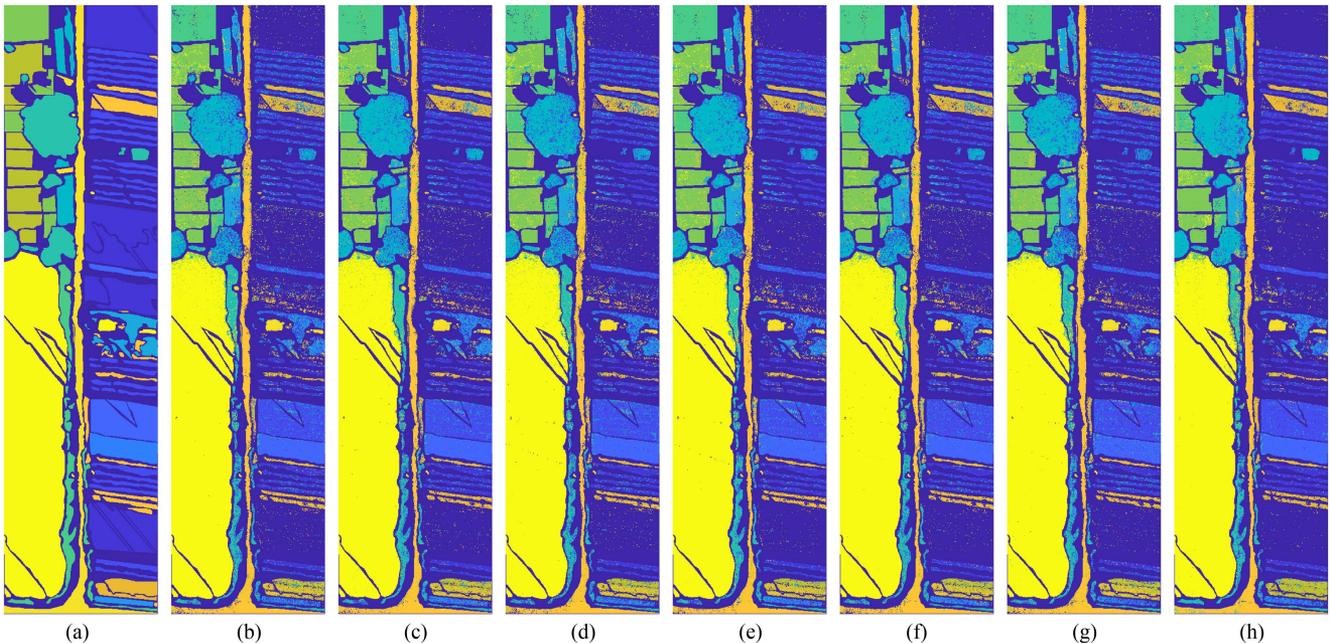


Fig. 4. Classification maps of WHU-Hi-HanChuan dataset obtained by different methods. (a) Ground truth. (b) OCF. (c) ISSC. (d) ONR. (e) ASPS. (f) FNGBS. (g) MGSR. (h) SPCA-AMGL.

that SPCA-AMGL can select a more informative band subset. In addition, ASPS and FNGBS demonstrate competitive performance by leveraging subspace partitioning strategies to reduce interband redundancy. However, these methods do not surpass the performance achieved by SPCA-AMGL. The MGSR model, which prioritizes band clustering performance, fails to select representative bands for classification, resulting in the lowest OA among all BS methods.

Second, the band subset size significantly impacts classification performance. Generally, as the band subset size increases, all BS methods tend to achieve better classification performance.

However, due to the limitations imposed by the availability of labeled samples, the Hughes phenomenon emerges as the band subset size further increases. For instance, in Fig. 1(b), the OA of ISSC starts to decline when the number of bands exceeds 35. Similarly, in Fig. 1(d), OCF and MGSR experience decreasing OA values when the number of bands surpasses 30 and 25, respectively. Despite this, SPCA-AMGL still avoids this phenomenon across all datasets, suggesting its capability to select more representative bands.

Third, when compared with using all bands, SPCA-AMGL consistently achieves better or comparable performance on

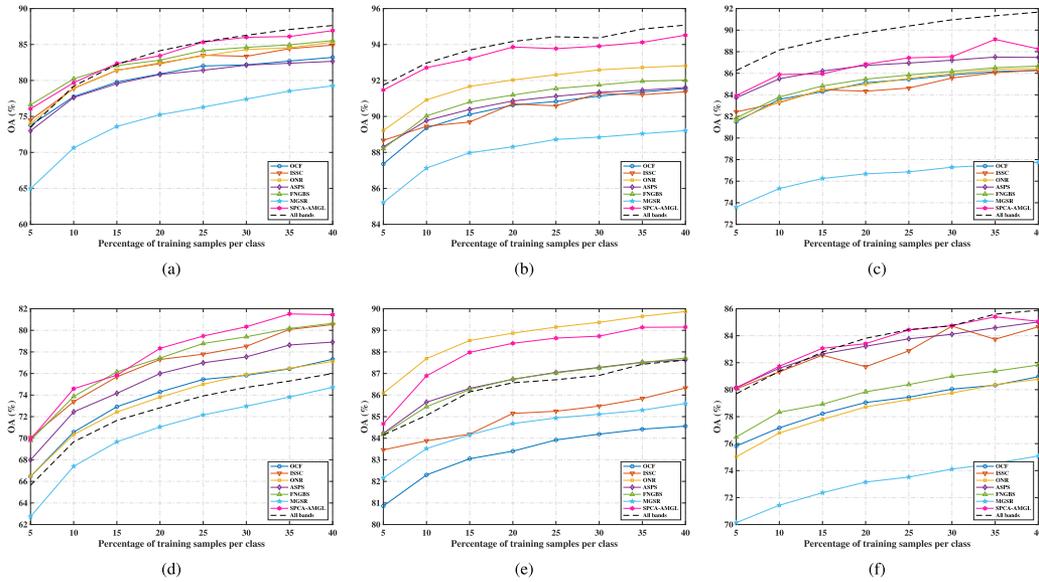


Fig. 5. Classification performance of different methods under different numbers of training samples. (a)–(c) Classification results of using SVM on Indian Pines, Pavia University, and WHU-Hi-HanChuan datasets, respectively. (d)–(f) Classification results of using KNN on Indian Pines, Pavia University, and WHU-Hi-HanChuan datasets, respectively.

all three datasets. For example, on the Indian Pines dataset, SPCA-AMGL matches the OA achieved using all bands with just 40 bands in SVM classification, and even with only ten bands in KNN classification, SPCA-AMGL surpasses the OA of using all bands. In contrast, other methods either fall short of or require more bands. This indicates that the complete band set contains noisy bands that can degrade classification performance. Similar trends are observed in the Pavia University and WHU-Hi-HanChuan datasets, where SPCA-AMGL tends to achieve better OA with fewer bands than other BS methods.

Fourth, SPCA-AMGL demonstrates superior performance across class accuracy, OA, AA, and Kappa. Tables I–III show that SPCA-AMGL is the best method under 50 bands. Specifically, on the Indian Pines dataset, SPCA-AMGL achieves the highest class accuracy for eight ground objects, and this number is five and nine for the Pavia University and WHU-Hi-HanChuan datasets, respectively. Moreover, SPCA-AMGL consistently achieves better performance in terms of OA, AA, and Kappa. While ONR, ASPS, and FNGBS also exhibit competitive classification performance among other BS methods, the proposed SPCA-AMGL consistently outperforms them. In addition, from Figs. 2–4, it is evident that our proposed method generates superior classification maps and fewer misclassified points, further substantiating its effectiveness.

For the experimental results, we summarize the following aspects to explain why our method achieved better accuracy.

- 1) From the perspective of reconstruction, our method of selecting high-quality bands is a global strategy. On the contrary, methods, such as OCF, ONR, ASPS, and FNGBS adopt a subspace partitioning strategy, which select the most representative bands in each group, resulting in a more localized BS. Because in some cases, multiple representative bands may be located in the same group, but only one band is ultimately selected. On the other hand,

some groups may contain many bands with poor quality, but some of them may still be selected.

- 2) When selecting bands, our method fully considers maintaining the local neighborhood structure of low dimensional data to enhance its separability. We achieve this by applying local manifold preserving constraint on the projection matrix. Furthermore, unlike MGSR, our method does not use a fixed similarity matrix. It integrates the similarity matrix and projection matrix into a unified framework, achieving mutual promotion and updating during the optimization process.

### C. Classification Performance With Different Numbers of Training Samples

This experiment compares the classification accuracies of SPCA-AMGL with other six methods by changing the training samples. The percentage of training samples is changed from 5% to 40%. The band subset sizes on the Indian Pines, Pavia University, and WHU-Hi-HanChuan datasets are 30, 30, and 30, respectively. All the parameters in the six methods are the same as in experiment in Section IV-B.

Fig. 5 shows the OA curves obtained by seven methods using SVM and KNN classifiers on three datasets. Similar to Fig. 1, SPCA-AMGL exhibits excellent performance in most cases. Among other methods, ISSC, ONR, ASPS, and FNGBS perform differently on different datasets. For example, FNGBS outperforms other comparison methods on the Indian Pines dataset. On the Pavia University dataset, ONR outperforms other comparison methods, and even outperforms SPCA-AMGL in the KNN classifier case [see Fig. 5(e)]. On the WHU-Hi-HanChuan dataset, ASPS performs well. But from an overall perspective, SPCA-AMGL achieves advanced performance and wide applicability to datasets.

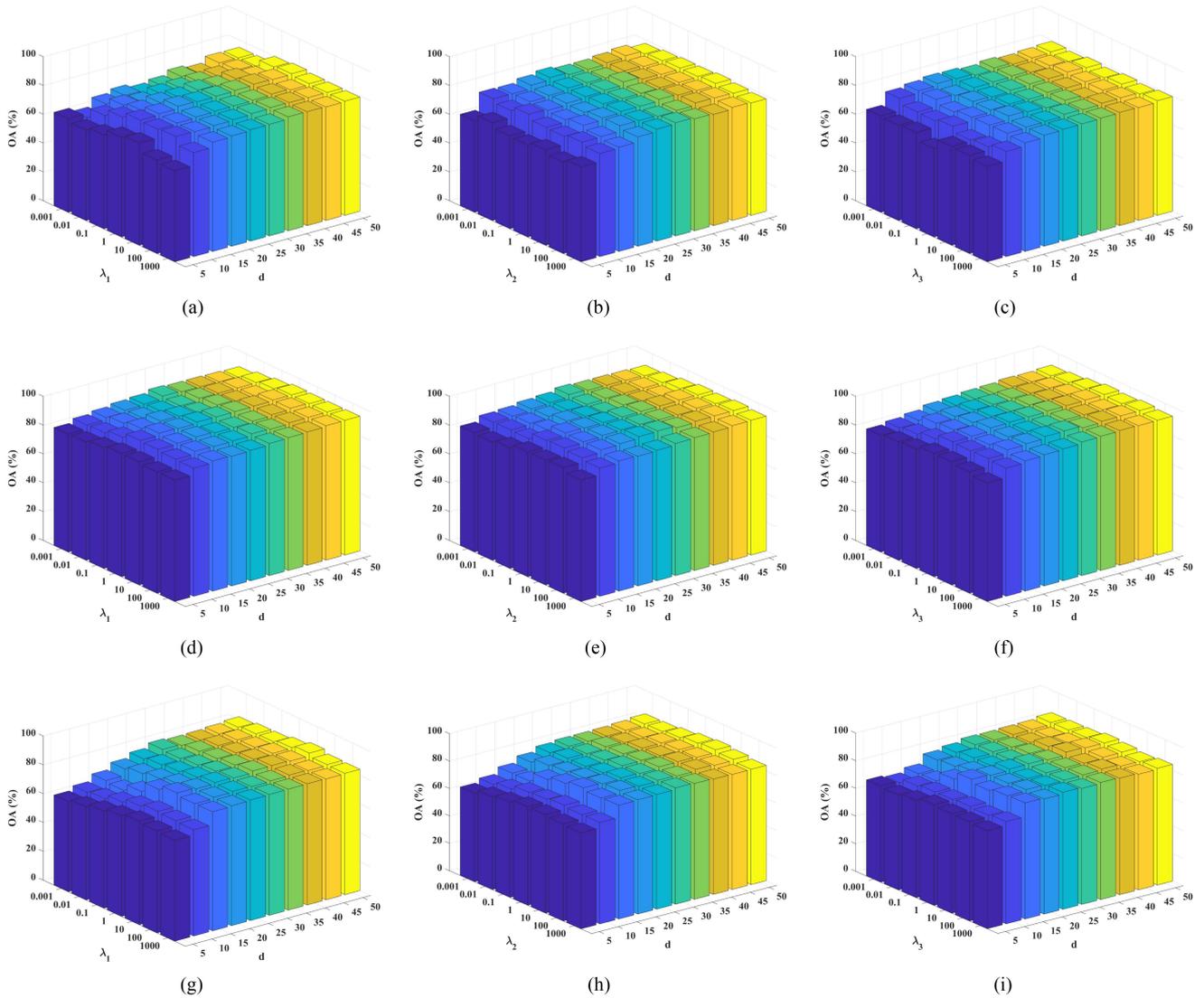


Fig. 6. OA under different choice of  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  on three datasets. (a)–(c) Indian Pines. (d)–(f) Pavia University. (g)–(i) WHU-Hi-HanChuan.

#### D. Analysis of Parameters Sensitivity

In this section, we conduct extensive parameter analysis on the proposed algorithm to evaluate the impact of different parameter settings on algorithm performance. Specifically, we focus on the following key parameters:  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ , and the value of  $k$  in the initial KNN graphs (note that this is not the same as  $k$  in the KNN classifier).

First, we calculate the effects of  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  on the values of OA, and the results are shown in Fig. 6. We use the grid search strategy to guide the optimal parameter setting. Specifically, in the sensitivity test of  $\lambda_1$ , the other two parameters are fixed. Then,  $\lambda_1$  takes a value from  $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3\}$ , and the number of selected bands ranges from 5 to 50. As depicted in Fig. 6, it is evident that the parameter  $\lambda_1$  has the most significant influence on OA. A substantial decline in OA is observed when  $\lambda_1$  is relatively small. For instance, as illustrated in Fig. 6(a), on the Indian Pines dataset with 30 selected bands, OA is notably lower when  $\lambda_1$  is 0.001 compared with the case when  $\lambda_1$  is

1. Another notable observation is that  $\lambda_2$  and  $\lambda_3$  have minimal impact on OA across all tested datasets. Moreover, after selecting an appropriate range for  $\lambda_1$ , SPCA-AMGL consistently achieves excellent OA values over these datasets. Therefore, we believe that after selecting suitable values for  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$ , they can be used on varying datasets, which demonstrates the widespread applicability of SPCA-AMGL. This stands in stark contrast to MGSR, which is highly sensitive to parameter choices, severely limiting its practicality. For our method, we suggest that the value of  $\lambda_1$  should be between 1 and 100, and the values of  $\lambda_2$  and  $\lambda_3$  should be greater than 0.

Second, to explore the impact of the initial KNN graphs on the SPCA-AMGL model, we conduct an evaluation with varying the value of neighbors. In this experiment, we construct multiple initial KNN graphs by setting  $k$  from 1 to 19, while selecting 15 and 30 bands. The experimental results are presented in Fig. 7. It is evident that across the three datasets, when the value of  $k$  is less than five or greater than 15, the OA is lower than when  $k$  falls within the range of 5 to 15. Our interpretation is that when

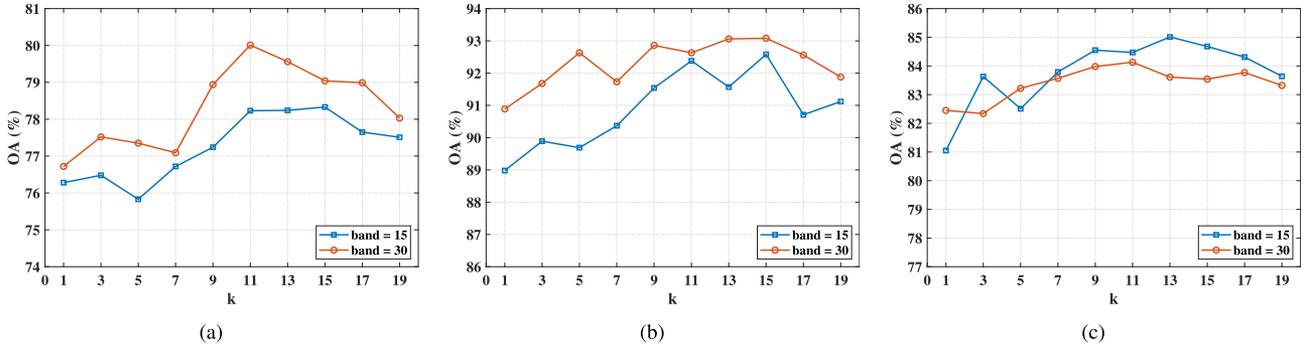


Fig. 7. Influence of  $k$  in KNN graph for (a) Indian Pines, (b) Pavia University, and (c) WHU-Hi-HanChuan.

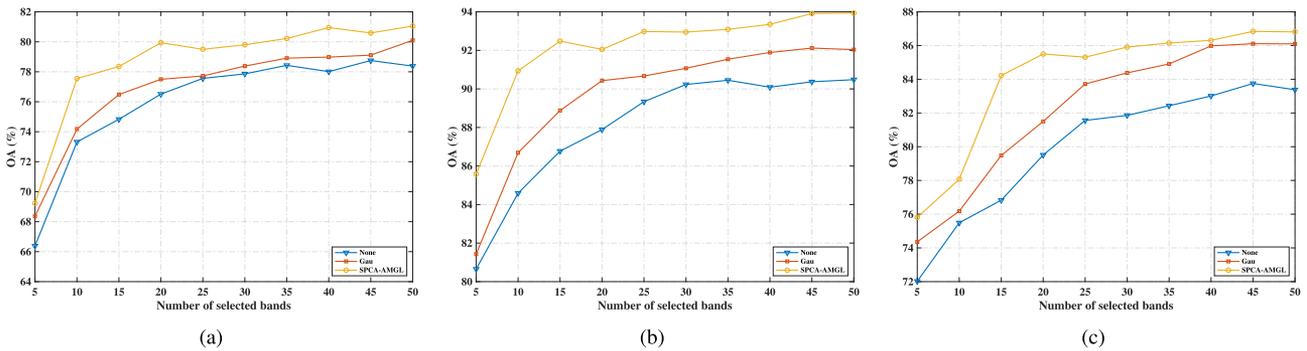


Fig. 8. Classification performance of different graph construction methods. (a) Indian Pines. (b) Pavia University. (c) WHU-Hi-HanChuan.

$k$  is excessively small, the neighboring structure depicted by the constructed graph remains relatively simplistic. Conversely, with overly large  $k$  values, the construction of graph becomes susceptible to the influence of noisy pixels. Consequently, based on empirical observations, setting  $k$  within the range of 5 to 15 is considered appropriate.

### E. Ablation Experiment

To investigate the efficacy of an AMGL, we conduct ablation experiments on the three datasets. Specifically, we conduct three comparative experiments, i.e., no any graph, a neighborhood graph based on Gaussian kernel, and three neighborhood graphs within SPCA-AMGL: Gaussian kernel, Laplacian kernel, and Cosine similarity. The results are presented in Fig. 8, where it is evident that methods employing multigraph learning outperform those relying solely on a single graph, and both of these approaches outperform the strategy that do not use graph. It is worth noting that the efficacy of employing multiple graphs is not a universal maxim, and arbitrary utilization of multiple graphs does not necessarily guarantee superiority over other methods. Different combinations of graphs need to be carefully considered for comprehensive evaluation.

### F. Comparison of Running Time

To quantify the impact of computational efficiency, we conduct a thorough runtime analysis of the proposed algorithm against existing state-of-the-art techniques. The experiments are

performed on a computer with Intel Core i5-12490F (4.2 GHz) CPU and 32 GB RAM, and all the methods are implemented in MATLAB R2021b. Table IV represents the running time of different methods when 15 bands are selected on the three datasets. All values are the average of ten independent runs.

The results of the runtime comparison, as depicted in Table IV, highlight notable differences among the methods. The proposed algorithm significantly reduces processing times on large-scale datasets compared with its counterparts. For instance, on the WHU-Hi-HanChuan dataset, the proposed method achieves a runtime reduction of 88% when compared with OCF and 60% when compared with FNGBS. This substantial improvement in runtime efficiency stems from the algorithm's implementation strategy of selecting a fixed number of data inputs. These optimizations not only accelerate the computational process but also enhance the scalability of the algorithm to handle larger datasets without sacrificing accuracy.

### G. Convergence Analysis

In order to assess the convergence behavior of our proposed algorithm, we conduct a series of convergence experiments across the three datasets. The purpose of these experiments is to analyze the rate and stability at which the algorithm converges toward an optimal solution. We initialize the algorithm and track the changes in objective function values over iterations.

The results, depicted in Fig. 9, indicate that our algorithm consistently converges toward a stable solution for all datasets (less

TABLE IV  
RUNNING TIME (S) OF DIFFERENT METHODS TO SELECT 15 BANDS ON THREE DATASETS

Dataset	OCF	ISSC	ONR	ASPS	FNGBS	MGSR	SPCA-AMGL
Indian Pines	0.16	0.14	0.42	0.08	0.04	5.86	0.37
Pavia University	0.51	0.64	0.47	0.33	0.17	10.27	0.31
WHU-Hi-HanChuan	3.59	3.86	3.32	1.56	1.06	23.37	0.42

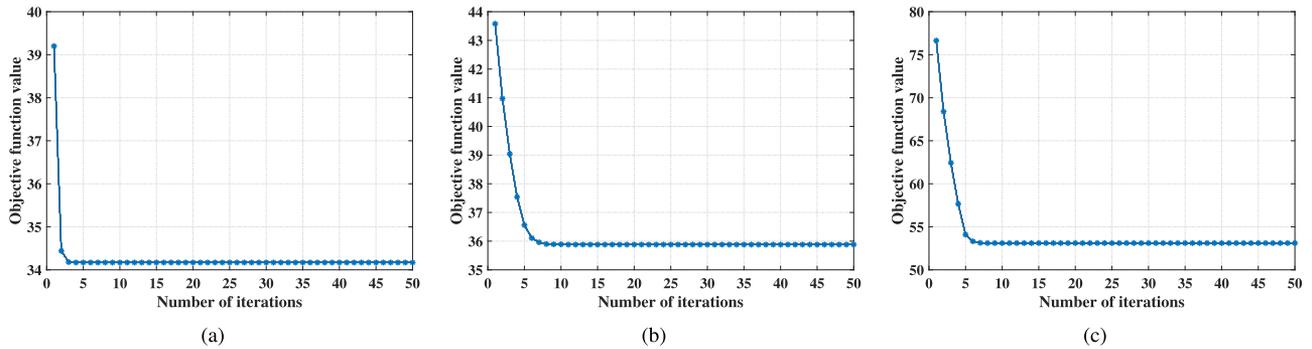


Fig. 9. Convergence curves of objective function values in SPCA-AMGL on three datasets. (a) Indian Pines. (b) Pavia University. (c) WHU-Hi-HanChuan.

than ten iterations). The convergence curves exhibit a smooth and rapid descent initially, followed by a gradual stabilization of the objective function values. The consistent convergence behavior across different datasets demonstrates the robustness and reliability of our algorithm. These findings underscore the algorithm's ability to efficiently optimize the BS process, ensuring the selection of informative and discriminative spectral bands for hyperspectral data classification applications.

## V. CONCLUSION

In this article, we represent a novel unsupervised BS method called SPCA-AMGL for HSI dimensionality reduction. The SPCA-AMGL assumes that there is a low-dimensional embedding subspace, where the most important information of HSI is retained. Therefore, this approach revolves around the idea that identifying the band significance for projecting the original HSI data into the low-dimensional subspace. To this end, SPCA-AMGL proposes to apply SPCA with  $L_{2,1}$  norm sparse constraint to learn the projection matrix. In this way, the most of information in the original data is retained and the orthogonal constraint also reduces the correlation between low-dimensional bands. Meanwhile, a local manifold preserving constraint based on LPP is proposed to preserve the local neighbor structure of HSI in the low-dimensional subspace. To obtain a sufficient description of the data geometry, an AMGL strategy is introduced to learn a better quality similarity matrix. Moreover, the proposed method integrates BS and multigraph learning into a framework so that they can promote each other during the process of optimization. Therefore, our method ensures that the bandwise information, interband correlation, and manifold-preserving capability of selected bands are simultaneously considered. Finally, the band significance is expressed in a projection weight matrix.

Related experiments are conducted on three HSI datasets: 1) Indian Pines, 2) Pavia University, and 3) WHU-Hi-HanChuan. Through experiments, we can conclude that SPCA-AMGL are

excellent in selecting high-quality bands, stability, and parameter sensitivity. Therefore, it can effectively reduce the dimension of HSI for subsequent hyperspectral tasks. In future work, we hope to further improve the AMGL model. For example, the 2-D spatial relationship of pixels in image data is naturally a measure of similarity, which can be used to construct a new graph to characterize the local nearest neighbor structure of HSI.

## REFERENCES

- [1] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.
- [2] A. F. Gao, B. Rasmussen, P. Kulits, E. L. Scheller, R. Greenberger, and B. L. Ehlmann, "Generalized unsupervised clustering of hyperspectral images of geological targets in the near infrared," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 4294–4303.
- [3] P. Kempeneers, S. De Backer, W. Debruyne, P. Coppin, and P. Scheunders, "Generic wavelet-based hyperspectral classification applied to vegetation stress detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 610–614, Mar. 2005.
- [4] R. J. Ellis and P. W. Scott, "Evaluation of hyperspectral remote sensing as a means of environmental monitoring in the St. Austell China clay (Kaolin) region, Cornwall, UK," *Remote Sens. Environ.*, vol. 93, no. 1/2, pp. 118–130, 2004.
- [5] W. Sun, L. Tian, Y. Xu, D. Zhang, and Q. Du, "Fast and robust self-representation method for hyperspectral band selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 5087–5098, Nov. 2017.
- [6] S. Yan, X. Xu, D. Xu, S. Lin, and X. Li, "Beyond spatial pyramids: A new feature extraction framework with dense spatial sampling for image classification," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 473–487.
- [7] X. Fang et al., "Approximate low-rank projection learning for feature extraction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5228–5241, Nov. 2018.
- [8] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.
- [9] F. Nie, S. Yang, R. Zhang, and X. Li, "A general framework for auto-weighted feature selection via global redundancy minimization," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2428–2438, May 2019.
- [10] X. Li, H. Zhang, R. Zhang, and F. Nie, "Discriminative and uncorrelated feature selection with constrained spectral analysis in unsupervised learning," *IEEE Trans. Image Process.*, vol. 29, pp. 2139–2149, 2019.

- [11] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 51–64, Jan. 2017.
- [12] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, Jan. 2011.
- [13] X. Cao, T. Xiong, and L. Jiao, "Supervised band selection using local spatial information for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 329–333, Mar. 2016.
- [14] X. Cao, C. Wei, Y. Ge, J. Feng, J. Zhao, and L. Jiao, "Semi-supervised hyperspectral band selection based on dynamic classifier selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1289–1298, Apr. 2019.
- [15] F. He, F. Nie, R. Wang, W. Jia, F. Zhang, and X. Li, "Semisupervised band selection with graph optimization for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10298–10311, Dec. 2021.
- [16] C. Sui, Y. Tian, Y. Xu, and Y. Xie, "Unsupervised band selection by integrating the overall accuracy and redundancy," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 185–189, Jan. 2015.
- [17] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.
- [18] J. Wang, H. Wang, Z. Ma, L. Wang, Q. Wang, and X. Li, "Unsupervised hyperspectral band selection based on hypergraph spectral clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5509905.
- [19] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.
- [20] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.
- [21] S. Li and H. Qi, "Sparse representation based band selection for hyperspectral images," in *Proc. IEEE 18th Int. Conf. Image Process.*, 2011, pp. 2693–2696.
- [22] A. Martínez-Usó, Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
- [23] W. Sun, J. Peng, G. Yang, and Q. Du, "Correntropy-based sparse spectral clustering for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 484–488, Mar. 2020.
- [24] Y. Cai, X. Liu, and Z. Cai, "Bs-Nets: An end-to-end framework for band selection of hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1969–1984, Mar. 2020.
- [25] J. Feng et al., "Dual-graph convolutional network based on band attention and sparse constraint for hyperspectral band selection," *Knowl.-Based Syst.*, vol. 231, 2021, Art. no. 107428.
- [26] J. Wang, C. Tang, X. Zheng, X. Liu, W. Zhang, and E. Zhu, "Graph regularized spatial-spectral subspace clustering for hyperspectral band selection," *Neural Netw.*, vol. 153, pp. 292–302, 2022.
- [27] X. Shang, C. Cui, and X. Sun, "Spectral-spatial hypergraph-regularized self-representation for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5504405.
- [28] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.
- [29] S. Yi, Z. He, X.-Y. Jing, Y. Li, Y.-M. Cheung, and F. Nie, "Adaptive weighted sparse principal component analysis for robust unsupervised feature selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 6, pp. 2153–2163, Jun. 2020.
- [30] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [31] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [32] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Adv. Neural Inf. Process. Syst.*, vol. 14, pp. 585–591, 2001.
- [33] L. Zhang and F. Luo, "Review on graph learning for dimensionality reduction of hyperspectral image," *Geo-Spatial Inf. Sci.*, vol. 23, no. 1, pp. 98–106, 2020.
- [34] X. He and P. Niyogi, "Locality preserving projections," *Adv. Neural Inform. Process. Syst.*, vol. 16, pp. 153–160, 2003.
- [35] Q. Lu, X. Li, and Y. Dong, "Structure preserving unsupervised feature selection," *Neurocomputing*, vol. 301, pp. 36–45, 2018.
- [36] A. Yuan, M. You, D. He, and X. Li, "Convex non-negative matrix factorization with adaptive graph for unsupervised feature selection," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 5522–5534, Jun. 2020.
- [37] P. Zhou, L. Du, X. Li, Y.-D. Shen, and Y. Qian, "Unsupervised feature selection with adaptive multiple graph learning," *Pattern Recognit.*, vol. 105, 2020, Art. no. 107375.
- [38] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
- [39] X. Chen, R. Chen, Q. Wu, F. Nie, M. Yang, and R. Mao, "Semisupervised feature selection via structured manifold learning," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 5756–5766, Jul. 2021.
- [40] M. You, A. Yuan, M. Zou, D. Jian He, and X. Li, "Robust unsupervised feature selection via multi-group adaptive graph representation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, pp. 3030–3044, Mar. 2023.
- [41] S. Yang, J. Hou, Y. Jia, S. Mei, and Q. Du, "Superpixel-guided discriminative low-rank representation of hyperspectral images for classification," *IEEE Trans. Image Process.*, vol. 30, pp. 8823–8835, Oct. 2021.
- [42] J. Yue, L. Fang, and M. He, "Spectral-spatial latent reconstruction for open-set hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 31, pp. 5227–5241, Aug. 2022.
- [43] M. You, A. Yuan, D. He, and X. Li, "Unsupervised feature selection via neural networks and self-expression with adaptive graph constraint," *Pattern Recognit.*, vol. 135, 2023, Art. no. 109173.
- [44] Y. Yuan, X. Li, Q. Wang, and F. Nie, "A semi-supervised learning algorithm via adaptive Laplacian graph," *Neurocomputing*, vol. 426, pp. 162–173, 2021.
- [45] F. Nie, R. Zhang, and X. Li, "A generalized power iteration method for solving quadratic problem on the Stiefel manifold," *Sci. China Inf. Sci.*, vol. 60, no. 11, pp. 1–10, 2017.
- [46] Y. Zhong et al., "Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 4, pp. 46–62, 2018.
- [47] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "Whu-HI: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *Remote Sens. Environ.*, vol. 250, 2020, Art. no. 112012.
- [48] W. Sun, L. Zhang, B. Du, W. Li, and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784–2797, Jun. 2015.
- [49] Q. Wang, F. Zhang, and X. Li, "Hyperspectral band selection via optimal neighborhood reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8465–8476, Dec. 2020.
- [50] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940–4950, Dec. 2019.
- [51] Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5028–5039, Jun. 2021.
- [52] Y. Zhang, X. Wang, X. Jiang, and Y. Zhou, "Marginalized graph self-representation for unsupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2021, Art. no. 5538513.



**Wenxian Zhang** (Member, IEEE) received the B.S. degree in software engineering from Nanjing University of Information Science and Technology (NUIST), Nanjing, China, in 2019. He is currently working toward the M.S. degree in computer science and technology with Northwest A&F University, Xi'an, China.

His research interests include machine learning, computer vision, and hyperspectral image analysis.



**Aihong Yuan** received the Ph.D. degree in signal and information processing from the Center for Optical Imagery Analysis and Learning (OPTIMAL), Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Beijing, China, in 2019.

He is currently a Lecturer with Computer Science Department, Northwest A&F University, Xianyang, China. His research interests include machine learning, image/video content understanding, and deep learning.



**Jinglei Tang** received the Ph.D. degree in agricultural electronization and automation from Northwest A&F University (NWSUAF), Xianyang, China, in 2010.

She is currently an Associate Professor with the College of Information Engineering, NWSUAF and a Researcher with the Key Laboratory of Shaanxi Province for Complex System Control and Intelligent Information Processing, Xi'an University of Technology, Xi'an, China. Her research interest includes machine learning, computer vision, pattern recognition, image analysis, decision support, and information

management.

**Xuelong Li** (Fellow, IEEE) is a Full Professor with the School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an, China.