A Geometry-Based Band Selection Approach for Hyperspectral Image Analysis

Wenqiang Zhang, Xiaorun Li, Yaxing Dou, and Liaoying Zhao

***Abstract*— Band selection (BS) is a special case of the feature selection problem, and it tries to remove redundant bands and select a few informative and distinctive bands to represent the whole image cube. The maximum ellipsoid volume (MEV) method regards the band subset with the maximum volume as the optimal band combination. However, the MEV method cannot be directly applied for hyperspectral imagery due to the high dimensionality of the data sets. Therefore, we first combine MEV with the sequential forward search (SFS) and propose a new unsupervised BS method called MEV-SFS. Furthermore, a subtle relationship between the ellipsoid volume of the band set and the orthogonal projections (OPs) of the candidate bands is observed. Based on this relationship, we propose another equivalent method, namely, the OP-based BS (OPBS) method. OPBS is the fast version of MEV-SFS, and it has a better computational efficiency and the potential to determine the number of bands to be selected. We specifically explain the rationality of the MEV-based methods (MEV-SFS and OPBS) and illustrate their theoretical significance and physical meaning from different aspects. Theoretical analysis also demonstrates that OPBS can be regarded as a model or framework for BS, and thus, we further propose a third novel BS method named the OPBS-information divergence (OPBS-ID) method, which is a variant of OPBS. OPBS-ID can achieve a better classification performance than OPBS in many cases. Experimental results on different hyperspectral data sets demonstrate that the proposed methods have high computational efficiency, and the selected bands can achieve satisfactory classification performances.**

***Index Terms*— Dimensionality reduction (DR), feature selection, hyperspectral remote sensing, orthogonal projection (OP), sequential forward search (SFS).**

I. Introduction

HYPERSPECTRAL images contain hundreds of bands with a fine resolution, e.g., 0.01 μm, which makes it possible to reduce the overlap between classes and therefore enhance the potential to discriminate subtle spectral differences [1], [2]. However, the high dimensionality of the data set also results in several problems, such as heavy computational and storage cost. In addition, the high resolution of the spectrum causes high correlation among the bands. Therefore, to process data effectively, dimensionality reduction (DR) is important and necessary. One approach of DR is feature extraction, which reduces the data dimensionality by extracting a set of new features from the original ones through some functional mapping. For instance, principal component analysis (PCA) [3] is one of the well-known feature extraction methods, and other feature extraction methods include nonnegative matrix factorization [4], independent component analysis [5], local linear embedding [6], maximum noise fraction [7], and so on. However, the physical meaning of the original data may be changed and some important information may be compromised and distorted during the transformation [8].

Another DR approach is band selection (BS) [9]. BS methods reduce the feature space by selecting a subset of the original features. BS is preferable for feature extraction in hyperspectral imagery because BS methods select a subset of bands without losing their physical meaning and have the advantage of preserving the relevant original information in the data. In terms of the object information availability, BS methods can be divided into two categories: supervised BS and unsupervised BS methods. Supervised BS methods try to find the most informative bands with respect to the available prior knowledge, whereas unsupervised methods do not assume any object information. However, prior knowledge is often unavailable in practice, and in these cases, supervised BS methods are not suitable. Therefore, we focus on unsupervised BS methods in this paper. In recent years, many unsupervised BS methods have been proposed. Some of them are based on band ranking, where different criteria are used to measure the importance of the bands. These include information divergence BS (IDBS) [10], constrained BS [10], linearly constraint minimum variance (LCMV) [10], maximum-variance PCA (MVPCA) [11], and mutual information [12]. Other BS methods take bands' correlation into consideration and resort to finding the bands combination with the optimal indexes, such as optimal index factor [13], maximum ellipsoid volume (MEV) [14], maximum information (MI) [15], minimum-dependent information [16], volume-gradient-based BS (VGBS) [17], and other similar methods [18], [19]. Recently, exploiting correlation through clustering algorithms has attracted increased attention in the field of BS [20]. Some methods based on clustering have been proposed, such as affinity propagation [21], exemplar component analysis [22],

Manuscript received September 11, 2017; revised November 6, 2017 and December 27, 2017; accepted February 18, 2018. Date of publication March 20, 2018; date of current version July 20, 2018. This work was supported in part by the National Nature Science Foundation of China under Grant 61671408 and Grant 61571170, in part by the Joint Fund Project of the Chinese Ministry of Education under Grant 6141A02022314, and in part by the Shanghai Aerospace Science and Technology Innovation Fund under Grant SAST2016028. *(Corresponding author: Xiaorun Li.)*

W. Zhang, X. Li, and Y. Dou are with the College of Electrical Engineering, Zhejiang University, Zhejiang 310027, China (e-mail: wqzhang@zju.edu.cn; lxrly@zju.edu.cn; 11510054@zju.edu.cn).

L. Zhao is with the Department of Computer Science, Hangzhou Dianzi University, Zhejiang 310027, China (e-mail: zhaoly@hdu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2811046

k-means-based BS methods [23], and so on. In the clustering-based methods, each band is considered as a data point, and the data set is partitioned into groups of similar bands (clusters) without any class label information. Furthermore, many BS methods based on advanced machine learning algorithms have been proposed; for instance, [24] proposed a novel BS method based on manifold ranking (MR), [25] proposed a multitask sparsity pursuit framework for BS, and in [26], a BS method based on multiple graph is proposed. These methods design new criteria for BS and apply efficient suboptimal searching methods such as clone selection [26] to avoid exhaustive search. It has been proved that these methods can achieve quite good performances for classification task [24]–[26].

Among all these BS methods, the MEV method attracts our attention. The MEV method regards the band set with the MEV as the best combination, and the volume of bands can be measured by the determinant of the bands' covariance matrix [17]. However, MEV was originally proposed for multispectral images and required an exhaustive search to find the optimal band combination. But an exhaustive search is impractical in hyperspectral remote sensing due to the high dimensionality of hyperspectral data sets. Therefore, Geng *et al.* [17] adopt the sequential backward search (SBS), which is one of the widely used suboptimal searching methods [27], to find a relatively good band combination and propose a novel BS method named the VGBS [17]. VGBS is derived from the MEV method, and it removes the band with the largest volume gradient iteratively, which is equivalent to repetitively removing the band that maximizes the remaining bands' covariance matrix determinant [17]. Moreover, to compute the volume gradient efficiently, VGBS performs singular value decomposition (SVD) to reduce the dimensionality of each band before BS, which reduces the computational complexity of VGBS significantly. Experimental results demonstrate that the VGBS method can achieve a good classification performance [17].

In this paper, we proposed three geometry-based BS methods, that is, MEV-sequential forward search (SFS), orthogonal-projection-based BS (OPBS), and OPBS-ID, which are also related with the MEV method. We first combined MEV with the SFS [27] and proposed a new BS method named MEV-SFS. The MEV-SFS method selects one band at one time and regards the band that maximizes the determinant of the selected bands' covariance matrix as the optimal band. Furthermore, a subtle relationship between the ellipsoid volume of the band set and the OPs of the candidate bands is found. Based on this relationship, we further propose another equivalent BS method, the OPBS method, which is a fast version of MEV-SFS. We find that the band with the largest distance to the hyperplane spanned by the selected bands is exactly the band that maximizes the bands' ellipsoid volume. From a viewpoint of the geometric theory of a simplex, the distance to the selected bands can be regarded as the height from the candidate band to the simplex spanned by the selected bands, and it is equal to the OP on the simplex base [28]. Therefore, OPBS calculates the OPs of all the candidate bands and tries to find the band with the largest OP as the desired band. Interestingly, the OPs of candidate bands can be obtained by using recursive formulas, which reduces the computational complexity of the OPBS method significantly. Thus, the OPBS method has the advantage of low computational cost. Additionally, the bands obtained by OPBS (or MEV-SFS) have large amounts of information and low band correlation, which is another advantage of the proposed methods. Moreover, OPBS is also a model or framework for BS; therefore, we further propose a third BS method, namely, the OPBS-ID method, which is one of the variants of OPBS and can achieve a better classification performance than OPBS in most cases. Additionally, we find that the OPBS method and its variants have the potential for determining the number of bands to be selected, which can also be regarded as one of the advantages of the proposed methods.

It can be found that OPBS is similar to VGBS, since both the methods are based on MEV. The major differences between them are as follows: VGBS adopts SBS whereas OPBS adopts SFS, and generally, SFS has a better computational efficiency than SBS; VGBS reduces the computational complexity by computing the volume gradient of each candidate band, whereas OPBS makes it by incrementally calculating the OPs of each candidate band, and the incremental calculation of OPs is easier than computing the derivative of the covariance matrix determinant. Additionally, VGBS requires to perform SVD to reduce the dimensionality of the whole data set before BS, which in fact is also a time-consuming process since SVD is usually computationally complex, whereas OPBS does not need to perform any DR preprocess to the data set. Therefore, OPBS is simpler and has a better computational efficiency compared with VGBS. As for the classification performance, although OPBS and VGBS apply different suboptimal searching strategies, they both resort to find the bands with the MEV, and in fact, the results obtained by them can be considered as the same in most cases. Additionally, OPBS has some other advantages; for instance, OPBS can be regarded as a model for BS and it is easy to modify or change the metrics in OPBS to obtain different variants, and some variants like OPBS-ID can achieve better performances than OPBS and VGBS. Moreover, OPBS has the potential to determine the number of bands to be selected. Therefore, compared with VGBS, OPBS has more advantages and has a wider range of applications.

The major contributions of this paper would be summarized as follows.

1) We combined MEV with SFS and proposed the MEV-SFS method. The adoption of SFS allows for using the MEV method in hyperspectral imagery.
2) A subtle relationship between the ellipsoid volume of the selected bands and the OPs of candidate bands is found. Furthermore, we derived the recursive formulas of OPs and proposed another efficient method, i.e., the OPBS method.
3) We explained the rationality of the MEV-based methods and illustrated their theoretical significance from three aspects. The theoretical analysis explains why the proposed methods can find the bands that are informative and distinctive.

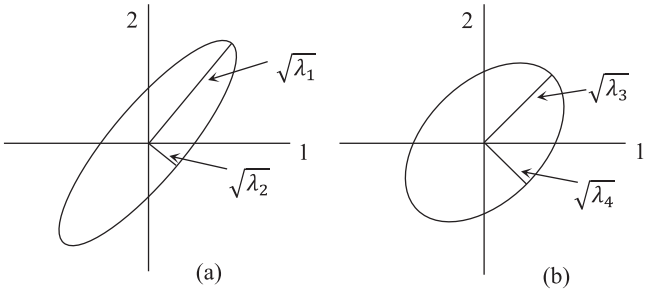


Fig. 1. Different measure criteria for information. (a) Sum of the square root of variances (information). (b) Product of the square root of variances.

4) A new joint correlation metric is proposed in this paper. Specifically, the angle between a single band and the hyperplane spanned by other bands can be used to evaluate the joint correlation between a single band and multiple bands.
5) OPBS is further extended as a model for BS, and thus one of its variants called OPBS-ID is introduced in this paper.

The remainder of this paper is organized as follows. Section II presents the MEV-SFS method. Section III specifically presents the OPBS method. Section IV shows the theoretical analysis of OPBS and presents the OPBS-ID method. The evaluation experiments, which are conducted on different real data sets, are described in Section V. Finally, Section VI shows some concluding remarks.

II. MEV-SFS

A. MEV BS

MEV is a BS method originally proposed for multispectral images. It attempts to find the bands that have a large amount of information and low correlation at the same time. Reference [17] gives an example in 2-D, which is shown in Fig. 1. The total variance in Fig. 1(a) is larger than that in Fig. 1(b), i.e., $\lambda_1 + \lambda_2 > \lambda_3 + \lambda_4$. However, the correlation of the bands in Fig. 1(a) is higher than that in Fig. 1(b); therefore, $\lambda_1 \times \lambda_2 < \lambda_3 \times \lambda_4$. The ellipse in Fig. 1(b) has a larger area than the ellipse in Fig. 1(a). When extended to a high-dimensional space, MEV attempts to select the ellipsoid with the MEV, which discourages the selection of band sets with high correlation. It can be proven that the volume of the ellipsoid equals the determinant of the covariance matrix of bands, except for a constant factor [17].

For a band set $X \in \mathbb{R}^{N\times L}$, where L and N are the numbers of bands and pixels in the set, assume that the mean value of each band has been removed, then the ellipsoid volume of X can be measured by the determinant of the covariance matrix (neglecting the constant factor) [17]. For computational convenience, the square root term is neglected in the following. Therefore, the criterion of the MEV method is selecting the band set with the maximum determinant of the covariance matrix.

B. SFS

The MEV method was originally designed for the Landsat-4 Thematic Mapper, which is a multispectral sensor and has seven bands [14]. Since there are only a total of 35 combinations for selecting three out of seven bands, an exhaustive search can be executed rapidly. However, for a hyperspectral image with hundreds of bands, exhaustive strategies cannot be used due to the huge computational time. Just as stated in [29] and [30], an exhaustive search for the optimal solution is prohibitive from a computational viewpoint. Therefore, many suboptimal search methods are widely used in the research on BS. The most commonly used search strategies can be found in [27] and [29]–[33]. In this paper, the SFS is used as the searching strategy. SFS is a simple greedy search algorithm that belongs to the heuristic suboptimal search methods. SFS starts with an empty set of features and adds the feature x that maximizes the cost function $f(Y_k + x)$ when combined with the features Y_k that have already been selected, until a feature subset with the desired quantity is obtained.

By combining the MEV method with SFS, we propose a new BS method, which is denoted as MEV-SFS. The detailed procedures are presented in Algorithm 1.

Algorithm 1 MEV-SFS Algorithm

Input: Observations $X = [x_1, x_2, \ldots, x_L] \in \mathbb{R}^{N\times L}$, the number of selected bands n.
Initialization: Calculate the covariance matrix of all bands, $C = X^T X$ (then, all entries in the following $\Phi_t^T \Phi$ can be acquired from C). Find the band with the maximum variance as the initial selected band set, i.e., $\Phi = \{x_{id(1)}\}$. Then set counter $i = 2$.
Band selection:
while $i < n + 1$ **do**
1: Tentatively add the candidate band to the selected band set by turns and determine the band to be selected, i.e.,

$$\Phi = \arg\max_{\Phi_t} det(\Phi_t^T \Phi)$$

where $\Phi_t = \Phi \cup \{x_t\}, t = 1, 2, 3, \ldots, L$. If the tth band x_t has already been selected, its Φ_t and corresponding determinant will not be calculated and compared.
2: $i \leftarrow i + 1$
end while
Output: Selected band set Φ.

III. OPBS

In the MEV-SFS method, each band is tentatively added into the selected band set by turns, and the corresponding ellipsoid volume constructed by these bands is calculated at the same time. In each round of lookup, we select the band that maximizes the selected bands' volume as the desired band and begin the next round of searching. The sequence is repeated in this manner until the desired number of bands has been obtained. In this section, a subtle relationship between the ellipsoid volume of the selected bands and the OPs of candidate bands is found, and thus, we further propose a fast

version of the MEV-SFS method, that is, the OPBS method, in which the computation of the volume (or determinant) is replaced by the incremental calculation of the OPs of the candidate bands. It should be noted that, in this paper, the mean value of each band has been removed unless otherwise stated, namely, for any band B, its mean value equals zero.

A. Relationship Between OPs and Volume

First, we derive the relationship between the OPs of candidate bands and the volume of selected bands. For the MEV-SFS method, the desired bands are obtained one by one. In each round of lookup, the band that maximizes the determinant of the bands' covariance matrix is regarded as the optimal band. Assume that we have obtained k selected bands, and the selected band set is denoted as $Z = [B_1, B_2, \ldots, B_k] \in \mathbb{R}^{N\times k}$, then, to find the $(k+1)$th desired band, for a candidate band B, we need to calculate its associated covariance matrix determinant as follows:

$$V = \det([Z, B]^T [Z, B]). \tag{1}$$

The band that maximizes V will be selected as the optimal band. Equation (1) can also be written as follows:

$$V = \det(Z^T Z) \times B^T [I - Z(Z^T Z)^{-1} Z^T] B \tag{2}$$

where I denotes an identity matrix. It can be found that, for any candidate band B, $\det(Z^T Z)$ is the same constant; therefore, we have a new objective function, i.e.,

$$h = B^T P^{\perp} B \tag{3}$$

$$P^{\perp} = I - Z(Z^T Z)^{-1} Z^T \tag{4}$$

where h denotes the square of the candidate bands distance to the selected bands, and the band that maximizes h will be regarded as the optimal band. It is worth noting that matrix $P^{\perp}$ is symmetric and idempotent, that is, $P^{\perp} = (P^{\perp})^T$ and $P^{\perp} = (P^{\perp})^2$. Therefore, (3) can also be written as

$$h = \|P^{\perp} B\|^2 \tag{5}$$

where $\|*\|$ denotes the Euclidean norm. In linear algebra and related fields of mathematics [34], the vector subspace spanned by the bands in Z can be denoted as $U = \mathrm{Span}\{B_1, B_2, \ldots, B_k\}$, and $U^{\perp}$ denotes the orthogonal complement subspace of U. Then, $P^{\perp} B$ actually denotes the OP of B onto the subspace $U^{\perp}$; therefore, $P^{\perp}$ is called the orthogonal projector. Similarly, the projector that maps on the subspace U can be obtained by

$$P = Z(Z^T Z)^{-1} Z^T \tag{6}$$

where $P^{\perp}$ is also the orthogonal complement of P, and it is easy to find that $P^{\perp} = I - P$ and $P^{\perp} P = 0$. Thence, to find the band with the largest distance to the selected bands is equivalent to selecting the band that has the maximum OP $\|P^{\perp} B\|$. This is the basic idea of OPBS.

B. Incremental Calculation

Then, for the OPBS method, when a new desired band has been selected, we need to update the projector $P^{\perp}$ (or P) and start the next iteration. However, we note that the dimension of the projector $P^{\perp}$ (or P) is N by N, where N is the number of the pixels and usually very huge, and thus, the computation and storage of the projectors are unacceptable in practice. Fortunately, we find that the projector $P^{\perp}$ (or P) can be updated by using recursive formulas, which allows us to achieve the incremental calculation of the OPs without computing and storing the projectors. To derive the recursive forms of (4) and (6), for the selected band set Z, we have the subspace $U_Z = \mathrm{Span}\{Z\}$ and the projector $P_Z = Z(Z^T Z)^{-1} Z^T$, and the newly selected band is denoted as B_{k+1}. Then, we need to update P_Z, and the new projector is denoted as $P_{Z,B}$. From the viewpoint of adaptive updating, $P_{Z,B}$ can be decomposed into two orthogonal parts: the nonadaptive updating part P_Z (or the known part) and the adaptive updating part P_w [34], i.e.,

$$P_{Z,B} = P_Z + P_w \tag{7}$$

$$\langle P_Z, P_w \rangle = 0 \tag{8}$$

where P_Z does not reflect any information about the newly selected band B_{k+1}, and the adaptive updating part P_w should include the effect of B_{k+1}. According to the linear subspace theory, P_w is also an orthogonal projector [34]. Without loss of generality, P_w can be written as

$$P_w = w(w^T w)^{-1} w^T \tag{9}$$

where vector w is a linear transformation of vector B_{k+1} because P_w reflects the effect of B_{k+1}. Then, according to (8), we have

$$\langle P_Z, P_w \rangle = P_Z w(w^T w)^{-1} w^T = 0. \tag{10}$$

Because $w \neq 0$, we have $P_Z w = 0$, which demonstrates that the OP of the vector w onto U_Z equals zero, that is, $w \in U_Z^{\perp}$. Moreover, vector w is a linear transformation of B_{k+1}, so w should be the OP of B_{k+1} onto the vector space $U_Z^{\perp}$, namely, $w = P_Z^{\perp} B_{k+1}$ (strictly speaking, w is a vector with the same direction as the vector $P_Z^{\perp} B_{k+1}$, namely, $w = \alpha P_Z{}^{\perp} B_{k+1}$, where α is a scale factor with nonzero real values and does not influence the derivation of the following equations). Thence, we have

$$P_w = P_Z^{\perp} B_{k+1} \left(B_{k+1}^T P_Z^{\perp} B_{k+1}\right)^{-1} \left(P_Z^{\perp} B_{k+1}\right)^T. \tag{11}$$

And therefore, we obtain the recursive forms of (4) and (6), i.e.,

$$P_{Z,B} = P_Z + P_Z^{\perp} B_{k+1} \left(B_{k+1}^T P_Z^{\perp} B_{k+1}\right)^{-1} \left(P_Z^{\perp} B_{k+1}\right)^T \tag{12}$$

$$P_{Z,B}^{\perp} = P_Z^{\perp} - P_Z^{\perp} B_{k+1} \left(B_{k+1}^T P_Z^{\perp} B_{k+1}\right)^{-1} \left(P_Z^{\perp} B_{k+1}\right)^T. \tag{13}$$

C. OPBS Method

By using the recursive formula in (13), we can directly achieve the incremental calculation of the OPs and it is unnecessary to compute and store the projectors. Suppose there is a data set $X = [x_1, x_2, \ldots, x_L] \in \mathbb{R}^{N\times L}$, the number of

desired bands is set to be n, and $id(i)$ $(i = 1, 2, \ldots, n)$ denotes the index of the ith selected band. Since the OPBS method adopts SFS, the desired bands are selected one by one. To find n desired bands, we need to perform n rounds of lookup. For convenience, the OP vector of the tth band x_t in the ith round can be denoted as follows:

$$y_t^{(i)} = P_i^{\perp} x_t \tag{14}$$

$$h_t^{(i)} = \left\| y_t^{(i)} \right\|^2 = \left(y_t^{(i)} \right)^T y_t^{(i)} \tag{15}$$

where $y_t^{(i)}$ denotes the OP vector of x_t and $h_t^{(i)}$ denotes the squared Euclidean norm of the OP vector of the candidate band x_t. Then, the band that has the largest $h_t^{(i)}$ (or OP) is selected as the desired band, and the next round of lookup begins. According to (13), the new orthogonal projector $P_{i+1}^{\perp}$ can be incrementally updated by using $P_i^{\perp}$ and $x_{id(i)}$, and thus, for the tth band in the $(i+1)$th round, we have

$$\begin{aligned} y_t^{(i+1)} &= P_{i+1}^{\perp} x_t \\ &= P_i^{\perp} x_t - P_i^{\perp} x_{id(i)} \left(x_{id(i)}^T P_i^{\perp} x_{id(i)} \right)^{-1} \left(P_i^{\perp} x_{id(i)} \right)^T x_t \\ &= y_t^{(i)} - y_{id(i)}^{(i)} \left(h_{id(i)}^{(i)} \right)^{-1} \left(y_{id(i)}^{(i)} \right)^T x_t. \end{aligned} \tag{16}$$

It is evident that the OP of the same candidate band can also be calculated by a recursive formula. Furthermore, according to $P^{\perp} = (P^{\perp})^T$ and $P^{\perp} = (P^{\perp})^2$, (16) can be justified as

$$y_t^{(i+1)} = y_t^{(i)} - y_{id(i)}^{(i)} \left[\left(y_{id(i)}^{(i)} \right)^T y_t^{(i)} / h_{id(i)}^{(i)} \right] \tag{17}$$

which demonstrates that the current OP vector $y_t^{(i+1)}$ is only associated with two variables: $y_t^{(i)}$ and $y_{id(i)}^{(i)}$, which are, respectively, the OP vectors of x_t and $x_{id(i)}$ in the previous round. All the variables in (17) have been computed in the previous iterations; moreover, $y_t^{(i)}$ and $y_{id(i)}^{(i)}$ are $N \times 1$ dimensional vectors and $h_{id(i)}^{(i)}$ is a scalar. Therefore, the calculation of (17) only involves low-complexity vector multiplication and scalar multiplication, which is more efficient than the high-order matrix multiplication and matrix inversion in (3) and (4). Furthermore, the term $(y_{id(i)}^{(i)})^T y_t^{(i)}$ is first computed in (17) and it is also a scalar, so this equation avoids the generation of high-order matrix variables during the calculation, which is useful for saving computing time and storage space. The basic steps of the OPBS method are presented in Algorithm 2.

IV. Theoretical Analysis

Although the experimental results in [17] have proven that the MEV-based methods consider the band correlation and show a good performance, the underlying rationality of this kind of methods is not well explained. Therefore, in this section, we give the detailed explanation of the rationality of the MEV-based methods (e.g., MEV-SFS and OPBS). Furthermore, we abstract the OPBS method as a model or framework for BS, which allows us to give one of the variants of OPBS, i.e., OPBS-ID. Additionally, we find that the OPBS and its variants have an interesting property, and the OP (or the value of the objective function) of each newly selected band is smaller than that of the previously selected band, which can be used to determine the quantity of bands to be selected.

Algorithm 2 OPBS Algorithm

Input: Observations $X = [x_1, x_2, \ldots, x_L] \in \mathbb{R}^{N \times L}$, the number of selected bands n.

Initialization: Find the band with the maximum variance as the initial selected band set, i.e., $\Phi = \{x_{id(1)}\}$. Let $Y = [y_1, y_2, \ldots, y_L] = X$, $h_{id(1)} = \|x_{id(1)}\|^2$ and set counter $i = 2$.

Band selection:

while $i < n + 1$ or (41) is not met **do**

1: Calculate the OP vector of the tth band $x_t, (t = 1, 2, \ldots, L)$.

$$y_t = y_t - y_{id(i-1)} [y_{id(i-1)}^T y_t / h_{id(i-1)}]$$

2: Select the band with the maximum OP as the optimal band and add it into Φ, i.e.,

$$\begin{aligned} h_t &= y_t^T y_t \\ id(i) &= \arg\max_t (h_t) \\ \Phi &= \Phi \cup \{x_{id(i)}\} \end{aligned}$$

If the tth band x_t has already been selected, its y_t and h_t will not be calculated and compared.

3: $i \leftarrow i + 1$

end while

Output: Selected band set Φ.

A. Explanation of OPBS

The OPBS method (or MEV-SFS) has the advantage of considering the information and band correlation simultaneously. In this section, we explain the theoretical significance and physical meaning of the OPBS method and illustrate why the proposed methods can find the bands with a large amount of information and low correlation. The OPBS method can be explained from three aspects.

1) Orthogonal Decomposition: The OPBS method is also an orthogonal decomposition based method. Just as shown in Fig. 2, OPBS decomposes the candidate band B into two orthogonal components: $B_{\|}$ and $B_{\perp}$, i.e.,

$$B = B_{\|} + B_{\perp} \tag{18}$$

where $B_{\|} = PB$ and $B_{\perp} = P^{\perp} B$, and (18) is called the orthogonal decomposition of B relative to U_Z. Obviously, the component $B_{\perp}$ is exactly the OP vector of the band B onto the vector space $U_Z^{\perp}$, so the objective function of OPBS can also be expressed as

$$h = \|P^{\perp} B\|^2 = \|B_{\perp}\|^2. \tag{19}$$

Thus, OPBS tries to find the band with the largest orthogonal component $\|B_{\perp}\|^2$. It is worth noting that OPBS neglects the contribution of another orthogonal component $B_{\|}$, and this is exactly the key characteristic of the OPBS method. Since $B_{\|}$ belongs to the vector space U_Z, it can be linearly represented by using the selected bands; therefore, it is reasonable to consider that $B_{\|}$ cannot provide extra (or useful) information for the selected band set; in other words, $B_{\|}$ denotes the information redundancy between the candidate band B and the

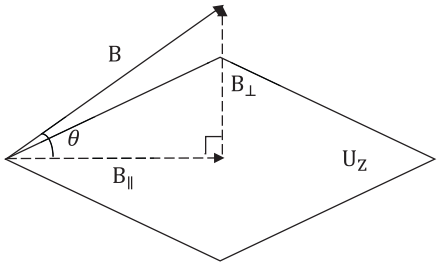


Fig. 2. 3-D example for illustrating the mechanism of the OPBS method. The candidate band B is orthogonally projected onto the vector space U_Z. $B_\parallel$ and $B_\perp$ are the two orthogonal components.

selected band set Z. In contrast, $B_\perp$ is perpendicular to any band in Z, so there is no redundancy between the components $B_\perp$ and Z. Additionally, because the mean value of each band has been removed, it can be proved that $\|B_\perp\|^2$ exactly denotes the variance of $B_\perp$ [since the average of any band equals zero, and according to (23), it is easy to prove that the averages of $B_\parallel$ and $B_\perp$ also equal zero, so the variance of $B_\perp$ is exactly $1/N$ of the value for $\|B_\perp\|^2$], and the variance (or the standard deviation) is often used as the information metric in BS, so the OP (i.e., $\|B_\perp\|^2$) can be regarded as the useful (or extra) information that the candidate band B can provide to the band combination. Therefore, from this point of view, OPBS can be explained as finding the band that is able to provide the largest amount of useful (or extra) information.

2) Linear Regression: The OPBS method is also based on the linear regression. As we previously mentioned, OPBS decomposes the candidate band B into two orthogonal parts: PB and $P^\perp B$. In fact, from the viewpoint of the linear regression, PB denotes the linear estimate or prediction of B using the bands in Z, and the objective function $\|B_\perp\|^2$ evaluates the prediction error.

Likewise, for the selected band set $Z \in \mathbb{R}^{N\times k}$, the OP of the candidate band B is measured by

$$\|B_\perp\| = \|P^\perp B\| = \|B - PB\| \tag{20}$$

where the term PB can be expressed as

$$\hat{B} = PB = Z(Z^TZ)^{-1}Z^TB = Z[(Z^TZ)^{-1}Z^TB] \tag{21}$$

where $\hat{B}$ denotes the linear estimate of the band B, and its meaning will be introduced in the following. Then, note that $[(Z^TZ)^{-1}Z^TB]$ is a $k \times 1$ vector, so it can be denoted as

$$a = [a_1, a_2, \ldots, a_k]^T = (Z^TZ)^{-1}Z^TB \tag{22}$$

where a_i denotes a scalar; then, (21) can be rewritten as

$$\hat{B} = Z \times a = a_1B_1 + a_2B_2 + \cdots + a_kB_k = \sum_{i=1}^{k} a_iB_i \tag{23}$$

which demonstrates that $\hat{B}$ (or PB) is exactly a linear combination of the selected bands in Z and a is a set of weights that determine how each selected band affects the prediction. Additionally, it can be proved that $a = (Z^TZ)^{-1}Z^TB$ is exactly a least squares solution. Thus, $\hat{B}$ (or PB) is the linear estimate or prediction of B using the bands in Z. Furthermore, the objective function of OPBS actually estimates the mean squared error (MSE) between the band B and its estimate $\hat{B}$, i.e.,

$$\mathrm{MSE}_B = \frac{1}{N}\|B - \hat{B}\|^2 = \frac{1}{N}h \tag{24}$$

where N is the number of pixels. Obviously, the MSE is exactly $1/N$ of the square norm of the candidate bands OP. Therefore, OPBS is also a linear regression-based method; it uses the selected bands to estimate the candidate band, and the band with the maximum MSE will be selected.

Additionally, when a band has a large MSE, it also demonstrates that the band is difficult to be linearly expressed by the selected bands and thus has a low correlation with them. From the viewpoint of linear correlation, if the candidate band can be described easily by a linear expression, the band is highly correlated with the selected bands and cannot provide sufficiently useful information to the selected band set. In the worst case, the candidate band is exactly linearly expressed by the selected bands, which demonstrates that the band is linearly correlated with them and thus can be considered as a totally redundant band. Therefore, from the perspective of linear regression, OPBS tries to find the band with a low redundancy with the selected bands, and the band with lower correlation is more likely to be selected.

3) Basic Rules of BS: Usually, there are two basic ideas guiding the design of an unsupervised BS method. First, because the process of DR almost inevitably results in the loss of information, to minimize the loss, we prefer retaining the bands with large amounts of information. Second, we also wish that the selected bands have a low redundancy with each other, which ensures that the selected band set can provide sufficiently useful information for further applications. A good BS method should consider both the information and band redundancy (usually measured by the band correlation). Interestingly, the OPBS method is perfectly consistent with these two rules, which ensures that it can achieve a good performance in practice.

The objective of OPBS is to find the band with the largest distance to the selected bands, and its objective function is shown in (5), and we can analyze it from the geometric perspective. Likewise, take Fig. 2 as an example, then it is easy to find that the OP of the candidate band B can also be obtained by

$$\|P^\perp B\| = \|B_\perp\| = \sin\theta \times \|B\| \tag{25}$$

$$\sin\theta = \|P^\perp B\|/\|B\| = \|P^\perp \overline{B}\| \tag{26}$$

where $\overline{B}$ is the normalized band of B (namely, $\overline{B}$ is the unit vector in the direction of B). Thence, it can be seen from (25) that the objective function of OPBS actually consists of two parts: $\sin\theta$ and $\|B\|$. The former denotes the sine function of the angle between the candidate band and the hyperplane spanned by the selected bands, and the latter represents the norm (or standard deviation) of the candidate band.

As we mentioned before, for any band B, the mean value equals zero, so it can be proved that the variance of B is exactly $1/N$ of its squared Euclidean norm $\|B\|^2$, where N is a constant and denotes the number of pixels in a band image.

Since $1/N$ is the same for any band, it can be neglected and thus the Euclidean norm $\|B\|$ can be regarded as the standard deviation of B. Therefore, in (25), $\|B\|$ actually estimates the amount of information of B, which means the band with a larger amount of information is more likely to be selected by OPBS. As for $\sin\theta$, it actually evaluates the joint correlation between the candidate band and the selected bands. The angle between the single band and the hyperplane spanned by other bands can measure the joint correlation between a single band and multiple bands. As we mentioned in Section IV-A2, if the angle θ is large, the candidate band is difficult to be linear expressed by the selected bands and thus has a low correlation with them. For instance, in the best case, θ equals 90°, which means the band is perpendicular to any band in the selected band set, and we can consider that there is no correlation between the band and the selected bands. Therefore, in (25), $\sin\theta$ evaluates the joint correlation between the candidate band and the selected bands; the larger the value for $\sin\theta$ is, the lower the correlation is, which means the band with lower correlation is more likely to be selected. It should be noted that, although it seems that $\sin\theta$ should be influenced by $\|B\|$ according to (26), the fact is that the value for $\sin\theta$ is only related with the direction of vector B instead of its norm $\|B\|$, because $\sin\theta = \|P^{\perp}\overline{B}\|$, which means $\sin\theta$ is not influenced by the norm of B and we can use the unit vector of B to calculate $\sin\theta$.

The metric θ (or $\sin\theta$) also reminds us of a widely used correlation metric, the correlation coefficient, i.e.,

$$\rho_{i,j} = \frac{\sum_{m=1}^{N}(x_{i,m}-\mu_i)(x_{j,m}-\mu_j)}{\sqrt{\sum_{m=1}^{N}(x_{i,m}-\mu_i)^2\sum_{m=1}^{N}(x_{j,m}-\mu_j)^2}} \quad (27)$$

where $\rho_{i,j}$ denotes the correlation coefficient between the bands x_i and x_j, μ_i and μ_j denote the average of x_i and x_j, respectively, and $x_{i,m}$ and $x_{j,m}$ denote the mth pixel of x_i and x_j. Coincidently, for the MEV-based methods (OPBS and MEV-SFS), we use the mean-shifted data in the BS process, because mean-shifting is necessary for the calculation of the covariance matrix; therefore, (27) can be simplified as

$$\rho_{i,j} = \frac{\sum_{m=1}^{N}x_{i,m}x_{j,m}}{\sqrt{\sum_{m=1}^{N}x_{i,m}^2\sum_{m=1}^{N}x_{j,m}^2}} = \frac{x_i^T x_j}{\|x_i\|\|x_j\|} = \cos\theta_{i,j}. \quad (28)$$

Obviously, the correlation coefficient also measures the band correlation by estimating the angle between two vectors (bands). Interestingly, when there is only one band in the selected band set, the cosine of the angle θ in our proposed method is exactly the correlation coefficient between the candidate band and the selected band. From this point of view, the cosine of the angle θ can be regarded as the extension of the correlation coefficient in a high-dimensional space. However, it should be noted that the correlation coefficient is a pairwise correlation metric; namely, it is used to estimate the correlation between two bands, whereas the metric θ (or $\sin\theta$) in this paper is able to measure the joint correlation between a single band and multiple bands. Thus, in the proposed methods, the band correlation is evaluated jointly instead of pairwise and the new joint correlation metric can also be regarded as one of the contributions of this paper.

Because the objective function of OPBS consists of an information metric and a correlation metric, only the band with a large amount of information and a low correlation with the selected bands can maximize the objective function, which explains why the proposed methods can consider the information and band correlation simultaneously. Moreover, (25) further explains the physical meaning of the proposed methods. As aforementioned, $\|B\|$ denotes the total information of the candidate band B, and it is reasonable to consider that the useful information provided by B should be less than its total information. We also notice that the value of $\sin\theta$ ranges exactly from 0 to 1, so $\sin\theta$ can be regarded as the percentage of the useful information. Therefore, the OPs (i.e., $\|P^{\perp}B\| = \sin\theta \times \|B\|$) can be regarded as the useful (or extra) information that the candidate band can provide. For instance, in the best case, B is perpendicular to the bands in Z, and all the information of B can be regarded as the useful information. Whereas in the worst case, B is linearly correlated with the selected bands and the useful information reduces to zero. Only the band with a large amount of information and a low redundancy will be selected by the proposed methods. In this manner, the proposed methods ensure that there is sufficiently useful information in the band combination.

B. OPBS Model

Just as shown in (25), the objective function of the OPBS method consists of two parts: the band correlation metric $\sin\theta$ and the information metric $\|B\|$, which indicates that we can replace the information metric $\|B\|$ with other similar criteria to obtain the variant versions of OPBS. From this point of view, OPBS can be regarded as a model or framework for BS, which includes various implementations. The objective function of the OPBS model can be expressed as

$$\text{Index} = \sin\theta \times \text{info}(B) \quad (29)$$

where Index denotes the value of the objective function and $\text{info}(B)$ denotes the amount of information of the candidate band B, and it can be measured by different criteria. For instance, we can use the ID [10] as the information metric and give one of the variants of the OPBS method, which is denoted as OPBS-ID. The ID is defined as follows:

$$D(B) = \sum_{i=1}^{N} p_i \log\frac{p_i}{q_i} + \sum_{i=1}^{N} q_i \log\frac{q_i}{p_i} \quad (30)$$

where $D(B)$ denotes the ID of the candidate band B, and $\text{info}(B) = D(B)$. In (30), $p = [p_1, p_2, \ldots, p_N]$ is the image histogram of band B and $q = [q_1, q_2, \ldots, q_N]$ is its associated Gaussian distribution with the mean and the variance determined by sample mean and sample variance of band B [10], and both p and q are normalized as a probability distribution. The ID is used to measure how much far away from a Gaussian distribution for a given band image [10]. The band with high non-Gaussianity can be regarded as the band with a large amount of information; thus, according to the viewpoints in Section IV-A, OPBS-ID also resorts to

finding the band with the largest amount of useful information. Of course, when compared with OPBS, OPBS-ID has a weaker theoretical significance; for instance, it cannot be perfectly explained from the perspective of linear regression.

For the OPBS-ID method, if we normalize all the bands and use the normalized bands to compute their associated OPs, the squared norm of the normalized bands' OP is exactly equal to the square of $\sin\theta$, i.e.,

$$h = \|P^{\perp}\overline{B}\| = (\sin\theta \times \|\overline{B}\|)^2 = (\sin\theta)^2 \tag{31}$$

where $\overline{B}$ is the normalized band of B (namely, $\overline{B}$ is the unit vector in the direction of B). Equation (31) demonstrates that $\sin\theta$ of each candidate band can also be incrementally calculated by using the recursive formula in (17). Additionally, the IDs of all bands are only calculated once and it is not computationally complex; therefore, OPBS-ID has the computational efficiency close to that of OPBS. Since it is easy to obtain the procedures of OPBS-ID by modifying the small parts of the OPBS algorithm, the OPBS-ID algorithm can be omitted for reducing the length of this paper.

C. Decline of OPs

OPBS and its variants have the potential to determine the suitable number of bands to be selected, which is another advantage of the proposed methods. Before we start introducing the method for determining the quantity of the selected bands, an important hypothesis for OPs needs to be illustrated first, that is, the OP of the newly selected band will be smaller than that of the previously selected bands (note that once the band is selected, its OP will not be updated). This viewpoint can be qualitatively explained according to the analysis in Section IV-A. The OPs of the candidate bands denote their useful information, and the useful information of the band obtained in this round should be less than that of the band obtained in the previous round. This occurs because the information redundancy between the candidate bands and the selected bands will become more significant during the BS process, as more bands are included in the selected band set. In other words, the useful information that the candidate bands can provide becomes less. Therefore, it is reasonable to think that the newly selected bands' OPs decline as the number of iterations increases. For instance, Fig. 3 shows the OP of the band obtained in each round for a real hyperspectral image. We can see that, initially, the selected bands' OPs decline rapidly as the number of iterations increases, and then, the rate of decrease declines, and eventually, the OPs become almost invariant (but are still slowly decreasing). However, the qualitative analysis and the special case cannot verify the rationality of our hypothesis perfectly; therefore, we provide the strict theoretical proof in the following. According to Section III-C, the OPs of two sequentially selected (or optimal) bands $x_{id(i)}$ and $x_{id(i+1)}$ are measured by $h_{id(i)}^{(i)} = \|y_{id(i)}^{(i)}\|^2$ and $h_{id(i+1)}^{(i+1)} = \|y_{id(i+1)}^{(i+1)}\|^2$, respectively. Then, our objective is to prove

$$h_{id(i)}^{(i)} \geq h_{id(i+1)}^{(i+1)}. \tag{32}$$

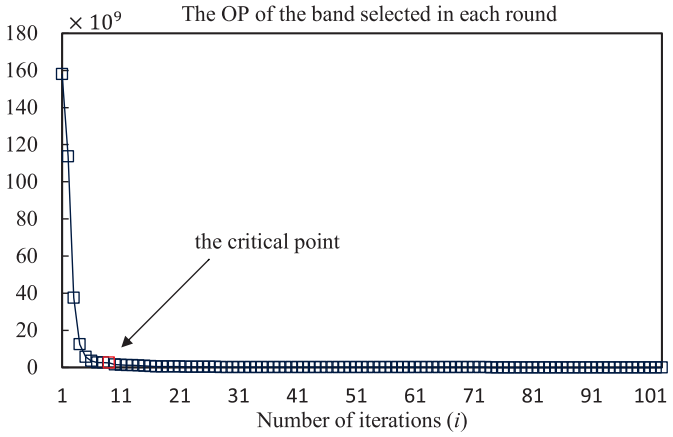


Fig. 3. OPs of the selected bands for the Pavia University data set, which has 103 bands (see Section V-C). As the number of iterations increases, the newly selected bands' OP declines. Red point: recommended number of the bands to be selected ($\varepsilon = 0.0015$).

However, $h_{id(i)}^{(i)}$ and $h_{id(i+1)}^{(i+1)}$ have no direct relationship, so we introduce a third variable, i.e.,

$$h_{id(i+1)}^{(i)} = \left\|y_{id(i+1)}^{(i)}\right\|^2. \tag{33}$$

Then, it is equivalent to proving that

$$h_{id(i)}^{(i)} \geq h_{id(i+1)}^{(i)} \geq h_{id(i+1)}^{(i+1)}. \tag{34}$$

Obviously, $h_{id(i)}^{(i)}$ is larger than $h_{id(i+1)}^{(i)}$ because the band $x_{id(i)}$ is the optimal band in the ith round. Therefore, we just need to prove

$$h_{id(i+1)}^{(i)} \geq h_{id(i+1)}^{(i+1)}. \tag{35}$$

We note that to prove (35) is equivalent to proving that for any band x_t, its OP will decrease as the number of iterations increases. That is, our final goal is to prove

$$h_t^{(i)} \geq h_t^{(i+1)}. \tag{36}$$

Therefore, we compute the equation as follows:

$$\Delta = \left\|y_t^{(i)}\right\|^2 - \left\|y_t^{(i+1)}\right\|^2 = \left(y_t^{(i)}\right)^T y_t^{(i)} - \left(y_t^{(i+1)}\right)^T y_t^{(i+1)} \tag{37}$$

and substitute (17) into (37) and yield

$$\Delta = \frac{\left\|\left(y_{id(i)}^{(i)}\right)^T y_t^{(i)}\right\|^2}{\left\|y_{id(i)}^{(i)}\right\|^2} \geq 0. \tag{38}$$

Therefore, (34)–(36) have been proved, and we obtain the relationship as follows:

$$h_{id(1)}^{(1)} \geq h_{id(2)}^{(2)} \geq \ldots \geq h_{id(i)}^{(i)} \geq \ldots \geq h_{id(n)}^{(n)} \tag{39}$$

$$h_t^{(1)} \geq h_t^{(2)} \geq \ldots \geq h_t^{(i)} \geq \ldots \geq h_t^{(n)}. \tag{40}$$

Equations (39) and (40) demonstrate that the values of the OPs decline during the BS, which has two meanings: first, for a newly selected band, its OP is smaller than that of the optimal band obtained in the previous round; second, for the same candidate band, its OP keeps declining as the number of iterations increases, until the band is selected by the proposed

method. Thence, the rationality of our previous hypothesis has been proved. Based on this property, OPBS can determine the suitable number of selected bands.

Considering that the newly selected bands' useful information decreases during the BS, when the number of the selected bands exceeds a specific size, adding more bands into the selected band set no longer increases the total information of the band combination significantly. In this case, the newly added bands' contribution can be ignored. This is the basic idea to determine the suitable number of desired bands. Here, we introduce a simple strategy to find how many desired bands are suitable. We notice that the declining rate of the OPs will be relatively slow when too many bands have been obtained (see Fig. 3). Therefore, when the declining rate of the OPs is lower than a specific threshold ε, the BS algorithm can be terminated. The termination condition is expressed as follows:

$$\sum_{i=k-1}^{k} \frac{h_{id(i-1)}^{(i-1)} - h_{id(i)}^{(i)}}{2h_{id(1)}^{(1)}} = \frac{h_{id(k-2)}^{(k-2)} - h_{id(k)}^{(k)}}{2h_{id(1)}^{(1)}} < \varepsilon \qquad (41)$$

where denominator $h_{id(1)}^{(1)}$ is used to scale the declining rates to a suitable range. Equation (41) denotes that when the average declining rate of the OPs of two sequentially selected bands is smaller than ε, the BS algorithm will be terminated and the critical point k can be regarded as the recommended number of bands to be selected (see Fig. 3). In other words, if the BS algorithm is terminated at the kth round, the number k will be regarded as the recommended number. The smaller the parameter ε, the more bands are obtained. Of course, the strategy can be modified or replaced depending on specific applications.

As for the variants of OPBS, their objective functions also exhibit the same phenomenon as the OPs. In fact, according to (19), (25), and (40), we can derive the relationship as follows:

$$\sin\theta_t^{(1)} \geq \sin\theta_t^{(2)} \geq \cdots \geq \sin\theta_t^{(i)} \geq \cdots \geq \sin\theta_t^{(n)} \qquad (42)$$

where $\theta_t^{(i)}$ denotes the angle between the candidate band x_t and the hyperplane spanned by the selected bands in the ith round. Then according to (29), it is easy to prove that the values for the objective function of OPBS-ID also decline during the BS. Therefore, the variants of OPBS also have the potential to determine the suitable number of bands to be selected and the strategy proposed in this paper can also be used for these methods.

D. Computational Complexity Analysis

Finally, we analyze the computational complexity of the proposed methods. OPBS is the fast version of MEV-SFS and has much better computation efficiency than MEV-SFS. As for OPBS-ID, its computational complexity is slightly higher than that of OPBS but is still much lower than that of MEV-SFS. Assume that we are selecting n bands from L bands, and each band has N pixels. The float-point operations (flops) are used to measure the computational complexity.

To select one desired band, MEV-SFS needs to calculate the bands' covariance matrix determinant. The complexity of the determinant computation for a kth-order matrix is $o(k^3)$ [17]. Therefore, for MEV-SFS, the number of flops for initialization and BS is NL^2 and $\sum_{i=2}^{n}(L-i+1)i^3$, respectively. Thus, the total flops for the MEV-SFS method are $NL^2 + \sum_{i=2}^{n}(L-i+1)i^3$.

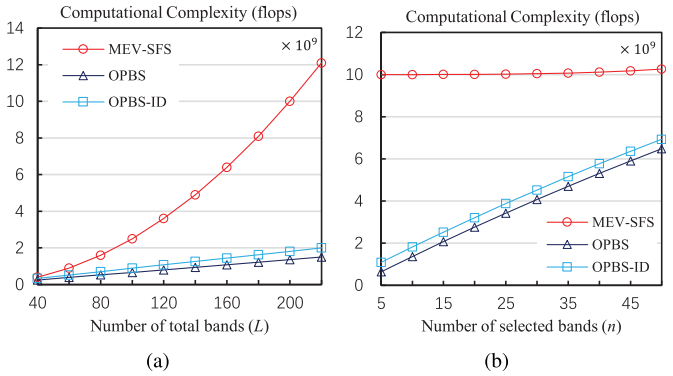


Fig. 4. Computational complexities of the MEV-SFS, OPBS, and OPBS-ID methods. (a) Computational complexity versus the number of total bands ($n = 10$). (b) Computational complexity versus the number of the selected bands ($L = 220$).

For the OPBS method, the incremental calculation of the OPs includes low-complexity vector multiplication and scalar multiplication. Specifically, the complexity of initialization is NL, and the complexity of BS is $\sum_{i=2}^{n}(L-i+1)(3N+1)$. Considering that N is usually much larger than L and n, the total complexity of OPBS is about $\sum_{i=2}^{n}(L-i+1)(3N)+NL$.

Compared with the OPBS method, the additional computation cost of OPBS-ID is mainly attributed to the band normalization and the computation of IDs, which results in about $8NL$ flops; thus, the total complexity of OPBS-ID is about $\sum_{i=2}^{n}(L-i+1)(3N)+9NL$.

The total complexities of the three methods are shown in Fig. 4. We can see that the OPBS and OPBS-ID methods are very similar in their computational complexities, and the values are much smaller than the values for the complexity of MEV-SFS. Fig. 4(a) shows that the number of total bands has little influence on the computational complexities of OPBS and OPBS-ID, whereas the complexity of MEV-SFS increases greatly with an increase in the number of total bands. Fig. 4(b) shows that although the MEV-SFS method is not sensitive to the number of the selected bands, OPBS and OPBS-ID are much superior to MEV-SFS, especially when selecting a small number of bands. Therefore, compared with MEV-SFS, OPBS and OPBS-ID have much smaller computational complexities, and the superiority will be more significant when a small number of bands are selected out of many bands. In fact, because the number of pixels is usually much larger than those of the selected bands and total bands, the computational complexity of OPBS is theoretically about $3n/L$ of that of MEV-SFS. For instance, if we select 10 bands out of 200 bands, the total computational complexity of OPBS is only about 15% (precisely, 13.7%) of that of MEV-SFS, and the total computational complexity of OPBS-ID is about 18.2% of that of MEV-SFS. Therefore, OPBS and OPBS-ID

have the significantly reduced computational complexities compared with MEV-SFS. Additionally, when compared with other BS methods that consider both information and band correlation, OPBS and OPBS-ID also show good computational efficiency (see Section V), which is also the advantage of the proposed methods.

V. Experiments

In this section, some comparative tests are conducted to evaluate the performance of the proposed methods. The experiments are based on three different hyperspectral data sets. Six different types of unsupervised BS methods, namely, MVPCA [11], LCMV band correlation constraint (LCMV-BCC) [10], LCMV band correlation minimization (LCMV-BCM) [10], MI [15], VGBS [17], and MR [24], are used for the comparison with the proposed methods. The comparison includes three aspects, i.e., classification accuracy, band correlation, and computing time. The pixel classifications of hyperspectral imagery are conducted, respectively, with two different classifiers: K-nearest neighborhood (KNN) [35] and support vector machine (SVM) [36]. Additionally, the recommended numbers of selected bands are also tested in this section. We give a brief description of these BS and classification methods that would be used in the experiments.

MVPCA is a classical BS method and is often used for the comparison of BS methods because of its simplicity and effectiveness. The MVPCA method selects the bands with high discriminative capabilities, and the discriminative capability of a band is evaluated by the variance-based band-power ratio of this band to that of total bands [11]. The LCMV-based methods aim to find the bands that best represent the whole image tube [10], and the representative ability is measured by the correlation with the whole image tube. Two criteria can be used as the BS criteria for LCMV, namely, BCC and BCM, and the associated methods are denoted as LCMV-BCC and LCMV-BCM, respectively. MI is a correlation-optimized-based BS method, in which the Kullback–Leibler (K-L) divergence is used to measure the information of a band, and by constructing an information matrix, the bands with a large amount of information and low correlation are selected [15]. MR is one of the state-of-the-art BS methods, which is based on the machine learning methods, including clustering algorithm, clone selection, and MR [24]. In the MR method, first all the bands are partitioned into several clusters by using the clustering algorithm such as k-means, then the clone selection [24] is used to find one representative band from each cluster to construct the initial selected band set, and finally, according to MR, more bands are added into the current selected band set to obtain the final band combination. The MR method can achieve a quite good performance for classification, because it includes the advantages of many effective machine learning methods. As for the VGBS method, it is also one of the state-of-the-art methods [17]; just as the OPBS method, VGBS is also derived from the MEV method. The major difference between OPBS and VGBS is that their subset searching strategies and specific selection criteria are different.

TABLE I
Number of Samples for Ground Objects in Indian Pines Data Set

Class	Training samples	Testing samples
1.Corn-notill	143	1291
2.Corn-mintill	83	751
3.Grass/Pasture	49	448
4.Grass/Trees	74	673
5.Hay-windrowed	48	441
6.Soybeans-notill	96	872
7.Soybeans-meantill	246	2222
8.Soybeans-clean	61	553
9.Woods	129	1165

The KNN classifier is one of the most fundamental and simple classification methods and should be one of the first choices when there is a little or no prior knowledge about the distribution of the data. The class of a new sample is determined by the labels of several training data points that are nearest to this sample [20]. In our experiments, the number of neighbors in KNN is set to be 3. An SVM is another widely used classifier in hyperspectral image processing; it has shown a good performance in the case of nonlinear separability and small training sample sets [20]. The optimal decision surface of an SVM is constructed from its support vectors with a maximum margin, which are conventionally determined by solving a quadratic programming problem. In our experiments, the kernel function of the SVM is Gaussian radial basis function, and the one-against-all scheme [37] is used for multiclass classification.

A. Indian Pines Data Set

To evaluate the performance of the proposed methods, three real data sets are used in the following experiments. The first hyperspectral image we used has been researched extensively because of its free availability on the internet. The image has $145 * 145$ pixels and 220 bands with a wavelength range from 400 to 2500 nm. In our experiments, bands 1–3, 103–112, 148–165, and 217–220 were removed due to atmospheric water vapor absorption and low signal-to-noise ratio [15], leaving a total of 185 valid bands to be used. From the 16 land-cover classes available in the original ground truth, seven classes can be removed because of a lack of sufficient samples [15]. Thus, for the remaining nine classes, we randomly choose 10% of the samples for each class to generate the training samples and the remainder are the test samples. The nine classes are listed in Table I.

1) Classification Performance Comparison: Just as recommended in [17] and [21], 15 bands are selected by different methods from the data set. Table II lists the 15 bands and Fig. 5 shows the overall classification accuracies using these bands. Because the OPBS method is theoretically equivalent to the MEV-SFS method, the bands obtained by them are the same, and Table II also verifies this point. Thence, the MEV-SFS method can be omitted in the comparison of classification accuracies. From Fig. 5, we can see that the classification based on the bands selected by MR and OPBS-ID achieves

TABLE II
15 BANDS SELECTED BY DIFFERENT METHODS FOR THE INDIAN PINES DATA SET

	Fifteen Bands														
LCMV-BCC	106	135	139	154	160	165	167	178	179	180	181	182	183	184	185
LCMV-BCM	103	104	106	107	110	117	120	140	143	154	160	165	168	178	179
MVPCA	18	19	20	21	22	23	24	25	26	27	28	29	30	38	39
MI	2	9	15	17	20	30	31	32	39	54	58	59	72	85	86
VGBS	10	15	17	20	26	29	31	32	36	54	58	59	72	85	86
MR	4	25	39	50	52	55	58	61	131	151	153	159	161	162	185
OPBS	10	15	17	20	26	29	31	32	39	54	58	71	72	85	86
OPBS-ID	10	32	38	54	81	84	85	88	89	93	101	116	143	168	183
MEV-SFS	10	15	17	20	26	29	31	32	39	54	58	71	72	85	86

TABLE III
CORRELATION OF 15 SELECTED BANDS AND COMPUTING TIME OF DIFFERENT METHODS FOR THE INDIAN PINES DATA SET

Methods	LCMV-BCC	LCMV-BCM	MVPCA	MI	VGBS	MR	MEV-SFS	OPBS	OPBS-ID
ACC	0.9816	0.9882	0.5950	0.2196	0.1995	0.2438	0.1815	0.1815	0.2769
Time (s)	1.27	1.97	0.08	15.11	2.39	3.41	2.12	0.16	0.67

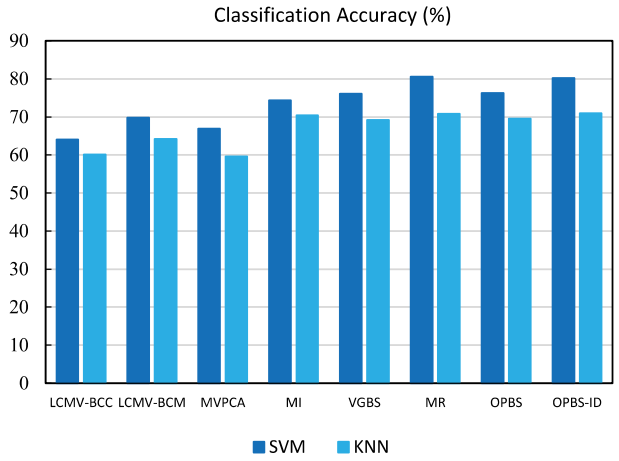


Fig. 5. Classification accuracy based on the 15 bands selected by different methods from the Indian Pines data set.

the highest accuracies. Followed by OPBS and VGBS, they obtain the similar classification accuracies, which are just slightly lower than those of OPBS-ID and MR. MI also shows a good performance, whereas the remaining three methods obtain relatively low accuracies. Specifically, when using the SVM classifier, MR and OPBS-ID have the highest accuracies of 80.64% and 80.19%, respectively. Followed by OPBS and VGBS, they perform almost the same, and the accuracies of them are 76.26% and 76.07%, respectively. MI also shows a good performance, and its accuracy equals 74.38%. As for MVPCA and the two LCMV-based methods, their accuracies are relatively lower when compared with other methods. The accuracies of MVPCA, LCMV-BCC, and LCMV-BCM are 66.97%, 64.08%, and 69.82%, respectively. For the KNN classifier, similarly, OPBS-ID and MR have the highest accuracies of 70.95% and 70.83%, respectively, followed by MI, which has an accuracy of 70.40%. As for OPBS and VGBS, they have the accuracies of 69.57% and 69.23%, respectively. Likewise, the remaining three methods perform not well, and the accuracies of MVPCA, LCMV-BCC, and LCMV-BCM are, respectively, 59.66%, 60.14%, and 64.22%. It is worth noting that OPBS and VGBS show almost the same performances in this experiment, and according to Table II, it can be found that most of the bands obtained by these two methods are the same. This occurs because they are both based on MEV, and the little differences between them are caused by different suboptimal searching strategies. In fact, more tests on other data sets further verify that the VGBS and OPBS methods have almost the same classification performances (see Sections V-B and V-C). Therefore, it can be considered that the OPBS method obtains the same results as the VGBS method in practice.

Finally, in this experiment, all the proposed methods show satisfactory classification performances: the classification accuracy of OPBS is almost the same with VGBS and their accuracies are just slightly lower than those of OPBS-ID and MR. As for OPBS-ID, it performs quite well, and its accuracy is quite close to that of MR and much higher than those of other methods. Therefore, this experiment demonstrates that the bands selected by the proposed methods can well represent the whole image tube and the proposed methods can obtain the classification accuracies which are comparable to the accuracies of other state-of-the-art methods.

2) Band Correlation Comparison: We compare the average correlation of the 15 bands obtained by these methods. The average correlation coefficient is used to evaluate the band correlation of a band set. The larger the value, the larger the band correlation. These results are listed in Table III, which indicates that the bands selected by MVPCA, LCMV-BCC, and LCMV-BCM are highly correlated. This occurs because most of the selected bands obtained by the MVPCA, LCMV-BCC, and LCMV-BCM methods are adjacent bands (see Table II), which are usually highly correlated with each other. In contrast, the other methods perform much better, and the bands obtained by OPBS (or MEV-SFS) have the lowest band correlation. As shown in Table II, most of the bands selected by the proposed methods are not adjacent bands; likewise, VGBS, MR, and MI also avoid selecting many adjacent bands, which demonstrate that these methods have taken band

correlation into consideration. From Fig. 5 and Table III, it can be found that the bands with high correlation usually result in low classification accuracies; for instance, the bands obtained by LCMV-BCC are highly correlated and thus obtain a relatively low classification accuracy; on the contrary, the bands obtained by the proposed methods have lower correlation and therefore achieve much better classification performances. This demonstrates that the information and band correlation should both be considered in BS. Therefore, the classification results and correlation comparison jointly demonstrate that the proposed methods can consider the information and band correlation simultaneously.

3) Computing Time Comparison: Furthermore, we compare the computing time of these methods, and the results are shown in Table III. Table III shows that the proposed methods run quite fast and only MVPCA is faster than OPBS and OPBS-ID. Considering that the OPBS and OPBS-ID can achieve much better classification accuracies than MVPCA, a little more time cost is acceptable. We also find that the computing time of OPBS and OPBS-ID is much less than that of MEV-SFS, which is consistent with the conclusions in Section IV-D. When compared with other methods, the computational cost of OPBS and OPBS-ID is still quite low. According to Table III, we can see that MEV-SFS and VGBS cost similar time. VGBS performs SVD to the whole data set and this process causes about NL^2 flops [for the data set with N pixels and L bands, the complexity of the SVD process is about $O(NL^2)$], so the theoretically computational complexity of VGBS is in fact close to that of MEV-SFS. Therefore, for the classification task, although OPBS and VGBS can obtain almost the same classification accuracies, the computational efficiency of OPBS is much better than that of VGBS, and this is the major superiority of the OPBS method relative to VGBS. As for the variant of OPBS, namely, the OPBS-ID method, it is not only faster but also more accurate than VGBS; therefore, OPBS-ID is superior to VGBS. MR costs more time than the methods above, because it includes clustering, clone selection, and MR, which are relatively time-consuming processes. As for MI, it computes the K-L divergences of all the band pairs in the data set, and the subset searching is also time-consuming, so the MI method costs the most time. Therefore, this experiment demonstrates that the three proposed methods have satisfactory computational efficiency, and they can find the desired bands in a reasonable time, especially the OPBS and OPBS-ID methods cost little time.

4) Recommended Number of Bands: Finally, if we use (41) as the stop criterion for the OPBS or OPBS-ID methods, the proposed methods are able to determine the suitable numbers of bands to be selected. For instance, we set the parameter ε as 0.0015, and the recommended number given by OPBS is 15 (see Fig. 6), which is exactly consistent with the recommended number in [17] and [21]. For the OPBS-ID method, the recommended number is equal to 7; to select more bands, the parameter needs to be set as a smaller value (e.g., 0.0001). Although how to determine the suitable value of ε requires further research, we can use the value ranges from 0.0001 to 0.0015 as the default value in most cases.

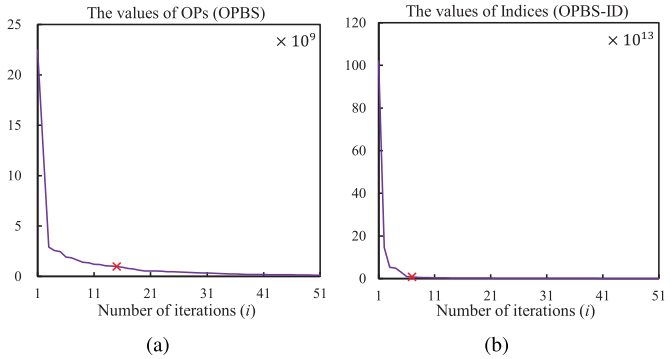


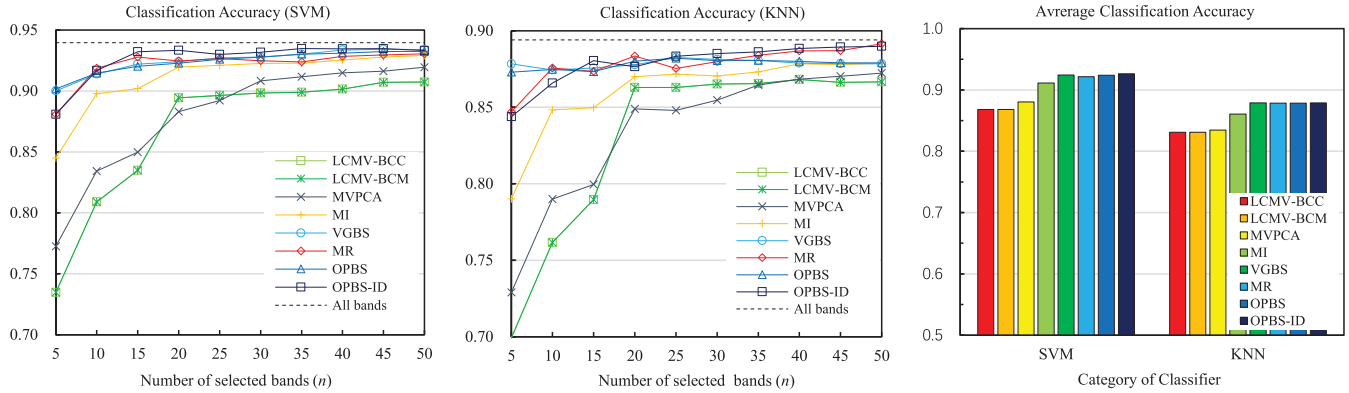
Fig. 6. Recommended numbers of selected bands for the Indian Pines data set. (a) OPs of the first 51 selected bands obtained by OPBS. (b) Indices of the first 51 selected bands obtained by OPBS-ID. Red point: recommended number.

B. Salinas Data Set

The second image is used to investigate the performance of the proposed methods in the case of selecting different numbers of bands. This data set was also collected by the 224-band AVIRIS sensor over Salinas Valley, CA, USA, and is characterized by a high spatial resolution (3.7-m pixels). The area covered comprises 512 lines × 217 samples. As with the Indian Pines scene, we discarded the 20 water absorption bands, which were the bands: 108–112, 154–167, and 224. In our experiments, all 16 classes in the Salinas data set are used, and likewise, 10% of the samples are randomly selected for training, and the rest are used for testing.

1) Classification Performance Comparison: In this experiment, the number of selected bands ranges from 5 to 50 and the classification results are shown in Fig. 7. The results of classification accuracy curves demonstrate that, for all the BS methods, the classification accuracies increase as the number of the selected bands increases. Among all the methods, the OPBS-ID shows the best overall performance, followed by MR, OPBS, and VGBS, and these methods also show comparable performances to that of OPBS-ID. MI also performs well, but the four methods above are always superior to it. For this data set, LCMV-BCC and LCMV-BCM show the same performances. Similar to the results on other data sets, the accuracies of the two LCMV-based methods and MVPCA cannot be comparable to those of the other state-of-the-art methods.

Specifically, when using the SVM classifier [see Fig. 7(a)], OPBS-ID obtains the highest accuracy, which can be competitive with the accuracy of using all bands. Followed by OPBS and VGBS, likewise, they perform almost the same, and the accuracies of them are just slightly lower than that of OPBS-ID. As for MR, it also shows a quite good performance; especially when selecting less than 20 bands, the accuracy of MR is very close to that of OPBS-ID. The MI method also performs well, but when selecting less than 20 bands, its accuracy cannot be competitive with the accuracies of the proposed methods, and MI requires to select more bands to obtain the accuracy that is close to the accuracies of the proposed methods. As for the other three methods, they perform not well, especially when selecting a small number of bands (e.g., 15 bands). On the other hand, when using the



(a) (b) (c)

Fig. 7. Classification results of the Salinas data set. (a) and (b) Accuracy curves by SVM and KNN, respectively. (c) Average accuracy bars.

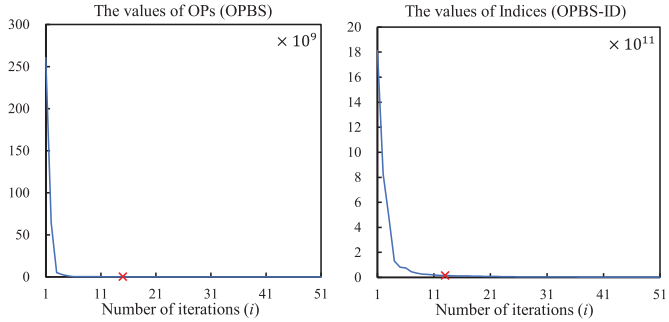
TABLE IV
CORRELATION OF 10 SELECTED BANDS AND COMPUTING TIME OF DIFFERENT METHODS FOR THE SALINAS DATA SET

Methods	LCMV-BCC	LCMV-BCM	MVPCA	MI	VGBS	MR	MEV-SFS	OPBS	OPBS-ID
ACC	0.8605	0.8605	0.9973	0.4823	0.4007	0.3132	0.3932	0.3932	0.2594
Time (s)	13.67	15.51	0.86	137.67	20.19	37.25	25.83	1.43	4.297

KNN classifier [see Fig. 7(b)], OPBS-ID and MR achieve the best performances and they are slightly superior to VGBS and OPBS; the superiority becomes more significant when more bands are selected from the data set. MI performs well, but its accuracy is slightly lower than those of the proposed methods, and likewise, the LCMV-based methods and MVPCA perform not well. Fig. 7(c) further shows the average classification accuracies of selecting different numbers of bands [the averages of the curves in Fig. 7(a) and (b)]. It is clear that the proposed methods achieve the performances that can be comparable to those of the state-of-the-art methods such as VGBS and MR; especially, the OPBS-ID method is even slightly better than them. Therefore, the experimental results further demonstrate that the proposed methods can achieve satisfactory performances for classification.

2) Band Correlation Comparison: Table IV shows the average band correlation of the 10 bands selected by different methods. Obviously, the bands selected by MVPCA, LCMV-BCC, and LCMV-BCM are highly correlated; in contrast, the bands obtained by other methods have much lower average correlation, which demonstrates that these methods have paid sufficient attention on the band correlation. Among all the methods, the bands obtained by OPBS-ID have the lowest band correlation; correspondingly, the OPBS-ID method achieves a relatively better classification performance on this data set. Similar to the results on Indian Pines data set, the bands with high correlation are generally corresponding to the low classification performances (see Fig. 7 and Table IV), which further indicate that a good BS method should consider both the information and band correlation. Therefore, this experiment further indicates that the proposed methods can consider the information and band correlation simultaneously.

3) Computing Time Comparison: The computing time of selecting 10 bands by different methods is compared in this experiment. The results are also listed in Table IV, which indicates that the MVPCA method costs the least time, followed by OPBS, OPBS-ID, the LCMV-based methods, VGBS, and MEV-SFS, whereas MR and MI require more time. Although MVPCA runs the fast, the computing time of OPBS is close to that of MVPCA. When compared with the competitors except MVPCA, the time cost of OPBS and OPBS-ID is quite low. For instance, the computing time of OPBS-ID is only about 11.5% of that of MR. Therefore, this experiment further indicates that the OPBS and OPBS-ID methods have high computational efficiency, run fast, and cost little time.



(a) (b)

Fig. 8. Recommended numbers of selected bands for the Salinas data set. (a) OPs of the first 51 selected bands obtained by OPBS. (b) Indices of the first 51 selected bands obtained by OPBS-ID. Red point: recommended number.

4) Recommended Number of Bands: The recommended numbers of selected bands are also shown in Fig. 8. The parameter ε for the OPBS and OPBS-ID methods is set as 0.0001 and 0.0015, respectively; then the BS algorithms are terminated at the 15th and 13th rounds, respectively, and thus, the recommended numbers are equal to 15 and 13. Although it is difficult to know how many bands should be selected in practice, a basic rule is that more bands need to be selected if an image scene is complicated and contains many classes, because the data dimensionality should be high enough to accommodate these classes for detection or classification. Therefore, in most cases, it is suitable to choose the number of bands close to the number of classes. Coincidently,

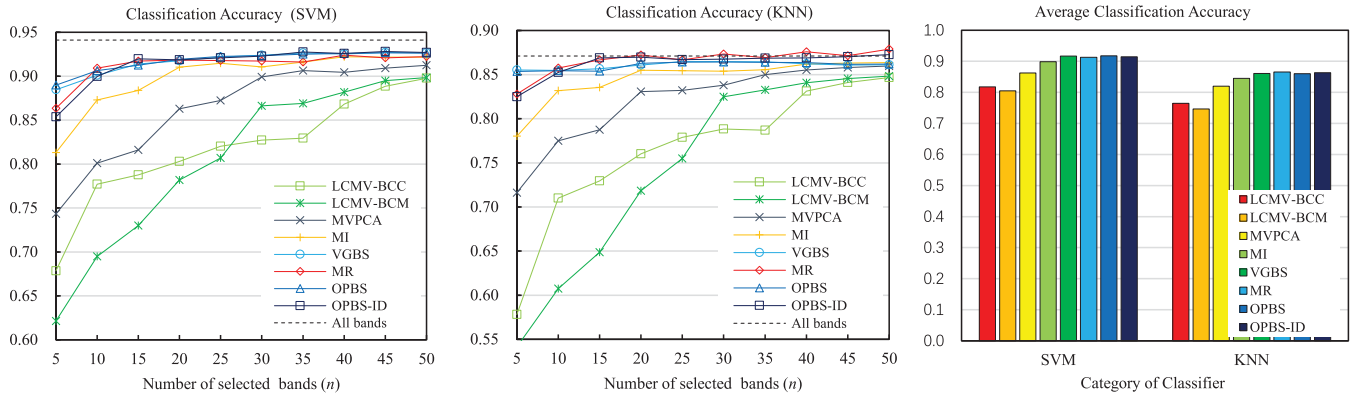


Fig. 9. Classification results of the Pavia University data set. (a) and (b) Accuracy curves by SVM and KNN, respectively. (c) Average accuracy bars.

the recommended numbers (15 and 13) for the OPBS and OPBS-ID methods are very close to the number of the classes in the Salinas data set (16 classes). Thus, the recommended number can be regarded as a reference value for the number of the bands to be selected.

C. Pavia University Image

The third image is a hyperspectral image at the University of Pavia acquired by the ROSIS-3 optical sensor. The data set has 103 spectral bands with a spectral range from 0.43 to 0.86 nm. The image size is 610×340 with a spatial resolution of about 1.3 m. In the image, nine classes are labeled: Asphalt, Meadows, Gravel, Trees, Painted Metal Sheets, Bare Soil, Bitumen, Self-Blocking Bricks, and Shadows [38]. For this data set, we also randomly choose 10% pixels for training and the rest for testing.

1) Classification Performance Comparison: Likewise, the number of the selected bands ranges from 5 to 50, and the overall classification accuracies of the selected bands are shown in Fig. 9. It can be observed that, for this data set, OPBS, OPBS-ID, VGBS, and MR show quite similar overall classification performances, followed by MI, MVPCA, LCMV-BCC, and LCMV-BCM. When using the SVM classifier [see Fig. 9(a)], OPBS, VGBS, and OPBS-ID show the best performances, which are quite similar to each other; especially the OPBS and VGBS methods perform almost the same. MR also achieves a good performance, which is very close to the performances of the proposed methods. The accuracy of MI is a little lower but is still much higher than those of MVPCA, LCMV-BCC, and LCMV-BCM. As for the KNN classifier [see Fig. 9(b)], results are a little different. MR and OPBS-ID perform the best, and they are slightly superior to OPBS and VGBS. And likewise, VGBS and OPBS perform almost the same and the accuracies of them are close to those of OPBS-ID and MR. Followed by MI, it also performs well, whereas the remaining three methods perform not well. Additionally, Fig. 9(c) shows the average accuracies of these methods. It demonstrates that the performances of the proposed methods can be comparable to the performances of the state-of-the-art methods such as MR and VGBS. For the SVM, OPBS and VGBS achieve the best overall performances, whereas for the KNN, OPBS-ID and MR achieve the best overall performances. Therefore, this experiment again verifies that the proposed methods can obtain the satisfactory classification performances, which are comparable to the performances of the state-of-the-art methods.

2) Band Correlation Comparison: The average band correlation of the 10 bands selected by different methods is listed in Table V. Similarly, for this data set, the LCMV-based methods and the MVPCA method obtain the bands with high correlation, and thus, they perform not well in the classification experiment. On the contrary, the bands selected by other methods are lowly correlated; therefore, they achieve much better classification performances. This experiment again verifies that the bands selected by the proposed methods are lowly correlated.

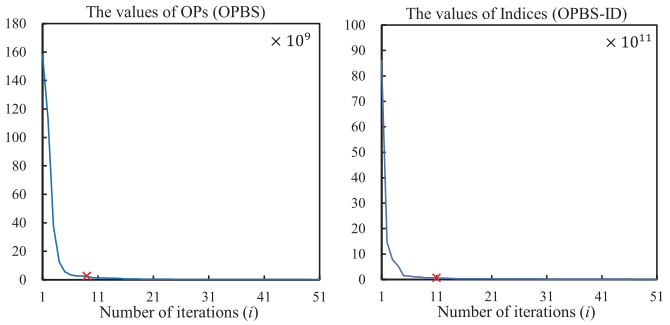
3) Computing Time Comparison: Also, the computing time of selecting 10 bands by different methods is shown in Table V. Likewise, it indicates that the time cost of the OPBS and OPBS-ID is much lower than that of the competitors except MVPCA, and we also notice that MEV-SFS also can select the desired bands in a reasonable time. Therefore, all the proposed methods have satisfactory computational efficiency; especially, the OPBS and OPBS-ID have quite low computational cost.

4) Recommended Number of Selected Bands: The recommended number of selected bands is also shown in Fig. 10. The threshold ε is set to be the default value of 0.0015 and the OPBS and OPBS-ID methods are terminated at the 9th and 11th rounds, respectively. We can find that the recommended numbers (9 and 11) for the OPBS and OPBS-ID methods are very close to the number of the classes in the Pavia data set (nine classes). Thus, the recommended number can be regarded as a reference value for the number of the bands to be selected.

From all the experiments on three different hyperspectral data sets, some important results can be summarized. In BS, the discriminative capability of each individual band and the correlation among them should both be considered, and our experiments demonstrate that the bands with low correlation usually result in good classification accuracies. The proposed methods have taken both the information and band correlation into consideration, and it is verified that the bands selected by them are informative for classification as well as lowly correlated with each other. Among the three proposed methods,

TABLE V
CORRELATION OF 10 SELECTED BANDS AND COMPUTING TIME OF DIFFERENT METHODS FOR THE PAVIA UNIVERSITY DATA SET

Methods	LCMV-BCC	LCMV-BCM	MVPCA	MI	VGBS	MR	MEV-SFS	OPBS	OPBS-ID
ACC	0.9879	0.9934	0.9980	0.6266	0.4750	0.6145	0.5267	0.5267	0.6396
Time (s)	11.47	12.59	0.75	72.41	12.96	22.24	13.28	1.31	4.25



(a) (b)

Fig. 10. Recommended numbers of selected bands for the Pavia University data set. (a) OPs of the first 51 selected bands obtained by OPBS. (b) Indices of the first 51 selected bands obtained by OPBS-ID. Red point: recommended number.

the OPBS method is the fast version of the MEV-SFS method; they are theoretically equivalent and yield the identical bands in practice. OPBS can also be regarded as a model for BS, and OPBS-ID is one of the variants of OPBS. In most cases, the OPBS-ID method is slightly superior to the OPBS method for classification, but the computing time of OPBS is always less than that of OPBS-ID. OPBS and OPBS-ID both have much better computational efficiency than MEV-SFS. When compared with other BS methods, all the proposed methods show satisfactory classification performances, which can be even competitive with the performances of the state-of-the-art methods like MR and VGBS. In fact, OPBS and VGBS perform almost the same; OPBS-ID is superior to VGBS, and both OPBS and OPSB-ID have a better computational efficiency than VGBS. Because MR is an advanced BS method based on several efficient machine learning algorithms, it shows amazingly good performances in our experiments, but the performance of OPBS-ID can be even comparable with that of MR, which indicates that the OPBS-ID method is also a highly efficient method. Additionally, the proposed methods always produce good and stable performance of classification in any hyperspectral data sets, which demonstrates that the proposed methods have strong robustness. As for the computational efficiency, owing to the use of recursive formulas, the OPBS and OPBS-ID methods achieve an incremental calculation of the objective functions and thus have low computational complexity, and the experimental results verify that the computational efficiency of them is much better than that of the similar methods considering band correlation, e.g., MI and MR. Therefore, the OPBS and OPBS-ID methods can obtain the desired bands in a quite short time. Finally, these experiments also verify that the OPBS and OPBS-ID have the potential to decide the number of the bands to be selected, and the recommended number can be regarded as a reference value of the bands to be selected. In conclusion, the effectiveness of the proposed methods has been proved.

VI. CONCLUSION

In this paper, we developed new unsupervised BS algorithms for the hyperspectral image analysis. MEV is a perfect selection criterion for BS, and by combined with SFS, it can be used for hyperspectral imagery and the new method is denoted as MEV-SFS. MEV-SFS iteratively added the band that maximizes the ellipsoid volume of selected bands into the selected band set, until the desired number of bands has been obtained. Furthermore, based on the relationship between the ellipsoid volume of the selected band set and the OPs of candidate bands, another equivalent method named OPBS is proposed. OPBS is the fast version of MEV-SFS; it tries to select the band that has the maximum OP as the desired band. Because OPs can be incrementally calculated, OPBS has a significantly reduced computational complexity compared with MEV-SFS. Moreover, theoretical analysis demonstrates that OPBS is also a framework for BS, so we further propose a third method, namely, the OPBS-ID method. OPBS-ID resorts to find the band with the largest amount of useful information, and it can achieve a better classification performance than OPBS in many cases. The advantages of these proposed methods are summarized as follows.

1) The three proposed methods can find the bands that are informative and lowly correlated and therefore can achieve good classification performances.
2) The OPBS and OPBS-ID methods have high computational efficiency, so they can find the desired bands in a short time.
3) The OPBS method has definite physical meaning and can be regarded as a model or framework for BS. It is easy to obtain the variants of OPBS.
4) OPBS and its variants (e.g., OPBS-ID) have the potential to determine how many bands should be selected. The recommended number has a reference significance for determining the number of the selected bands.

Finally, the experimental results on different real hyperspectral images have verified that the proposed methods are highly efficient and accurate unsupervised BS methods. Our future research interest is the adaptive determination of the parameter ε.

ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their detailed comments and valuable suggestions, which helped us to improve this paper.

REFERENCES

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
[2] D. Landgrebe, *On Information Extraction Principles for Hyperspectral Data*, vol. 34. West Lafayette, IN, USA: Purdue Univ., 1997.
[3] I. T. Jolliffe, "Principal component analysis and factor analysis," in *Principal Component Analysis*. New York, NY, USA: Springer, 1986, pp. 115–128.

[4] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
[5] A. Hyvärinen, J. Hurri, and P. O. Hoyer, "Independent component analysis," in *Natural Image Statistics*. London, U.K.: Springer, 2009, pp. 151–175.
[6] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[7] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-26, no. 1, pp. 65–74, Jan. 1988.
[8] J. Zhang, Y. Cao, L. Zhuo, C. Wang, and Q. Zhou, "Improved band similarity-based hyperspectral imagery band selection for target detection," *J. Appl. Remote Sens.*, vol. 9, no. 1, p. 095091, 2015.
[9] A. J. Brown, "Spectral curve fitting for automatic hyperspectral data analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1601–1608, Jun. 2006.
[10] 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.
[11] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. 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.
[12] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, Oct. 2006.
[13] P. Chavez, G. L. Berlin, and L. B. Sowers, "Statistical method for selecting Landsat MSS," *J. Appl. Photogr. Eng*, vol. 8, no. 1, pp. 23–30, 1982.
[14] C. Sheffield, "Selecting band combinations from multispectral data," *Photogrammetric Eng. Remote Sens.*, vol. 51, pp. 681–687, 1985.
[15] X.-S. Liu, L. Ge, B. Wang, and L.-M. Zhang, "An unsupervised band selection algorithm for hyperspectral imagery based on maximal information," *J. Infr. Millim. Waves*, vol. 31, no. 2, pp. 166–176, 2012.
[16] J. M. Sotoca, F. Pla, and J. S. Sánchez, "Band selection in multispectral images by minimization of dependent information," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 2, pp. 258–267, Mar. 2007.
[17] X. Geng, K. Sun, L. Ji, and Y. Zhao, "A fast volume-gradient-based band selection method for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7111–7119, Nov. 2014.
[18] L. Wang, X. Jia, and Y. Zhang, "A novel geometry-based feature-selection technique for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 4, no. 1, pp. 171–175, Jan. 2007.
[19] A. Zare and P. Gader, "Hyperspectral band selection and endmember detection using sparsity promoting priors," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 256–260, Apr. 2008.
[20] P. Mitra, C. A. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 301–312, Mar. 2002.
[21] Y. Qian, F. Yao, and S. Jia, "Band selection for hyperspectral imagery using affinity propagation," *IET Comput. Vis.*, vol. 3, no. 4, pp. 213–222, Dec. 2009.
[22] K. Sun, X. Geng, and L. Ji, "Exemplar component analysis: A fast band selection method for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 998–1002, May 2015.
[23] M. Ahmad, D. I. U. Haq, Q. Mushtaq, and M. Sohaib, "A new statistical approach for band clustering and band selection using K-means clustering," *IACSIT Int. J. Eng. Technol.*, vol. 3, no. 6, pp. 606–614, Dec. 2011.
[24] 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.
[25] Y. Yuan, G. Zhu, and Q. Wang, "Hyperspectral band selection by multitask sparsity pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 631–644, Feb. 2015.
[26] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral band selection by discovering diverse subset in multiple graphs," *IEEE Trans. Image Process.*, to be published.
[27] P. Pudil, F. J. Ferri, J. Novovicova, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proc. 12th IAPR Int. Conf. Pattern Recognit., Conf., B, Comput. Vis. Image Process.*, vol. 2. Oct. 1994, pp. 279–283.
[28] C.-I. Chang, C.-C. Wu, W.-M. Liu, and Y.-C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2804–2819, Oct. 2006.
[29] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.
[30] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. San Diego, CA, USA: Academic, 2013.
[31] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Trans. Comput.*, vol. C-26, no. 9, pp. 917–922, Sep. 1977.
[32] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Optimization*, vol. 7. Hoboken, NJ, USA: Wiley, 2000.
[33] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1360–1367, Jul. 2001.
[34] X.-D. Zhang, *Matrix Analysis and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2017.
[35] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*, vol. 31. Springer, 2013.
[36] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[37] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *J. Mach. Learn. Res.*, vol. 5, pp. 101–141, Jan. 2004.
[38] 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.

Wenqiang Zhang received the B.S. degree in automation from Zhejiang University, Hangzhou, China, where he is currently pursuing the Ph.D. degree in control theory and control engineering.

His research interests include image processing and machine learning.

Xiaorun Li received the B.S. degree from the National University of Defense Technology, Changsha, China, in 1992, and the M.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 1995 and 2008, respectively.

Since 1995, he has been with Zhejiang University, where he is currently a Professor with the College of Electrical Engineering. His research interests include hyperspectral image processing, signal and image processing, and pattern recognition.

Yaxing Dou, photograph and biography not available at the time of publication.

Liaoying Zhao received the B.S. and M.S. degrees from Hangzhou Dianzi University, Hangzhou, China, in 1992 and 1995, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, in 2004.

Since 1995, she has been with Hangzhou Dianzi University, where she is currently a Professor with the College of Computer Science. Her research interests include hyperspectral image processing, signal and image processing, pattern recognition, and machine learning.