Received January 17, 2020, accepted January 28, 2020, date of publication February 4, 2020, date of current version March 6, 2020.

Regularized Sparse Band Selection via Learned Pairwise Agreement

ZHI XI FENG\textsuperscript{1,2}, (Member, IEEE), SHUYUAN YANG\textsuperscript{1,2}, XIAOLONG WEI\textsuperscript{3}, QUAN WEI GAO\textsuperscript{1,2}, (Senior Member, IEEE), AND LI CHENG JIAO\textsuperscript{1,2}, (Fellow, IEEE)

\textsuperscript{1}School of Artificial Intelligence, Xidian University, Xi’an 710071, China
\textsuperscript{2}Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi’an 710071, China
\textsuperscript{3}Science and Technology on Plasma Dynamics Laboratory, Air Force Engineering University, Xi’an 710038, China

Corresponding author: Shuyuan Yang (syyang@xidian.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61906145, Grant 61771376, and Grant 61771380, in part by the National Natural Science Foundation of Shaanxi Province under Grant 2019JQ-663, and in part by the China Postdoctoral Science Foundation under Grant 2018M643587.

ABSTRACT Desired by sparse subset learning, in this paper, a hyperspectral band selection method via pairwise band agreement with spatial-spectral graph regularizer, referred as Regularized Band Selection via Learned Pairwise Agreement (RBS-LPA), was proposed. The process was formulated as a graph-regularized row-sparse constrained optimization problem, which select a few representative bands to code the all bands based on the learned pairwise band agreement. In RBS-LPA, a spatial-spectral informative graph, constructed by spatial-spectral neighbor relationship, is incorporated to encode both the spatial and spectral geometrical structure. By combining the learning procedure with graph regularizer jointly, the graph regularizer can adaptively change with the sparse representative bands, which ensures that the selected bands can well preserve the local manifold structure. Experimental results on three real urban hyperspectral data demonstrate the efficiency of the proposed RBS-LPA and achieve convincing performance.

INDEX TERMS Feature selection, self-representative learning, sparse optimization, graph regularizer.

I. INTRODUCTION

The advancement in remote sensors has made hyperspectral data collection much easier and more informative than ever before, which makes hyperspectral imaging become one of the most promising techniques for various successful applications, such as precision agriculture, environmental science, and military applications [1], [2]. Consisting of two spatial dimensionality and one high resolution spectral dimensionality, a hyperspectral image cube provides sufficiently valuable observations at hundreds of contiguous and very narrow spectral bands. Intuitively, having more spectral bands implies more discriminative power in classification and recognition [3], [4]. However, this is not always work in practical application, because not all the features presented in high-dimensional data set help in class prediction. The presence of irrelevant and redundant features not only increase the computational cost and memory requirement but also be possibly detrimental to their performance due to the noise effect and insufficient number of samples with respect to the ambient space dimension, commonly referred to as the “curse of dimensionality”, to describe the problems faced in the analysis of high dimensional data [5]–[8].

Finding low-dimensional structures in the data not only helps to reduce the computational cost and memory requirements of algorithms, but also reduce the side effects of noise in the high-dimensional data and improve the performance of classification tasks [9]. High-dimensional data sets are inherently sparse and hence, can be projected into lower dimension space without losing too much discriminative information, and the projection operation is so called dimensionality reduction (DR). In [10], Bioucas-Dias, Josíe \textsuperscript{1} M., et al. said that the high spectral resolution of hyperspectral data with original spectral features contains high redundancy. Specially, there is a high correlation between adjacent bands and the number of the original spectral features may be too high for classification purposes. In addition, the original spectral features may not be the most effective ones to separate the objects of interest from others. Therefore, dimensionality reduction (DR) techniques, often performed prior in some practical hyperspectral tasks, which reduce
the redundancy in original subspace while preserving the informative and representative information about objects for specific purpose, have been playing a paramountly important role in both researches and applications. Generally, DR techniques transform the original high-dimensional data into a smaller and more manageable representation with reduced dimensionality. Algorithms that have been developed can be broadly categorized into two groups: feature extraction and feature selection. Feature extraction approaches create a new set of features from the original feature space by transformation. Among them, principal component analysis (PCA) [11] and linear discriminant analysis (LDA) [11], are two well-known linear feature extraction algorithms which are widely used. While feature selection methods select, from the original feature set, an informative, compact and accurate subset features, which are highly effective in discriminating classes. Like FAST in [12] and Ranking based sparse SVM in [13]. Compared with feature extraction methods, feature selection not only has an easy implementation but also preserve the original semantic and physical property of the features, thereby offering the advantage of interpretability by a domain expert.

Feature selection can be conducted in a supervised or unsupervised mode, in term of whether the label information is utilized to guide the selection process [14]. Generally, supervised feature selection methods require a large amount of label information, in real hyperspectral scenario, supervised information is expensive to obtain. In this paper, we only consider unsupervised feature selection algorithms. Unsupervised feature selection is of great value where there is no prior knowledge of phenomena present. Recently, many unsupervised hyperspectral band selection methods have been proposed to support efficient practical application [15]. Existing techniques can be broadly classified into five categories: ranking based [16]–[18], clustering based [19], [20], learning based [21], evolution selection based [22], and information metric based [23] in these scenes. The main idea of ranking based methods is to find the most distinctive and informative bands. Various ranking based methods for unsupervised band selection are present in the literature, like information divergences (ID) [17], maximum variance based principal component analysis (MVPCA) [16], and similarity-based band selection method [18]. Clustering based methods [19] perform clustering operation on bands (treated as patterns) to group them according to their correlation, and select one band from each cluster representing the whole group. A few clustering based band selection techniques for hyperspectral images exist in the literature e.g., Ward’s linkage strategy using divergence (WaLuDi) [19], or using mutual information (WaLuMI) [19], and band selection using affinity propagation (AP) [20]. In addition, learning based method [15], [21], multi-task sparsity pursuit (MTSP) learning framework is introduced to improve the selection performance. Also, some sparsity inspired feature learning methods, which assume the features can be expressed as a linear combination of only some representative features, which formulates FS as a joint sparse recovery problem. and Li et al. proposed a sparse representation-based band selection method (SpaBS), which assumes that the interested bands contain the maximum information [21]. These algorithms are not well work where features lie in nonlinear models, evolution selection based [22], and information metric based [24].

Despite that many unsupervised band selection approaches have been proposed for HIC, there are still some issues to be further studied. First, it is highly demanded that the pairwise samples relationship should be preserved in the selection procedure; Second, spatial information buried in the hyperspectral image should be utilized, which can be easily explored by incorporate a local smoothness regularization, that is the spatial-spectral neighbor pixels have similar representative subset bands.

In this paper, we propose an unsupervised band selection algorithm to exploit the geometrical structure information in the image by using the manifold regularizer and pairwise agreement, referred to as Regularized Band Selection via Learned Pairwise Band Agreement (RBS-LPA). The pairwise agreement between each band can be computed by using a certain metric, such as Euclidean distance, mutual information, Kullback-Leibler divergence etc., or by a learning processing. In our work, the pairwise agreement defined as the refined low-rank representation coefficients with spatial relative position constraints. Using pairwise agreement to encode the original bands relationships, the band selection process can be implemented by solving a row-sparsity with graph regularizer constrained trace minimization problem by using Alternating Direction Method of Multipliers (ADMM) [25]. The key idea behind the proposed algorithm is what so-called self-expressiveness property of the data [26] that is we can use few informative bands to well encode the original all bands. Thus we cast the assumption that the original bands can be expressed as a linear combination of some representative and informative bands, and then it can be formulated as a sparse optimal problem that finding few representative bands from the original band. Meanwhile, the geometrical information of image is important for discrimination [27]. And the geometric structure information can be exploited by using the manifold assumption, which has been shown effective in classification and clustering tasks [28]–[30]. Thus geometrical spatial-spectral graph regularizer is introduced in this paper to exploit geometrical local invariant structure of the hyperspectral data. [26], i.e. the nearby pixels in the original feature space are likely to have similar embedding, to formulate a pixel-wise graph regularizer. We define an undirected graph to reveal the spatial-spectral geometric structure.

The contribution of the proposed method can be summarized as:

1. The RBS-LPA improves from sparse representation and it allows one to consider models beyond linear subspaces, it does not enforce the hyperspectral bands sampled from some underlying independent subspace.
2. The learned pairwise agreement via refined low-rank representation with spatial relative position can well encode
the latent geometric structure of the spatial-spectral bands and also robust to band noises.

(3) The proposed algorithm utilizes both the sparse learning mechanism and manifold structured regularizer, thus the graph spectral vectors and sparse responsibility matrix are jointly learned at the same time, the graph spectral vectors can be adaptively changed with the sparse responsibility matrix, and the selected representative bands are reselected to preserve the graph structure.

(4) By incorporating the geometrical information of the data space via constructing a neighbor graph regularizer, we ensure that the spatial spectral neighbors have similar representative band subset, which explicitly considers the local invariance.

The remainder of this paper is organized as follows. Section II first describes the works that related to representative selection and then describes the details of the proposed RBS-LPA. Section III evaluates the effectiveness of the proposed method by experimenting on some real urban hyperspectral dataset. The conclusion and discussion are drawn in Section IV.

II. THE REGULARIZED SPARSE BAND SELECTION VIA LEARNED PAIRWISE AGREEMENT

Given a hyperspectral data set \( X = \{x_1, x_2, \ldots, x_N\} \in R^{d \times N} \) consisting of \( N \) instances and features such that \( x_i \in R^d \) each vector \( x_i \) is, corresponding to a real hyperspectral pixel, a \( d \) dimensional vector of band features, hence \( x_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T \). And \( x^k \) is a \( N \) dimensionality vector that denotes the \( k-th \) row \( X \) corresponding to the band of the given hyperspectral data. And \( X \in R^{d \times N} \) denotes the whole data set. Without loss of generality, feature selection can be formulated as:

\[
s_i = S \left( W^T x_i \right)
\]

where \( W \in R^{d \times d} \) is a transform matrix, and force \( W \) is a transform matrix, and force \( W \) equal zeroes. And \( S(\cdot) \) denotes a selection operator that select the nonzero entries of input, that is to say, the features/bands that correspond to the indices of non-zero rows of \( W \) are selected. So, the dimension of \( s_i \) equals to the number of nonzero rows of \( W \), i.e. the dimensionality of the low-dimensional space.

A. A BRIEF REVIEW OF RELATED WORKS ON SPARSE REPRESENTATIVE SELECTION BASED ON PAIRWISE AGREEMENT

The performance of machine learning methods is heavily dependent on the choice of data representation (or features) on which they are applied. A good data representation should be compact, representative and informative. Inspired by the idea of DS3 [26], [31], we cast the assumption that the bands of pixels can be well represented only by a few ‘informative and representative’ bands. Thus the informative bands can form a compact representation of the complete original bands. Given hyperspectral data set \( X \in R^{d \times N} \), it is easy to construct a nonnegative pairwise agreement metric \( A_{ij} \) \( i=1, \ldots, d \) between the bands. Each \( A_{ij} \) indicates how well \( i-th \) band to represents \( j-th \) band. i.e., the larger value of \( A_{ij} \), the better \( x^i \) encode \( x^j \). The pairwise agreement can be arranged into a matrix form as:

\[
A = \begin{bmatrix}
A_{11} & A_{12} & \cdots & A_{1d} \\
A_{21} & A_{22} & \cdots & A_{2d} \\
\vdots & \vdots & \ddots & \vdots \\
A_{d1} & A_{d2} & \cdots & A_{dd}
\end{bmatrix} \in R^{d \times d}
\]

Given agreement metric \( A \), the goal of RBS-LPA is to find a small representative and informative feature subset of \( X \) that well represents the originally entire \( X \). Agreement metric can be computed according to a predefined metric measurement function, such as Euclidean distance or by a certain learning criteria. In our work, the agreement measure is obtained by low rank representation coefficients refined by spatial irrelevant relationship, the reason lies in the fact that the hyperspectral pixels are usually mixed due to the low spatial resolution and noise. To better handle the mixed hyperspectral data, here we lesson a more general framework “nonnegative Low Rank Representation (LRR)” to obtain the pairwise relationship between each band.

\[
\min_{D,E} \text{rank}(A) + \alpha \|E\|_1 \quad s.t. \; X^T = X^TA + E, \; A \succeq 0
\]

where \( X^T \) is a “dictionary” that linearly spans the data space. The optimal representation \( A \) can be obtained via Alternating Direction Method of Multipliers, which can be regarded as the optimal pairwise representation of bands. Meanwhile, the neighbor bands usually have similar response, thus we modified the relationship by using the sigmoid function. Mathematically, \( D_{ij} = A_{ij} \cdot \frac{-1}{1+e^{-\frac{-D_{ij}}{\beta}}} \). Fig.1 shows the spatial disagree weight with respect to spatial relative position.

We consider an optimization program with responsibility variable \( Z_{ij} \) associates with agreement \( D_{ij} \). We denote the matrix of all variables as:

\[
Z = \begin{bmatrix}
Z_{11} & Z_{12} & \cdots & Z_{1d} \\
Z_{21} & Z_{22} & \cdots & Z_{2d} \\
\vdots & \vdots & \ddots & \vdots \\
Z_{d1} & Z_{d2} & \cdots & Z_{dd}
\end{bmatrix} \in R^{d \times d}
\]
Z_{ij} is regarded as the responsibility of x^i being a representative band for x^j band and restrict Z_{ij} ∈ [0, 1]. Notice that x^i can have multiple representative bands, in which case ∑_{i=1}^N Z_{ij} > 0 for every x^i that represent x^j. We further restrict ∑_{j=1}^d Z_{ij} = 1 to ensure that the probability of x^i being represented by X is equal to 1. The problem of selecting a few bands to represent the all bands according to the agreements can be formulated by a row-sparsity regularized trace minimization problem on Z. First, selecting representative bands that can well encode X based on the agreement. If x^i being the a representative band for x^j band with responsibility Z_{ij}, then the cost of encoding x^i is −D_{ij}Z_{ij}. Hence the total cost of encoding x^i using its all representative bands is − ∑_{j=1}^d ∑_{i=1}^d D_{ij}Z_{ij}. Second, we would like to as few as bands of X to represent Y. It can be easily found that if the row of Z equal to 0, the index of the feature does not contribute to coding the output. Putting these two goals together, the selection procedure can be formulated as the following optimization problem:

$$\min_{\{Z_{ij}\}} \lambda \sum_{i=1}^d \|Z_i^T\|_q - \sum_{j=1}^d \sum_{i=1}^d D_{ij}Z_{ij}$$

s.t. \sum_{i=1}^d Z_{ij} = 1, Z_{ij} ≥ 0

where \|Z_i^T\|_q denotes the l_q norm and I() denotes the indicator function $I(x) = \begin{cases} 0, & x = 0 \\ 1, & \text{others} \end{cases}$, Z^i denotes the i-th row of Z. The first term in (3) counts to the number of selected bands and the second term corresponds to the total cost of encoding X via the selected representative bands. The parameter λ sets the trade-off between the two terms. Since the minimization of Eq.(3), is NP-hard, we consider the following convex relaxation formulation:

$$\min_{\{Z_{ij}\}} \lambda \sum_{i=1}^d \|Z_i^T\|_q - \sum_{j=1}^d \sum_{i=1}^d D_{ij}Z_{ij}$$

s.t. \sum_{i=1}^d Z_{ij} = 1, Z_{ij} ≥ 0

We rewrite the above formulation as matrix form:

$$\min_{Z} \lambda \|Z\|_{1,q} - \text{tr}(D^T Z)$$

s.t. $I^T Z = I^T, Z ≥ 0$  

where \|Z\|_{1,q} = \sum_{i=1}^d \|Z_i^T\|_q and I denotes a column vector with all elements equal 1. Once we get the optimal value, we can select the representative bands indices from the nonzero rows of solution Z^+. By optimizing the above objective function Eq.(5), we can obtain a representative feature subset that can well approximate the original feature space.

**B. REGULARIZED BAND SELECTION VIA LEARNED PAIRWISE BAND AGREEMENT (RBS-LPA)**

The above optimization Eq.(5) fails to discover the intrinsic geometrical structure of the image data space, which is essential to the real-world hyperspectral application. One might further hope that the selected features can well preserve the intrinsic spatial structure of the hyperspectral cube. It is assume that if two pixels $x_i, x_j$ are close in the original feature space, $s_i, s_j$ the representation of these two pixels with respect to the new feature space, are also close to each other. This assumption is usually referred to as local smoothness assumption [32]. Recent studies on manifold learning theory have demonstrated that the local geometric structure can be effectively modeled through a neighbor relationship graph on a scatter of data points. Consider a graph with N vertices, which each vertex corresponds to an instance/pixels. For each pixels $x_i$, we find its $p$ spatial-spectral nearest neighbors and add an edge between $x_i$ and its neighbors. In our works, we set $15 \times 15$ since the experiments reveal that the number of window size does have little effect on the final classification, and $p$ equals 6. Figure 2 shows the spatial-spectral neighbor relationship. There are many choices to define the weight matrix G on the graph. Most commonly used are 0-1 Weighting, Gaussian Kernel Weighting and Dot-Product Weighting. In this proposed framework, we use Euclidean distance to measure the neighbor relationship. Then we define the following regularized term:

$$R_1 = \frac{1}{2} \sum_{i,j=1}^N \|s_i - s_j\|^2 \delta_{ij} = \sum_{i=1}^N s_i^T Q s_i - \sum_{i,j} s_i^T s_j \delta_{ij}$$

where $\delta_{ij} = \begin{cases} 1, & x_i \in N_K(x_j) or x_j \in N_K(x_i) \\ 0, & \text{others} \end{cases}$

and $\delta_{ij}$ denotes the spatial Euclidian distance between $x_i$ and $x_j$, where Tr () denotes the trace of a matrix and is the degree matrix of G, which is a diagonal matrix whose entries are the column sum of $G$. $Q = \sum_i \delta_{ij}$ and $L = Q - G$ is the Laplician matrix of a graph. Minimizing the regular term,

![Figure 2](image-url)
we expect that if two pixels $x_i$ and $x_j$ are close, their embedding $s_i$, $s_j$ are also close to each other. Combining this geometric-based regularizer with DRS, the object function can be formulated as:

$$\min_{Z} \frac{\lambda}{2} \|Z\|_{1,q} - tr \left( D^T Z \right) + tr \left( \left( Z^T X^T \right) L (XZ) \right)$$

$$s.t. \|Z\|_{1,q} = 1^T, Z \geq 0$$

Eq.(7) can be solved by Alternating Direction Method of Multiplier (ADMM) optimization framework [22]. The overall procedure of the ADMM algorithm is to introduce appropriate auxiliary variables into the optimization program, and iteratively optimize the Lagrangian with respect to the variables. We introduce an auxiliary variable matrix $M \in \mathbb{R}^{d \times d}$ to Eq.(7) and reformulate as:

$$\min_{Z} \frac{\lambda}{2} \|Z\|_{1,q} - tr \left( D^T M \right) + tr \left( \left( M^T X^T \right) L (XM) \right)$$

$$s.t. M = 1^T, Z = M$$

where $\langle \rangle$ denotes trace operator of a given matrix. The ADMM approach is an iterative procedure that proceeds as summarized in Table 1.

### III. EXPERIMENTS AND DISCUSSION

This section will verify the performance of the proposed method in hyperspectral classification. The experimental results correspondingly consist of two types: 1): a comparison between the propose framework with some classical and state-of-art unsupervised band selection methods; 2): test the robustness and insensitive to the noise and parameter of our proposed framework. The simulations are carried out in MATLAB7.10.0 (R2010a) environment running in Intel(R) i5-6500, 3.20-GHZ CPU with 16.0-GB RAM. All the numerical results are obtained based on the average of thirty independent trials.

### A. HYPERSONTICAL IMAGE DATA

1) INDIAN PINES DATASET

This first scene, Indian Pines Dataset, was gathered by AVIRIS sensor over the Indian Pines test site in North-western Indiana and consists of $145 \times 145$ pixels and 224 spectral reflectance bands in the wavelength range $(0.4–2.5) \times 10^{-6}$ meters. The Indian Pines scene contains two-thirds agriculture, and one-third forest or other natural perennial vegetation. There are two major dual lane highways, a rail line, as well as some low density housing, other built structures, and smaller roads. The ground truth available is designated into sixteen classes and is not all mutually exclusive. Figure 3 shows the true color image of this dataset overlaid with the ground truth map.

2) PAVIAU DATASET

This hyperspectral dataset was acquired in 2001 by the ROSIS instrument, flown over the city of Pavia, Italy. The scene contains pixels, but some of the samples in this images contain no information and have to be discarded before the analysis. which includes 9 classes, is centered at the
University of Pavia. The geometric resolution is 1.3 meters. After removing 12 bands due to noise and water absorption, this hyperspectral dataset comprises of 103 spectral channels. A true color image and the ground truth map are shown in Figure 4.

3) KSC DATASET
This dataset was collected over the Kennedy Space Center (KSC), Florida, on March 23, 1996 by The NASA AVIRIS. AVIRIS acquires data in 224 bands of 10 nm width with center wavelengths from 400 - 2500 nm. The KSC data, acquired from an altitude of approximately 20 km, have a spatial resolution of 18 m. After removing water absorption and low SNR bands, 176 bands were used for the analysis. Discrimination of land cover for this environment is difficult due to the similarity of spectral signatures for certain vegetation types. For classification purposes, 13 classes representing the various land cover types that occur in this environment were defined for the site.

B. CONVERGENCE RESULTS
In practice, how fast the algorithm converge is crucial for the whole computational efficiency. Hence we conduct some experiments to test the convergence of the proposed algorithm. Figure 5 (a) and (b) give the some numerical results on IndianPine dataset to show the convergence behavior of the proposed iteration procedure. As seen from Figure 4, the objective of RBS-LPA all converge to a fixed value. The number of iteration when \( q = \infty \) is much less than that of \( q = 2 \). Figure 5(c) and (d) give the numerical results on KSC dataset to show the convergence behavior of the proposed algorithm.

C. COMPARISONS WITH STATE-OF-ART METHODS
In this part, we utilize five state-of-art methods to make holistic comparisons with our methods, including:

1): Clustering-based Band Selection (CBS) [19]. CBS is a hierarchical clustering structure to group bands to minimize the intracluster variance and maximize the intercluster variance by using a certain information measures. And also k-medoids, is a clustering algorithm related to the k-means algorithm and the medoidshift algorithm, which is attempt to minimize the distance between points labeled to be in a cluster and a point designated as the center of that cluster. In contrast to the k-means algorithm, k-medoids chooses data points as centers. The other method is Affinity Propagation (AP), which takes as input measures of agreement between pairs of bands, messages are exchanged between bands until a high-quality set of exemplars and corresponding clusters gradually emerges. We used affinity propagation to cluster hyperspectral bands. The Affinity Propagation (AP) algorithm [6], [31], [32] tries to find representatives from pairwise similarities between data points by using an approximate message passing algorithm. While it has been shown empirically that AP performs well in problems such as unsupervised image categorization [23], there is no guarantee for AP to find the desired solution and, in addition, it works only with a single dataset.

2): Maximum-Variance Principal Components Analysis (MVPCA) [16]. This method based on eigenvalue decomposition of a data matrix, and selects bands with higher variance.

3): Exemplar Component Analysis (ECA) [32]: Exemplars component analysis chooses the exemplars as the selected bands, and exemplar is a band which has a high local density and at a large distance from the band with a higher density.

4): Sparsity based band selection (SpaBS) [21]: This band selection cast a sparse representation of the hyperspectral image data through K-SVD, that decomposes the image
data into the multiplication of an overcomplete dictionary (or signature matrix) and the coefficient matrix. The coefficient matrix, that possesses the sparsity property, reveals how importantly each band contributes in forming the hyperspectral data. By calculating the histogram of the coefficient matrix, we select the top K bands that appear more frequently.
D. Evaluation Metrics

Since the proposed RBS-LPA method does not convolve with specific classifiers. Therefore, we expect the bands selected by the proposed framework have good performance on various types of classifiers. To test this, we consider two widely used classifiers, i.e., K-Nearest Neighbors (KNN), Support Vector Machine (SVM). The SVM classifier implements the radial basis function as the kernel function, with the variance parameter and penalization factor estimated via cross validation. The Euclidean distance is used in the KNN classifier with the neighborhood size equal to 1 and 3.

In our experiments, the regularized parameters are associated with the selected number of bands. Fig. 6 gives the classification results of Pavia University Scene with different classifiers. The accuracy curve is sketch by varying the selected band number in the range from 1 to 30. As observed than others to serve the need for dimensionality reduction and at the same time preserving the physical meaning of the selected bands.
from Figure 7(a-e), the performance of proposed RBS-LPA dominate the competitors on the employed classifiers. Figure 8 shows the selected band indices of the proposed methods with its competitors. The comparative selected band indices for Pavia University Scene are illustrated in Figure 8, which presents the band produced by ECA [32], MVPCA [16], WaLuMI [19], WaLuDI [19] and our proposed algorithm, respectively. As observed from figures, our proposed algorithm selected the band across all the band space. In addition, the selected bands indices of our proposed algorithm RBS-LPA \((q = \infty)\), RBS-LPA \((q = 2)\) only 3 different band among the 20.

Figure 9 gives the classification results of KSC Scene with different classifiers. The accuracy curve is sketched by varying the selected band number in the range from 1 to 30. As observed from Figure 9(a-c), the performance of proposed DSRS-GR dominate the competitors when the number of bands is small, but the performance are basically flat with other competitors when the number of bands increased to 20.

Figure 10 shows the classification map and numerical result with 10 selected bands when 20 samples are labeled. The proposed method achieved the best results than the compared methods and have better spatial consistency than the other methods.

IV. CONCLUSION

This paper proposes an effective sparsity-induced band selection methods named as RBS-LPA. By refining the learned pairwise agreement with spatial relative position, the final agreement can well encode the intrinsic agreement between each band. Given pairwise agreement between each band, band selection can be considered as a representation learning process. Compared with sparsity-based band selection, we not only make any assumption about the independence between bands but also the band information and pixels information are exploited jointly to find a few representative bands that can efficiently encode the target bands.

REFERENCES

[1] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, “Advances in spectral-spatial classification of hyperspectral images,” Proc. IEEE, vol. 101, no. 3, pp. 652–675, Mar. 2013.
[2] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, “Hyperspectral remote sensing data analysis and future challenges,” IEEE Geosci. Remote Sens. Mag., vol. 1, no. 2, pp. 6–36, Jun. 2013.
[3] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, “Nearest neighborhood subspace for hyperspectral classification,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 1, pp. 477–489, Jan. 2014.
[4] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification. Hoboken, NJ, USA: Wiley, 2012.
[5] L. Van Der Maaten, E. Postma, and J. Van den Herik, “Dimensionality reduction: A comparative,” J. Mach. Learn. Res., vol. 10, nos. 66–71, p. 13, 2009.
[6] N. Kwak and C.-H. Choi, “Input feature selection for classification problems,” IEEE Trans. Neural Netw., vol. 13, no. 1, pp. 143–159, Aug. 2002.
[7] K. Javed, H. A. Babri, and M. Saeed, “Feature selection based on class-dependent densities for high-dimensional binary data,” IEEE Trans. Knowl. Data Eng., vol. 24, no. 3, pp. 465–477, Mar. 2012.
[8] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” J. Mach. Learn. Res., vol. 3, pp. 1157–1182, Jan. 2003.
[9] M. Dash and H. Liu, “Feature selection for classification,” Intell. Data Anal., vol. 1, nos. 1–4, pp. 131–156, 1997.
[10] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, “Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 5, no. 2, pp. 354–379, Apr. 2012.
[11] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” Chemometrics Intell. Lab. Syst., vol. 2, nos. 1–3, pp. 37–52, 1987.
[12] Q. Song, J. Ni, and G. Wang, “A fast clustering-based feature subset selection algorithm for high-dimensional data,” IEEE Trans. Knowl. Data Eng., vol. 25, no. 1, pp. 1–14, Jan. 2013.
[13] L. Laporte, R. Flamary, S. Canu, S. Dejean, and J. Mothe, “Nonconvex regularizations for feature selection in ranking with sparse SVM,” IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 6, pp. 1118–1130, Apr. 2014.
[14] Z. Xu, I. King, M. R-T. Lyu, and R. Jin, “Discriminative semi-supervised feature selection via manifold regularization,” IEEE Trans. Neural Netw., vol. 21, no. 7, pp. 1033–1047, Jul. 2010.
[15] Y. Yuan, G. Zhu, and Q. Wang, “Hyperspectral band selection by multi-task sparsity pursuit,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 2, pp. 631–644, Feb. 2015.
[16] C.-I. Chang, Q. Du, T.-L. Sun, and M. 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.
[17] 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.
[18] Q. Du and H. Yang, “Similarity-based unsupervised band selection for hyperspectral image analysis,” IEEE Geosci. Remote Sens. Lett., vol. 5, no. 4, pp. 564–568, Oct. 2008.
[19] A. Martínez-UsóMartínez-Uso, F. Pla, J. M. Sotoca, and P. García-Sevilla, “Clustering-based hyperspectral band selection using information measures,” IEEE Trans. Geosci. Remote Sens., vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
[20] Y. Qian, F. Yao, and S. Jia, “Band selection for hyperspectral imagery using affinity propagation,” IET Comput. Vis., vol. 3, no. 4, p. 213, 2009.
[21] S. Li and H. Qi, “Sparse representation based band selection for hyperspectral images,” in Proc. 18th IEEE Int. Conf. Image Process., Sep. 2011, pp. 2693–2696.
[22] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, “Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 5, pp. 2956–2969, May 2015.
[23] E. Elhamifar, G. Sapiro, and S. S. Sastry, “Dissimilarity-based sparse subset selection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 11, pp. 2182–2197, Nov. 2016.
[24] W. Sun, L. Zhang, L. Zhang, and Y. M. Lai, “A dissimilarity-weighted sparse self-representation method for band selection in hyperspectral imagery classification,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 9, pp. 4374–4388, Sep. 2016.
[25] S. Boyd, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” FNT Mach. Learn., vol. 3, no. 1, pp. 1–122, 2010.
[26] E. Elhamifar, “Sparse modeling for high-dimensional multi-manifold data analysis,” Ph.D. dissertation, Johns Hopkins Univ., Baltimore, MD, USA, 2012.
[27] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, “Graph regularized sparse coding for image representation,” IEEE Trans. Image Process., vol. 20, no. 5, pp. 1327–1336, May 2011.
[28] L. Ma, M. M. Crawford, and J. Tian, “Local manifold learning-based k-nearest-neighbor for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 48, no. 11, pp. 4099–4109, Aug. 2010.
[29] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, “Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning,” IEEE Signal Process. Mag., vol. 31, no. 1, pp. 55–66, Jan. 2014.
[30] Y. Y. Tang, H. Yuan, and L. Li, “Manifold-based sparse representation for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 12, pp. 7606–7618, Apr. 2014.
ZHIXI FENG (Member, IEEE) received the B.A. degree in automation from the Lanzhou University of Technology, in 2012, and the Ph.D. degree in intelligent information processing from Xidian University, in 2018. He is currently a Lecturer of artificial intelligence with Xidian University. His research interests include machine learning and remote sensing information processing.

SHUYUAN YANG received the B.A. degree in electrical engineering, and the M.S. and Ph.D. degrees in circuit and system from Xidian University, Xi’an, China, in 2000, 2003, and 2005, respectively. She has been a Professor of artificial intelligence with Xidian University. Her research interests include machine learning and image processing.

XIAOLONG WEI received the B.S. degree in electronic information engineering from the Xi’an University of Posts and Telecommunications, Xi’an, China, in 2010, and the Ph.D. degree in aerospace science and technology from Air Force Engineering University, Xi’an, in 2016. In 2017, he joined the Science and Technology on Plasma Dynamics Laboratory, Air Force Engineering University, as a Lecturer. His research interests include deep learning, nondestructive testing of materials, and machine vision.

QUANWEI GAO (Senior Member, IEEE) received the B.S. degree in electronic information science and technology from the Shandong University of Technology, in 2017. He is currently pursuing the master’s degree in electronics and communication engineering with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China, Xidian University, Xi’an, China. His current research interests include target detection and data fusion.

LICHENG JIAO (Fellow, IEEE) received the B.S. degree from Shanghai Jiao Tong University, in 1982, and the M.S. and Ph.D. degrees from Xi’an Jiaotong University, Xi’an, China, in 1984 and 1990, respectively. He has been a Professor with the School of Artificial Intelligence, Xidian University, Xi’an, where he is currently the Director of the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education of China. His research interests include image processing, natural computation, machine learning, and intelligent information processing.