Hyperspectral Band Selection by Multitask Sparsity Pursuit

Yuan Yuan, *Senior Member, IEEE*, Guokang Zhu, and Qi Wang

***Abstract*—Hyperspectral images have been proved to be effective for a wide range of applications; however, the large volume and redundant information also bring a lot of inconvenience at the same time. To cope with this problem, hyperspectral band selection is a pertinent technique, which takes advantage of removing redundant components without compromising the original contents from the raw image cubes. Because of its usefulness, hyperspectral band selection has been successfully applied to many practical applications of hyperspectral remote sensing, such as land cover map generation and color visualization. This paper focuses on groupwise band selection and proposes a new framework, including the following contributions: 1) a smart yet intrinsic descriptor for efficient band representation; 2) an evolutionary strategy to handle the high computational burden associated with groupwise-selection-based methods; and 3) a novel MTSP-based criterion to evaluate the performance of each candidate band combination. To verify the superiority of the proposed framework, experiments have been conducted on both hyperspectral classification and color visualization. Experimental results on three real-world hyperspectral images demonstrate that the proposed framework can lead to a significant advancement in these two applications compared with other competitors.**

***Index Terms*—Band selection, compressive sensing (CS), hyperspectral image, immune clonal strategy (ICS), machine learning, multitask learning (MTL).**

I. Introduction

HYPERSPECTRAL images have attracted much attention from both academia and industry over the past few years. A hyperspectral image, or an image cube, can provide information from a wide range of wavelengths with fine spectral resolution. This characteristic makes it beneficial to a wide range of applications. For example, it has already been successfully introduced in environmental monitoring [1], biological analysis [2], medical imaging [3], production quality inspection [4], etc.

However, the increased volume of data contained in the hyperspectral image also entails researchers the inconvenience and inefficiency in information storage, display, transmission, and processing [5], [6]. Moreover, the hyperspectral information in an image cube often contains a large amount of redundancy and therefore has some undesirable statistical and geometrical properties [7], [8]. These drawbacks are mainly because that the hyperspectral image is commonly represented by hundreds of bands, which are highly correlated among neighboring bands. For some cases, increasing the spectral bands can enhance the performance, but that does not mean that all the bands take the role. The most critical factor might be only several few bands. For this reason, it is desirable to develop a feature selection technique that can automatically select only a few decisive and physically meaningful bands to represent the whole image cube without losing effectiveness. This technique is also known as *band selection*, which has been an attractive topic in hyperspectral remote sensing.

Hyperspectral band selection is often compared with another kind of redundancy reduction technique—feature extraction [9]. The latter is typically achieved by projecting the original information to a lower dimensional feature space [10]–[13]. In the field of hyperspectral image processing, band selection is often preferable to feature extraction, since it takes advantage of preserving the original information from the raw image cube [14], [15]. Reliable selection results can not only facilitate the storage, display, transmission, and processing but also guide effective data acquisition. As for this work, we mainly focus on the problems of hyperspectral classification and color visualization in hyperspectral image processing.

Recently, many hyperspectral band selection methods have been proposed to support efficient hyperspectral classification [16], [17] and color visualization [18], [19]. However, the performances of these methods are still far from satisfying. Most of these methods rely on independent selection that tackles bands individually, ignoring the interdependencies among them, whereas in practical applications, the selected bands will work jointly. In addition, these methods also suffer from the high computational burden due to lack of a compact band descriptor, although the independent selection strategy is employed. Beneficial from recent advances in machine learning, there are many techniques such as sparse learning [20]–[22], subspace selection [23], [24], and feature mining [25], which can provide some refreshing potentials to band selection and improve it for a better performance. In this paper, we propose a *multitask sparsity pursuit* (MTSP) framework for unsupervised hyperspectral band selection, which contributes the following.

- Formulate the representative band selection as a multitask sparse learning problem, where learning the representation of each band is viewed as a single task. The associated

Manuscript received April 12, 2013; revised July 3, 2013, September 22, 2013, December 26, 2013, and March 20, 2014; accepted May 16, 2014. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707104; by the National Natural Science Foundation of China under Grant 61172143, Grant 61105012, and Grant 61379094; and by the Fundamental Research Funds for the Central Universities under Grant 3102014JC02020G07.

Y. Yuan and G. Zhu are with the Center for OPTical IMagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: yuany@opt.ac.cn; zhuguokang@opt.ac.cn).

Q. Wang is with the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: crabwq@nwpu.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.2014.2326655

joint sparsity constraint problem is optimized using an *accelerated proximal gradient* (APG) method that achieves the global solution with quadratic convergence [26], [27]. To the best of our knowledge, there is no other work with regard to exploiting *multitask learning* (MTL) in hyperspectral band selection.

- Design a *compressive sensing* (CS)-based descriptor for efficient yet intrinsic hyperspectral band representation. Although the raw information of a band could be directly used as a description, this strategy often goes with unacceptable computational burden due to the huge dimensionality. In this paper, we define a new hyperspectral band descriptor by exploiting the sparse signal recovery power of CS.
- Construct an efficient searching strategy for candidate hyperspectral band combinations based on *immune clonal strategy* (ICS). To choose the best bands from the large volume, some conventional methods proposed to employ the subset forward-searching strategies [6], [28]. These strategies need to preliminarily determine one or several bands for initialization and are suitable for independent selection. However, this kind of strategies cannot be applied to our framework, which exploits interdependencies among different tasks for MTSP. Fortunately, we found that the powerful global searching ability of ICS can provide us an ideal solution without the need of examining perhaps billions of possible hyperspectral band combinations.

The rest of this paper is organized as follows. Section II reviews the works on the topic of hyperspectral band selection. Section III describes each component of the proposed framework in detail. Section IV presents the extensive experiments to prove the superiority of the proposed framework, and the conclusion follows in Section V.

II. Related Work

With the problems caused by the raw hyperspectral images, many efforts have been made in both theory and practice to refine the redundant data for efficient and effective storage, transmission, processing, etc. There are many selection criteria that have been introduced. The most popular ones include linear prediction, orthogonality, spectral angle, spectral correlation, spectral derivation, mutual information, and Kullback–Leibler divergence. Based on these criteria, two major types of strategies, i.e., pointwise selection and groupwise selection, are used to determine the effective band combination.

Pointwise-selection-based methods begin with an empty candidate set and then gradually augment it, or start with a full candidate set and then sequentially remove unfitted bands, until the desired size is reached. For example, Chang and Wang [29] proposed to sequentially exclude correlated bands using a divergence measure. Their method linearly constrains bands while minimizing correlations or dependencies among them. Du and Yang [30] initialized a pair of most distinctive bands through orthogonal subspace projection and then sequentially added bands most dissimilar to the determined ones, according to the cost of linear prediction and orthogonal subspace construction. In addition, Yang *et al.* [28] proposed a supervised method based on class signature prior. In their work, a sequential forward selection strategy is also employed based on the same initialization procedure as [30], and the expected band set is the one yielding the high classification accuracy.

On the other hand, in groupwise-selection-based methods, the bands selected out are simultaneously determined. For example, Martínez-Usó *et al.* [9] started from clustering bands to minimize the intracluster variance while maximizing the intercluster variance based on the measure of mutual information and Kullback–Leibler divergence. Then, the most representative band combination is determined as the one constituted by the bands with highest average correlations in the corresponding clusters. Dissimilarly, Su *et al.* [31] proposed to group all the bands by semisupervised k-means clustering with the use of class signature prior. Then, they excluded the clusters with the centroid bands dissimilar to others and selected the remaining cluster centers as the final output. More recently, Yin *et al.* [32] have introduced a computational evolutionary strategy into the field of supervised band selection. In their work, the candidate band combinations are evaluated through an affinity function driven by hyperspectral classification accuracy.

In addition to the aforementioned methods being general for a wide range of applications, there are also many task-driven band selection methods dedicated only to hyperspectral image visualization. These methods aim at providing the most representative three bands of tristimulus display suitable for human visual perception. For all these methods, the most basic requirement is that the displayed false color image should keep the discriminative ability of objects [19], [33]. Many signal processing techniques have been introduced to this type of task-driven band selection. For instance, Demir *et al.* [18] proposed to utilize *one-bit transform* (1BT) to select three most suitable spectral channels for *RGB* display. According to the transformation results, they first determined some well-structured bands as candidates and then further refined the candidate set using correlation measure to obtain the most dissimilar three bands. Moreover, information measure has been also proved powerful for hyperspectral image visualization. In [19], Moan *et al.* employed three different orders of information measures to compare spectral channels to determine the most important triple components. They first sought two most dissimilar bands as the R and B channels and then selected the one minimizing the third-order information measure among this band and the previously identified two as the G channel. Recently, supervised band selection method has been also introduced in hyperspectral image visualization. In [34], the band subset yielding the largest class separability is selected with a forward-searching strategy and divided into three spectral ranges. Then, the most uncorrelated three bands are further selected and mapped into RGB color space.

III. Proposed Framework

This section details the proposed MTSP-based framework for unsupervised groupwise band selection. The overall flowchart is summarized in Fig. 1. In this framework, the most crucial component is the MTSP-based criterion for the evaluation of

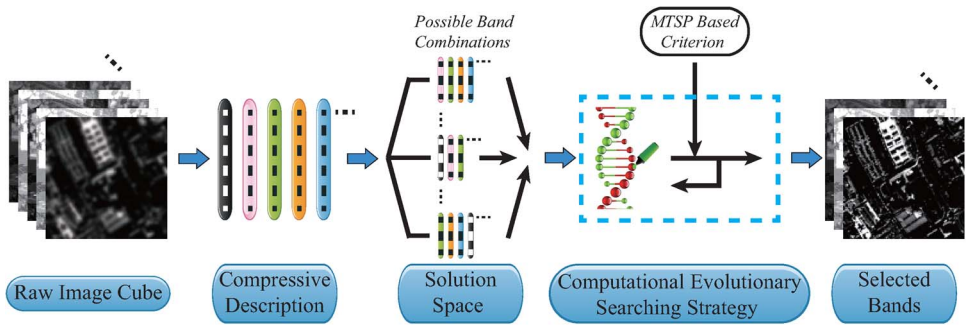

Fig. 1. Flowchart of the MTSP-based band selection framework.

each possible band combination. In addition, as it is found that using raw band information for MTL is not computationally acceptable in practice, a random-projection-based descriptor is designed, which exploits the sparse signal recovery power of CS to effectively represent each hyperspectral band before MTL. Then, a computational evolutionary strategy is proposed to efficiently search for the desired band combination from perhaps billions of candidates. The following paragraphs will individually describe all these functional aspects.

A. Compressive Hyperspectral Band Description

For the groupwise hyperspectral band selection, naively describing each band using the raw image information as descriptor is computationally too expensive for an ordinary terminal processor. This obstacle exists even for processing a medium-sized image. Therefore, a smart yet intrinsic descriptor is of prime importance for the success of hyperspectral image processing. This issue has been attracting much attention in recent years. Here, an effective and efficient descriptor is proposed to represent hyperspectral bands in the compressed domain.

1) Sparse-Measurement-Matrix-Based Random Projection: Based on CS theory [35], [36], a small number of randomly generated linear measurements can preserve most of the salient information encoded in the original high-dimensional signal, if it is compressible. This technique provides an alternative to Shannon–Nyquist sampling [37], [38]. The projections satisfying the *restricted isometry property* (RIP) [39] are proved to be adequate to preserve the information when projecting the compressible signals to a low-dimensional compressed subspace. RIP characterizes matrices that can act as nearly orthonormal bases, at least when operating on sparse vectors, and is proved sufficient to guarantee perfect recovery [39]. Baraniuk *et al.* [40] proved that the random measurement matrix satisfying the Johnson–Lindenstrauss lemma[1] can also satisfy the RIP in CS. To be specific, for the projection

$$\mathbf{x} = \mathbf{R}\mathbf{y} \tag{1}$$

the original high-dimensional data description $\mathbf{y} \in \mathbb{R}^d$ can be almost perfectly reconstructed from the low-dimensional description $\mathbf{x} \in \mathbb{R}^k$ $(k \ll d)$, if $\mathbf{y}$ is compressive and the random measurement matrix $\mathbf{R} \in \mathbb{R}^{k\times d}$ satisfies the Johnson–Lindenstrauss lemma. In other words, this projection can ensure that $\mathbf{x}$ preserves almost all the information in $\mathbf{y}$.

Limited by the requirement of satisfying RIP, a typically feasible measurement matrix is the random Gaussian matrix $\mathbf{R} \in \mathbb{R}^{k\times d}$ with entries $r_{ij} \sim N(0,1)$ [42], where the subscripts indicate the positions of the entries in $\mathbf{R}$. However, as it is a dense matrix, the storage and computational costs are still high when k is with a large value. In order to perform a more practical information-preserving and dimensionality reduction for hyperspectral band description, a very sparse measurement matrix satisfying the RIP is employed in [43]. This measurement matrix facilitates efficient projection from the original feature space to a low-dimensional compressed subspace, which has the entries defined as

$$r_{ij} = \begin{cases} \sqrt{s}, & \text{with probability } \frac{1}{2s} \\ 0, & \text{with probability } 1 - \frac{1}{s} \\ -\sqrt{s}, & \text{with probability } \frac{1}{2s}. \end{cases} \tag{2}$$

It is proved that this measurement matrix with $s = O(d)$ $(\mathbf{y} \in \mathbb{R}^d)$, even when $s = d/\log(d)$, can support a random projection with almost the same accuracy as the traditional scheme where $r_{ij} \sim N(0,1)$ [43]. As recommended in [42], this work sets $s = d/4$ to construct a very sparse random matrix for reliable projection.

2) Low-Dimensional Compressive Descriptor: In our band selection framework, bands are described by low-dimensional descriptors. These descriptors are extracted from the multiscale integral image feature space based on the aforementioned random projection. The overall flowchart of our feature extraction scheme is summarized in Fig. 2. This scheme mainly consists of three functional steps.

First, for each hyperspectral band, a dyadic Gaussian pyramid with four spatial scales is created to exploit the multiscale information in the band. This procedure is implemented by successively low-pass filtering and subsampling the input image and, finally, yields four octaves with horizontal and vertical scale reduction factors ranging from 1 (scale zero) to 0.125 (scale three).

Second, four integral images are calculated for the generated octaves. This treatment is found to take into account the global relationships between pixels within the same octave in our experiments. Then, each hyperspectral band can be represented as a very high dimensional multiscale feature vector $\mathbf{y} \in \mathbb{R}^d$, where d is directly the number of pixels in the dyadic Gaussian

[1]The Johnson–Lindenstrauss lemma [41] states that the distances between the points can be nearly preserved if they are embedded into a randomly selected subspace with suitable dimensions.

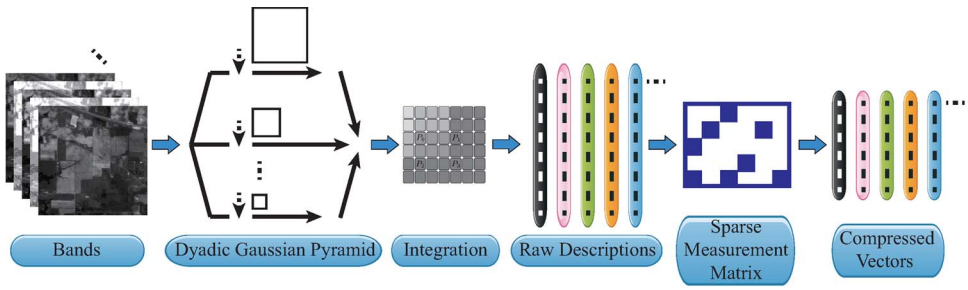

Fig. 2. Flowchart of the proposed compressive band description algorithm.

pyramid. Typically, d is on the order 10^4-10^6 for a hyperspectral band.

Finally, as discussed in Section III-A1, a very sparse random measurement matrix $\mathbf{R}$ with the setting $s = d/4$ is adopted. This random matrix is then used to project the original hyperspectral band description $\mathbf{y}$ onto a compressive vector $\mathbf{x} \in \mathbb{R}^k$ in a low-dimensional feature space using (1). In this paper,k is experimentally fixed to 200.

For a hyperspectral image cube, the random measurement matrix needs to be constructed only once and kept fixed for all the component bands. This matrix is very spare, and the corresponding storage and computational cost is very light. Factually, there are nonzero entries in $\mathbf{R}$, except the zero ones, and only the corresponding positions in the integral images need assigning the storage and computational resources in practice. Furthermore, benefiting from the very low-dimensional description $\mathbf{x}$, our overall groupwise band selection framework can be efficiently implemented within acceptable cost.

B. MTSP-Based Criterion

The proposed framework is based on groupwise selection that tackles each possible band combination as a complete unit. Before choosing the candidate band sets for consideration, the evaluation criterion should be introduced first. Here, an MTSP-based criterion is proposed to evaluate the appropriateness of an examined band combination. The following will describe each stage of this criterion in detail.

1) Multitask Representation of Hyperspectral Image: In many practical situations, a learning problem can be tackled as several related subtasks that share dependencies in some latent factors. Hence, employing the MTL framework that jointly solves these subtasks by utilizing their inherent relationships will have more advantages than tackling each one independently [44]. Recently, MTL has been successfully applied to various applications, such as image tagging [45], image classification [46], saliency detection [47], and object tracking [48]. In this paper, we formulate the evaluation of a band combination as an MTL problem, where the representation of each band in the image cube is considered as a subtask.

With the high band correlation and data redundancy, a hyperspectral band is usually very similar to the neighboring bands. In other words, a band in a hyperspectral image cube will be well linearly reconstructed by some others in general. Some traditional works have already taken into account this characteristic to select the bands dissimilar to the others [31], [49]. In this paper, the same characteristic is exploited in another way, which assumes that each band in the hyperspectral image cube can be represented as a sparse linear combination of some representative bands.

To be specific, in our MTSP-based criterion, a hyperspectral image cube is denoted in matrix form as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, where each column in this matrix is a vectorial representation in $\mathbb{R}^k$ of a band, whereas the desired band combination is viewed as a dictionary and denoted by $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_m]$ with desired band $\mathbf{d}_i$. Each band $\mathbf{x}_i$ can be represented as a linear combination $\mathbf{z}_i$ of the dictionary $\mathbf{D}$, such that $\mathbf{X} = \mathbf{DZ}$.

Moreover, the high redundancy and correlation are known to be contained between neighboring bands. This characteristic culminates in the subtask of individual band representations being jointly sparse. Specifically, in our case, band representations are encouraged to be individually sparse and share dictionary bands that can reliably reconstruct them by joint sparsity pursuit.

The most common strategy to impose joint sparsity in MTL framework is to use mixed norm, such as $\ell_{2,1}$ and $\ell_{\infty,1}$. These two regularizers are traditionally the sparsity-inducing norms to yield the sharing parameters among the individual subtasks. An example of how the joint sparsity regularizer works is presented in Fig. 3. In this paper, the convex $\ell_{2,1}$ and $\ell_{\infty,1}$ regularizers are investigated to address the problem of MTL in the representation of a hyperspectral image cube. In addition, the conventional ℓ_1 norm that regularizes each representation task independently will be also taken into account as a benchmark. Without losing generality, we introduce the notation $\ell_{p,1}$, $p \in \{1, 2, \infty\}$, which is defined as $\|\mathbf{Z}\|_{p,1} = \sum_i \|\mathbf{Z}_i\|_p$, where $\|\mathbf{Z}_i\|_p$ is the ℓ_p norm of $\mathbf{Z}_i$, the ith row of matrix $\mathbf{Z}$. Based on this definition, our MTL problem can be formulated as a convex optimization problem in (3), where the parameter λ controls the tradeoff between the sparsity of the solution and the fidelity of the approximation to $\mathbf{X}$, i.e.,

$$\min_{\mathbf{Z}} \|\mathbf{X} - \mathbf{DZ}\|_F^2 + \lambda \|\mathbf{Z}\|_{p,1}. \tag{3}$$

2) APG Method: The method employed in this paper to solve MTL is based on the APG approach [27]. This approach has been successfully used to provide a fast convergence rate solution for convex optimization problems with nonsmooth terms. To be specific, the APG approach is originally developed for the following minimization:

$$\min f(\mathbf{Z}) + g(\mathbf{Z}) \tag{4}$$

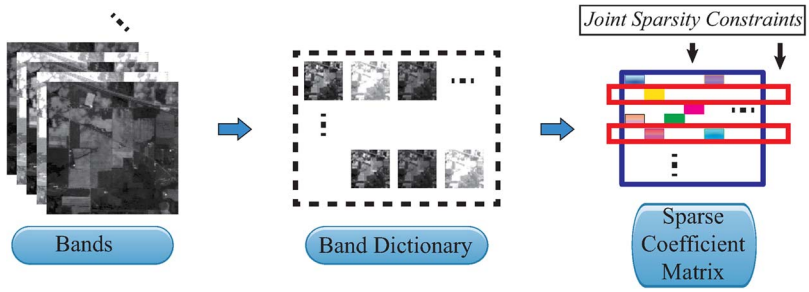

Fig. 3. Illustration of how the joint sparsity regularizer works. The ith column of the sparse coefficient matrix constitutes a representation in $\mathbb{R}^m$ for $\mathbf{x}_i$. A joint sparsity regularizer not only encourages each column vector to be sparse but also constrains coefficients in each row.

where $f(\mathbf{Z})$ is a smooth convex function, and its gradient is Lipschitz continue, whereas $g(\mathbf{Z})$ is a continue function but possibly nonsmooth. Compared with traditional projected subgradient optimizers that achieve sublinear convergence, the advantage of APG is justified by its quadratic convergence rate $O(1/t^2)$ for global solution. Concretely, each APG iteration involves two pivotal steps: a *gradient mapping* step, which updates the representation matrix $\mathbf{Z}^{(t)}$ with the fixed aggregation matrix $\mathbf{G}^{(t)}$, and an *aggregation* step, which linearly combines $\mathbf{Z}^{(t+1)}$ and $\mathbf{Z}^{(t)}$ to construct $\mathbf{G}^{(t+1)}$.

In the gradient mapping step, the algorithm will obtain $\mathbf{Z}^{(t+1)}$ by solving (5) with the currently estimated $\mathbf{G}^{(t)}$, i.e.,

$$\mathbf{Z}^{(t+1)} = \arg\min_{\mathbf{W}} \frac{1}{2}\|\mathbf{W}-\mathbf{V}\|_F^2 + \eta g(\mathbf{W}) \tag{5}$$

$$\mathbf{V} = \mathbf{G}^{(t)} - \eta \nabla f\left(\mathbf{G}^{(t)}\right) \tag{6}$$

where the temporary $\mathbf{V}$ is set for the sake of simplicity, and η is a small step parameter.

As for the aggregation step, the linear combination strategy for updating the aggregation matrix usually varies according to the specific requirements of applications. In the MTL framework, the most commonly used strategy is as follows:

$$\mathbf{G}^{(t+1)} = \mathbf{Z}^{(t+1)} + \frac{1-\alpha^{(t)}}{\alpha^{(t)}}\alpha^{(t+1)}\left(\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\right) \tag{7}$$

where $\alpha^{(t+1)}$ is congenitally calculated by $\alpha^{(t+1)} = 2/(t+3)$.

3) Fast Numerical Method for Solving (3): In our case, the APG method can be applied to (3) with

$$\begin{cases} f(\mathbf{Z}) = \|\mathbf{X}-\mathbf{D}\mathbf{Z}\|_F^2 & (8) \\ g(\mathbf{Z}) = \lambda\|\mathbf{Z}\|_{p,1}. & (9) \end{cases}$$

Then, the expected representation matrix $\mathbf{Z}^{(t+1)}$ will take the following form:

$$\mathbf{Z}^{(t+1)} = \arg\min_{\mathbf{W}} \frac{1}{2}\|\mathbf{W}-\mathbf{V}\|_F^2 + \eta\lambda\|\mathbf{W}\|_{p,1} \tag{10}$$

where $\mathbf{V}$ can be obtained by rewriting (6) as

$$\mathbf{V} = \mathbf{G}^{(t)} - 2\eta\mathbf{D}^T\left(\mathbf{D}\mathbf{G}^{(t)} - \mathbf{X}\right). \tag{11}$$

So far, the minimization problem (3) has been converted to solving (10) in each APG iteration. However, directly finding a solution from (10) is not yet a workable scheme. Here, we suggest a look back of the $\ell_{p,1}$ definition, which can help decouple (10) into m disjoint subproblems. Each subproblem corresponds to a row vector $\mathbf{Z}_i$ and can be easily solved by exploiting the structure of the ℓ_p ball with the complexity depended on p. More specifically, all the disjoint subproblems are formulated by the following equation:

$$\mathbf{Z}_i^{(t+1)} = \arg\min_{\mathbf{W}_i} \frac{1}{2}\|\mathbf{W}_i-\mathbf{V}_i\|_2^2 + \eta\lambda\|\mathbf{W}_i\|_p \tag{12}$$

where $\mathbf{V}_i$ is the corresponding ith row of matrix $\mathbf{V}$. Then, the solutions of (12) for $p \in \{1, 2, \infty\}$ can be easily obtained by

$$\mathbf{Z}_i^{(t+1)} = \begin{cases} \operatorname{sign}(\mathbf{V}_i)\max\left(0, \|\mathbf{V}_i\|_1 - \eta\lambda\right), & p=1 \\ \max\left(0, 1-\frac{\eta\lambda}{\|\mathbf{V}_i\|_2}\right)\mathbf{V}_i, & p=2 \\ \max\left(0, 1-\frac{\eta\lambda}{\|\mathbf{V}_i\|_1}\right)\mathbf{a}, & p=\infty \end{cases} \tag{13}$$

where $\mathbf{a}_j = \operatorname{sign}(\mathbf{V}_{ij})\min(\|\mathbf{V}_i\|_1, (\sum_{r=1}^{\hat{j}} u_r - \eta\lambda)/\hat{j})$, $j=1, 2, \ldots, n$. The temporary parameters u_r and $\hat{j}$ are obtained by setting $u_j = |\mathbf{V}_{ij}|$ and sorting these values in decreasing order $(u_1 \geq u_2 \geq \cdots \geq u_n)$. Then, $\hat{j}$ is calculated as $\hat{j} = \max\{j : \sum_{r=1}^{\hat{j}}(u_r - u_j) < \eta\lambda\}$.

Note that APG convergence in our case is achieved when the relative change rate in the cost of (3) falls below a given tolerance $\tau = 10^{-4}$ within a predefined step size $h = 10$. To be specific, the following ratio is calculated in each iteration t:

$$\phi = 1 - \frac{\min\limits_{t\leq i\leq t+h} f\left(\mathbf{Z}^{(i)}\right) + g\left(\mathbf{Z}^{(i)}\right)}{\max\limits_{t\leq i\leq t+h} f\left(\mathbf{Z}^{(i)}\right) + g\left(\mathbf{Z}^{(i)}\right)}. \tag{14}$$

The procedure will stop when $\phi \leq \tau$.

4) Criterion for Representativeness: As the final step, a selection of original bands must be identified according to some given evaluation criterion. In this paper, the finally preserved bands are those as representative as possible based on the aforementioned MTL result. In other words, the selected bands are those yielding the minimum cost in (3).

This scheme is similar to one popular criterion—linear prediction [31], [49]—which is originally used to evaluate the distinctiveness of an examined band by jointly measuring the similarity between this single band and other multiple bands. The proposed criterion differs from traditional linear-prediction-based methods mainly in two aspects: First, in our criterion, there is joint sparsity constraint imposed on the coefficients of linear combination. Second, the bands selected by our criterion are no longer the ones with the maximum reconstruction error using the linear combination of other selected bands.

C. Computational Evolutionary Band Selection

For a groupwise hyperspectral band selection task, there are perhaps billions of candidate band combinations to be analyzed. For efficient selection, some traditional methods choose to employ the subset forward-searching strategies [6], [28]. However, these strategies are suitable only for independent selection and cannot be applied to the situation that exploits interdependencies among different tasks. Fortunately, for this type of optimization problem, the computational evolutionary strategies can now provide us an ideal solution. Among these strategies, ICS is probably the one of greatest popularity due to its excellent global searching ability. Therefore, in this paper, ICS is also employed to solve the hyperspectral band selection problem.

1) Affinity Function Based on the Proposed Criterion: The ICS derives from the theory of clonal selection, which describes the dynamic learning mechanisms of the biological immune system counteractive to antigenic invasions [50]. In a typical ICS, the antigens are generally regarded as the target problems to be solved, whereas the antibody cells in the immune system are imitated as the solutions. It is obvious that different antibody cells may eliminate the antigens with different powers. This difference can help discriminate the effectiveness of an antibody cell and is measured by the affinity function in the computational system. The main process of ICS is to efficiently search for the appropriate antibody with the best affinity for the invaded antigen.

For the ultimate purpose, the finally preserved band combination is the one complying with the minimum reconstruction criterion introduced in Section III-B4. Driven by this consideration, in our ICS, the problem of hyperspectral image representation is converted as the antigen, whereas each band combination is regarded as an antibody cell. Then, the affinity for an antibody can be easily obtained by the following function:

$$A(\mathbf{D}) = e^{-\|\mathbf{X}-\mathbf{D}\mathbf{Z}\|_F^2 - \lambda\|\mathbf{Z}\|_{p,1}}. \tag{15}$$

For convenient interpretation of this function, a retrospect of notation definition should be presented. As defined in Section III-B1, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ is the description for a hyperspectral image with each column corresponding to a band, $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_m]$ is an examined band combination denoted with the same form as $\mathbf{X}$, and $\mathbf{Z}$ is the sparse matrix consisting of the reconstruction coefficients obtained by our MTL framework. It should be noted that each $\mathbf{D}$ is a subset of $\mathbf{X}$ and is actually specified by a binary vector in implementation, indicating which column of $\mathbf{X}$ is selected. A larger value of $A(\mathbf{D})$ indicates a better band combination.

2) Immune Operator for Effective Band Selection: In the biological immune system, when a new type of antigens has invaded a system, the organism can perform immune clonal multiplication to evolve the high-affinity antibody for defense. This process mainly involves clone, mutation, and selection. Correspondingly, in the ICS, the expected antibody cells are also selected through these three immune operators. In this paper, the possible band combinations are the antibody populations. A set of antibody population $\mathbb{D} = \{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_N\}$, where N is heuristically set as 10, will undergo the state transformation of ICS given as follows to select the highest affinity case:

$$\mathbb{D}(t) \xrightarrow{T_c^C} \mathbb{D}'(t) \xrightarrow{T_c^M} \mathbb{D}''(t) \xrightarrow{T_c^S} \mathbb{D}(t+1)$$

where T_c^C, T_c^M, and T_c^S represent the clone, mutation, and selection operators, respectively; and $\mathbb{D}'(t)$, $\mathbb{D}''(t)$, and $\mathbb{D}(t+1)$ are the correspondingly evolved antibody populations.

The clone operator T_c^C is, in fact, directly the self-copy of the antibodies. During this process, the clone number $N_C(\mathbf{D}_i)$ of each examined antibody is determined by its affinity for the given antigen. In this paper, $N_C(\mathbf{D}_i)$ is given by

$$N_C(\mathbf{D}_i) = \text{Int}\left(N \cdot \frac{A(\mathbf{D}_j)}{\max\limits_{j\in[1,N]} A(\mathbf{D}_j)}\right) \tag{16}$$

where $\text{Int}(\cdot)$ is a rounding-up function.

The mutation operator T_c^M aims at enriching the diversity of the antibody population. The strategy used in this stage is also related to the affinities of the antibodies. In this paper, we randomly pick out $N_M(\mathbf{D}_i')$ elements in the copied antibody $\mathbf{D}'(i)$ and then replace them with the equivalent numbers of other available candidates. $N_M(\mathbf{D}_i')$ is a random number from $[1, N_C(\mathbf{D}_i)]$, where $N_C(\mathbf{D}_i)$ is the clone number of the parent antibody cell $\mathbf{D}_i$. In addition, a supplement process is set within the mutation operator to avoid repeatedly constructing antibodies with the identical elements.

The selection operator T_c^S corresponds to a process of affinity maturation. This operator can preserve the most appropriate antibodies with the highest affinities as memory cells. In this paper, the probability of a mutated antibody cell $\mathbf{D}_i''$ replacing its parent $\mathbf{D}_i$ is calculated by the following equation:

$$p(\mathbf{D}_i \leftarrow \mathbf{D}_i'') = \begin{cases} 1, & A(\mathbf{D}_i) < A(\mathbf{D}_i'') \\ e^{-\frac{A(\mathbf{D}_i) - A(\mathbf{D}_i'')}{N_C(\mathbf{D}_i)}}, & A(\mathbf{D}_i) \geq A(\mathbf{D}_i''). \end{cases} \tag{17}$$

Then, the updated $\mathbf{D}_i$ $(i = 1, 2, \ldots, N)$ will constitute $\mathbb{D}(t+1)$ for the next iteration. The updating procedure stops when the relative change rate in the highest affinity during the last 50 steps falls below a predefined tolerance 10^{-4}.

IV. Experiments

This section will verify the performance of the proposed framework applied to both hyperspectral classification and color visualization. The experimental results correspondingly consist of two types of comparisons between our method and other competitors. In each comparison, we will examine whether the proposed framework can outperform the competitors over different numbers of selected bands. The following will start with the descriptions of the employed data sets and competitors and then present the comparative results in detail.

A. Data Sets

Three real-world hyperspectral image data, which are gathered by two different remote imaging systems, are used for experimental verifications.

1) Salinas Scene: This image was gathered by the 224-band AVIRIS system in 1998 over Salinas Valley, California. The data consist of 512 × 217 pixels with spectral coverage within 0.4–2.5 μm and are composed of vegetables, bare soils, and vineyard fields with a spatial resolution of 3.7 m. Previous to the practical analysis, 20 atmospheric and water bands, i.e., 108–112 (1.37–1.41 μm), 154–167 (1.83–1.93 μm), and 224 (2.50 μm) were discarded due to the low signal-to-noise ratio. In this paper, the employed Salinas Scene data are the ones with well-labeled ground reference containing 16 classes.[2]

2) Pavia University Scene: Different from the preceding image, the Pavia University data were acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) system, which generates 115 bands ranging from 0.43–0.86 μm with a geometric resolution of 1.3 m. The scene was acquired around the Engineering School at the University of Pavia, northern Italy, and consists of 610 × 340 pixels, each containing 103 bands with 12 noninformational bands discarded. In this paper, the employed Pavia University data are the one with groundtruth land cover map containing nine classes.[3]

3) Indian Pines Scene: This image is also well-known 224-band AVIRIS data collected on June 12, 1992, over the Indian Pines test scene, 6 mi west of West Lafayette. The provided image file is a subset of a larger scene and consists of 145 × 145 pixels with 220 corrected spectral reflectance bands in the wavelength range of 0.4–2.5 μm.[4] It now has been widely used with only 206 bands by removing bands 150–163 (1.79–1.89 μm), which are known as covering the region of atmospheric and water absorption.

B. Competitors

In our comparative experiments, the performance of the proposed framework is verified at the occasions in both hyperspectral classification and color visualization. Two kinds of band selection methods for each application are introduced as competitors. With regard to selecting bands for hyperspectral classification, the proposed method is compared with the following works having four specific implementations. Details of these approaches are as follows.

1) *Constrained Band Selection* (CBS) [29]. This method selects bands by imposing constrained band correlation and dependence minimization and is implemented by two different strategies. One is derived by the concept of the *constrained energy minimization* (CEM). This strategy converts a band image to a vector and linearly constrains a desired band while minimizing interfering effects caused by other bands. Since a band image is generally with a large size, using CEM will be accompanied by the computational problem. In order to cope with this dilemma, an alternative approach that reinterprets CEM-CBS with *linearly constrained minimum variance* (LCMV) has been also introduced in [29]. LCMV-CBS constrains a band image as a matrix and imposes a constraint vector on each of its column. In these two implementations, four specific criteria are employed. They are *band correlation minimization* (BCM), *band correlation constraint* (BCC), *band dependence minimization* (BDM), and *band dependence constraint* (BDC), respectively.
2) *Clustering-Based Band Selection* (CBBS) [9]. CBBS is a hierarchical-clustering-based method using an agglomerative strategy to join similar bands together. In particular, the employed clustering algorithm is based on Ward's linkage, which has the property of guaranteeing grouped partitions with minimum variance in their level of similarity. As a result, CBBS can construct a family of derived groups with minimized intracluster variance and maximized intercluster variance. Then, the bands with the highest average correlations in the corresponding clusters are selected as the final output. In addition, there are two different criteria used for measuring the degree of similarity between two bands: mutual information and Kullback–Leibler divergence. The corresponding two implementations are notated as CBBS-MI and CBBS-KLD, respectively.

In order to assess the performance of the proposed method on the aspect of hyperspectral image visualization, there are also two different band selection methods implemented as benchmarks. Details of these two competitors are as follows.

1) *1BT-Based Visualization* (1BTBV) [18]. This method provides a low-complexity selection for the RGB color display, which utilizes 1BT-based band representation. The initially preserved bands are those with well structure evaluated through the responses of a 1BT filter. Based on this strategy, the finally determined three color components are the ones most dissimilar to all the other well-structured bands.
2) *Information-Measure-Based Visualization* (IMBV) [19]. Dissimilar to the traditional mutual-information-based methods, IMBV employs three information measures of different orders for band selection. This method first excludes irrelevant bands by means of a center-surround entropy comparison. Then, the refined spectrums are segmented to three groups by adaptively thresholding the CIE standard observer color matching function. Finally, a triplet of bands from the segmented groups is determined with maximum second- and third-order informative contents.

C. Comparison for Hyperspectral Classification

Here, a supervised hyperspectral pixel classification process is implemented as an application to evaluate the performance of the proposed band selection method. To this end, four widely used classifiers are employed to produce the comparison results. They are *classification and regression trees* (CART), *k-nearest neighborhood* (k-NN), *naive Bayes*, and *support vector machine* (SVM), respectively. All the produced classification results

[2]http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

[3]http://tlclab.unipv.it/sito_tlc/people.do?id=pgamba

[4]https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html

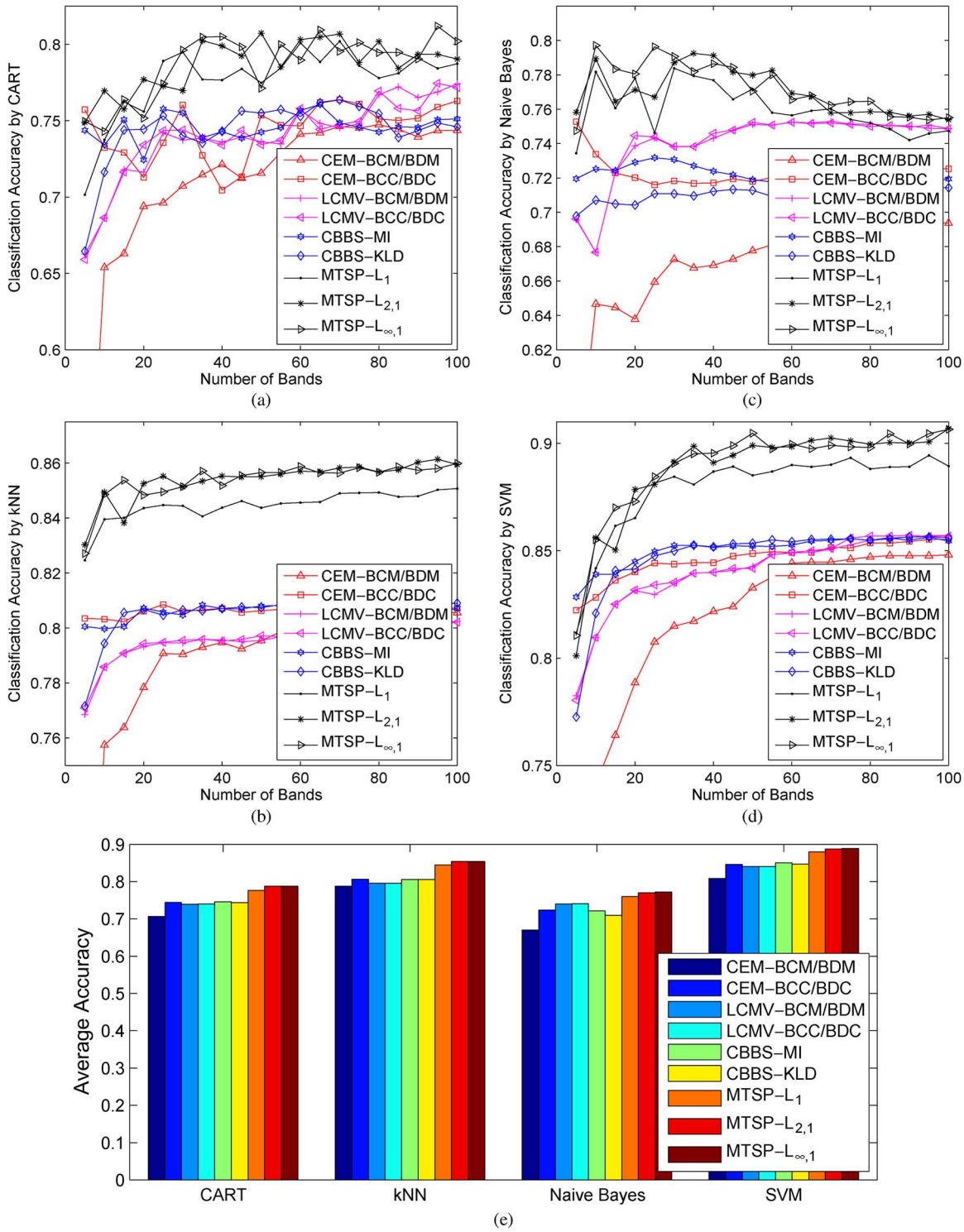

Fig. 4. Classification results of the Salinas scene. (a)–(d) Accuracy curves by CART, k-NN, naive Bayes, and SVM, respectively. (e) Average accuracy bars.

are analyzed by two different metrics: *accuracy curve* and *average accuracy bar*. The accuracy curve is sketched by varying the selected band number m, where m ranges from 5 to 100 with interval 5, and the average accuracy bar is plotted by averaging the classification accuracy rates over all the employed m.

The comparative results for the Salinas scene are illustrated in Fig. 4, where Fig. 4(a)–(e) presents the accuracy curves produced by CART, k-NN, naive Bayes, SVM, and the average accuracy bars, respectively. As observed from Fig. 4(a)–(d), the performances of the proposed MTSP methods clearly dominate the competitors on all the employed classifiers. Among the MTSP methods, MTSP-$\ell_{2,1}$ and MTSP-$\ell_{\infty,1}$ have nearly equivalent performance, while significantly outperforming MTSP-ℓ_1. Moreover, as a more general comparison, in Fig. 4(e), the average accuracy bars of MTSP-$\ell_{\infty,1}$, MTSP-$\ell_{2,1}$, and MTSP-ℓ_1 are with top three positions at the rank of classification performance, which further prove the effectiveness of the proposed framework.

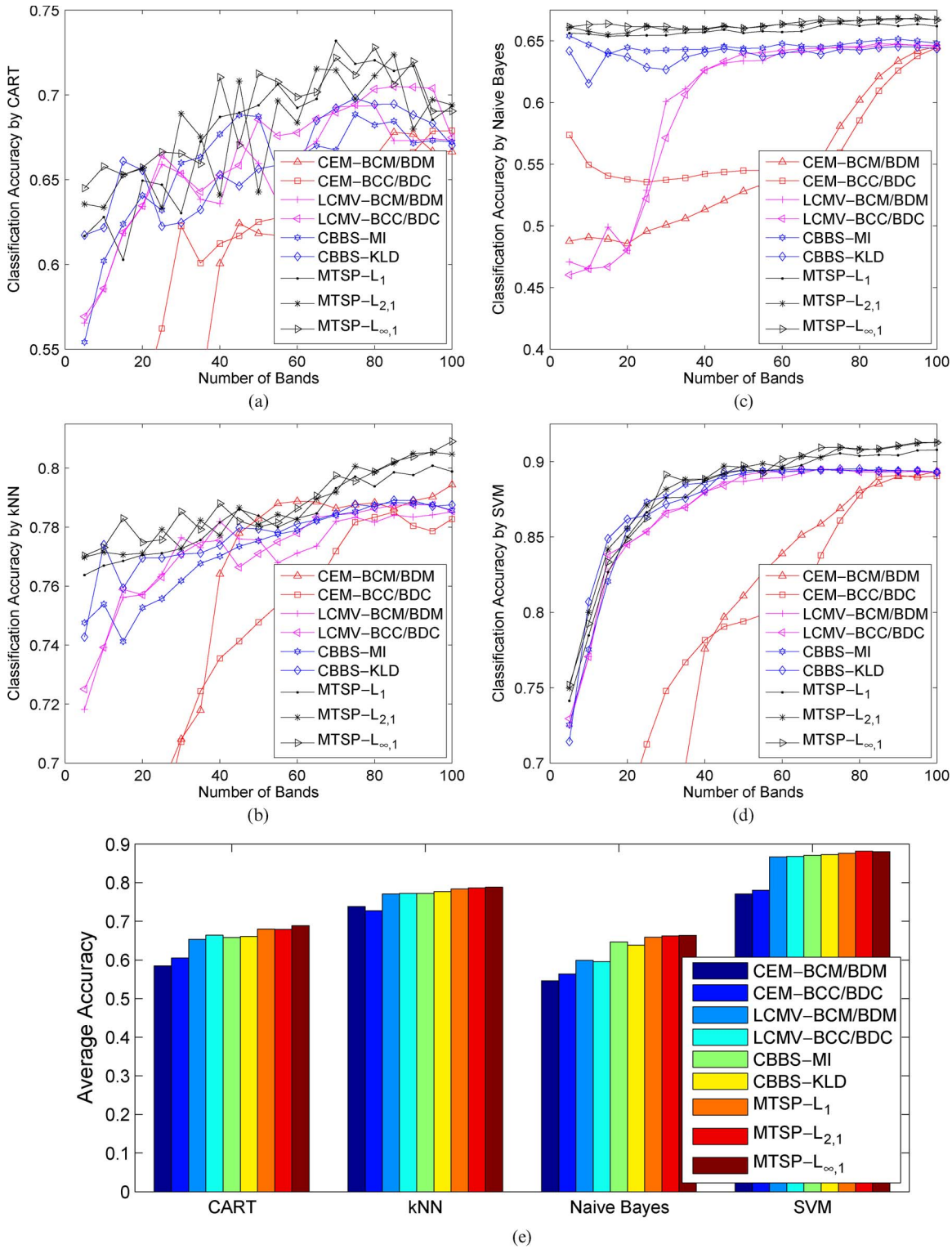

Fig. 5. Classification results of the Pavia University scene. (a)–(d) Accuracy curves by CART, k-NN, naive Bayes, and SVM, respectively. (e) Average accuracy bars.

With regard to the Pavia University scene, the comparison results are shown in Fig. 5 with the same form as in Fig. 4. For this hyperspectral image, as presented in Fig. 5(a) and (c), the proposed MTSP methods have significant advantages when adopting CART and naive Bayes classifiers. However, as shown in Fig. 5(b) and (d), if k-NN or SVM is employed, the accuracy curves can no longer provide clues discriminative enough to support comparative analysis. In such a case, the average accuracy bar will be the more important indicator. The corresponding results are demonstrated in Fig. 5(e). It is manifested that the proposed MTSP-$\ell_{\infty,1}$ and MTSP-$\ell_{2,1}$ clearly dominate all the other methods, and the former is more preferable in most cases. In addition, MTSP-ℓ_1 can also slightly outperform the existing methods, except the equivalent performance when the SVM classifier is utilized.

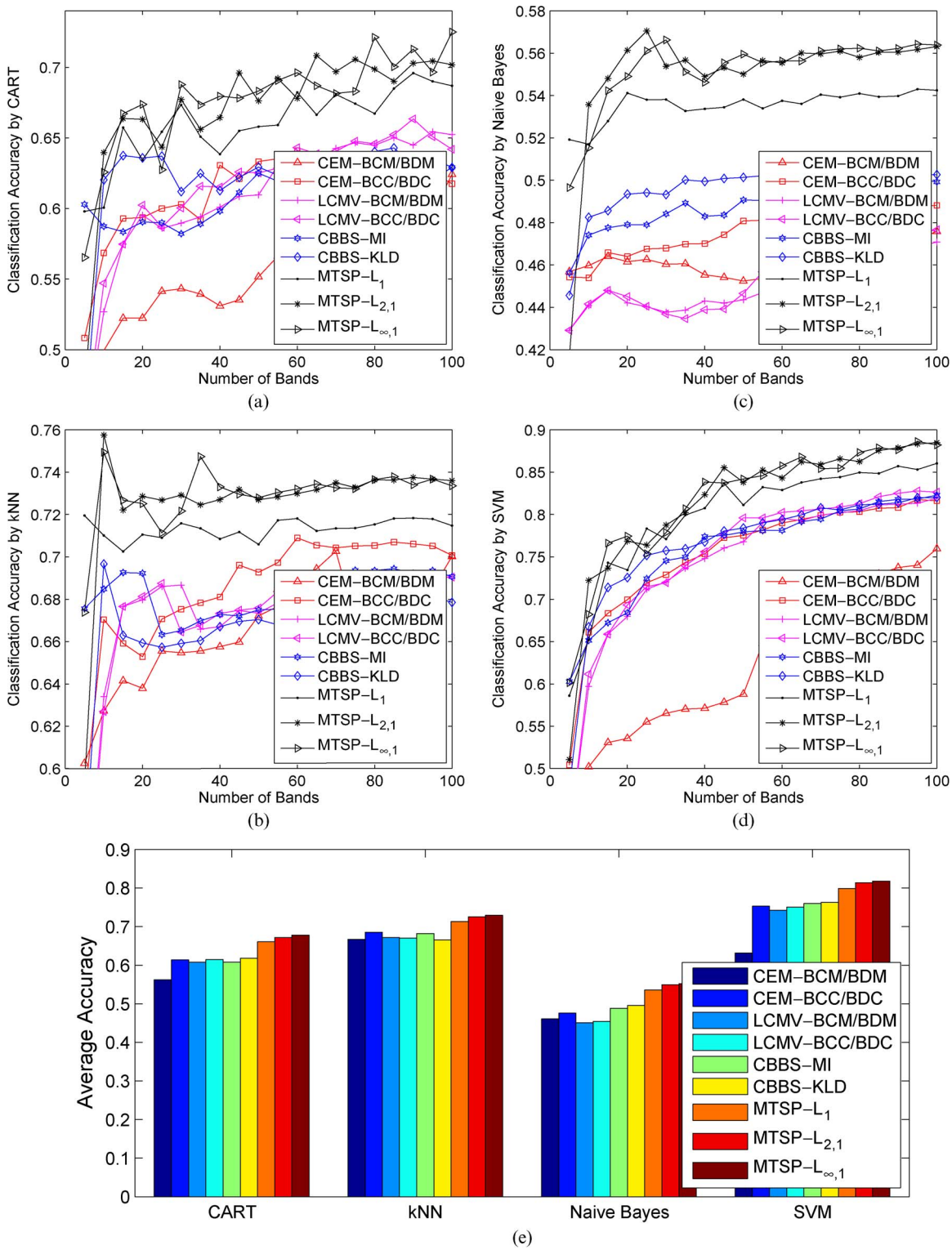

Fig. 6. Classification results of the Indian Pines scene. (a)–(d) Accuracy curves by CART, k-NN, naive Bayes, and SVM, respectively. (e) Average accuracy bars.

For the Indian Pines scene, the comparison results are very similar to those of the Salinas scene. As shown in Fig. 6(a)–(d), the proposed MTSP methods can outperform on all the conducted classification cases compared with the competitors. Therefore, in Fig. 6(e), the MTSP methods achieve significant superiorities in the average accuracy bars. In addition, it is observed that MTSP-$\ell_{\infty,1}$ clearly outperforms MTSP-ℓ_1 and is slightly better than MTSP-$\ell_{2,1}$.

Additionally, a series of statistical tests are also conducted for the classification results. These tests can show whether the classification accuracies of the proposed MTSP methods are better than those of the competitor with statistical significance. Since the distributions of the classification accuracies may be independent and not normal most of the time, the Student's t-test is employed here. The superiorities of the proposed MTSP methods are verified on the 0.05 significance level. If

TABLE I
STATISTICAL SIGNIFICANCE OF THE MTSP METHODS RELATIVE TO THE COMPETITORS OVER THE EMPLOYED HYPERSPECTRAL IMAGES

	CEM		LCMV		CBBS	
	BCM/BDM	BCC/BDC	BCM/BDM	BCC/BDC	MI	KLD
MTSP-ℓ_1	9.2×10^{-16}	1.1×10^{-7}	7.3×10^{-5}	1.8×10^{-4}	1.9×10^{-3}	1.4×10^{-3}
MTSP-$\ell_{2,1}$	2.9×10^{-18}	1.2×10^{-9}	2.1×10^{-6}	6.2×10^{-6}	7.8×10^{-5}	5.4×10^{-5}
MTSP-$\ell_{\infty,1}$	1.4×10^{-19}	1.3×10^{-10}	4.1×10^{-7}	1.3×10^{-6}	1.8×10^{-5}	1.2×10^{-5}

a significance value is less than 0.05, the null hypothesis stating that the performance of our method is not preferable can be rejected. The test results are presented in Table I, which provides each cell a significance value. These results prove that the performances of the proposed MTSP methods are significantly better than those of the competitors.

From the results, one can observe that the proposed MTSP methods with $\ell_{\infty,1}$ and $\ell_{2,1}$ regularizers are much effective than the compared methods. From this result, the following can be inferred: 1) the proposed MTSP-based criterion can be more distinguishable than others; 2) the proposed compressive hyperspectral band descriptor can provide each band an intrinsic description; and 3) the employed computational evolutionary strategy is powerful enough to help provide an ideal solution in our band selection framework. Therefore, the proposed MTSP-based framework has the ability to select the truly valuable bands in a greater probability for further processing.

D. Comparison for Hyperspectral Image Visualization

In the conducted hyperspectral image visualization experiments, four performance evaluation criteria, namely, *interclass perceptual distance* (ICPD) [19], [33], *image entropy* (IE) [51], *averaged gradient* (AG) [51], and *averaged color component correlation* (ACCC), are employed for the quantitative validation. In order to calculate ICPD, each produced false color image is first associated with the corresponding groundtruth land cover map. The following calculation is then divided into three steps. Fist, for each land cover class, all the contained false colors are directly averaged as the centroid color. Next, each centroid color will be projected into a perceptual uniform color space, which is CIELAB color space in our experiments. In the end, the Euclidean distances calculated from the couples of centroid colors are averaged as the final ICPD value. The entropy of the produced false image is also an important measure in the literature [51], which indicates the amount of information contained in the image. IE could provide a valuable clue for the assessment of visualization performance, but it will fail in the case of noise images. To this end, the criterion AG is recommended, which describes the image sharpness in terms of an average over the horizontal and vertical gradient values of the image. In addition, as the natural color images are typically with high correlation between RGB components, the correlation coefficients between RGB components can be also used as a measure for the hyperspectral image visualization methods [18], [52]. For the sake of simplicity, in this paper, the averaged correlation over the pairs of RGB components is calculated as an integrated metric for each visualization result. Larger ICPD, IE, AG, and ACCC values indicate better performances.

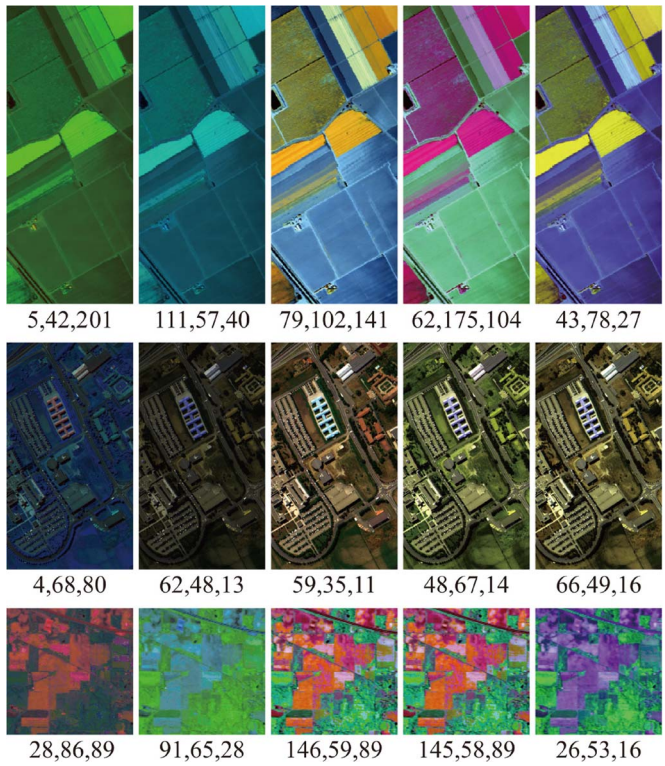

Fig. 7. Hyperspectral image visualization results. From top to bottom, each row corresponds to the Salinas, Pavia University, and Indian Pines scenes, respectively. From left to right, each column presents the false images produced by 1BTBV, IMBV, MTSP-ℓ_1, MTSP-$\ell_{2,1}$, and MTSP-$\ell_{\infty,1}$, respectively. The marked numbers are the indices of selected bands.

The comparison results are presented in Fig. 7 and Table II, where the former illustrates the visual comparisons, and the latter provides the ICPD, IE, AG, and ACCC values. These results show that, for all the processed hyperspectral images, the proposed MTSP methods strongly dominate the others. MTSP-$\ell_{\infty,1}$, MTSP-$\ell_{2,1}$, and MTSP-ℓ_1 can yield false images with more colorful and discriminative appearances and, of course, with significantly higher ICPD, IE, AG and ACCC values.

Additionally, it is interesting that, for both the Indian Pines and Pavia University scenes, the false color images produced by the proposed three MTSP methods are displayed with quite similar appearances. This status is due to the fact that the selected band combinations of these methods are extracted from the neighboring wavelength. However, these band combinations have not conveyed the identical information and therefore yielded different ICPD, IE, AG, and ACCC values. To sum up, among the proposed MTSP methods, employing $\ell_{\infty,1}$ regularizer will help select three bands most preferable for hyperspectral image visualization, and then, $\ell_{2,1}$ and ℓ_1 follow closely behind, compared with other competitors.

TABLE II
QUANTITATIVE COMPARISON OF DIFFERENT HYPERSPECTRAL IMAGE VISUALIZATION METHODS

	Salinas				Pavia University				Indian Pines			
	ICPD	ACCC	IE	AG	ICPD	ACCC	IE	AG	ICPD	ACCC	IE	AG
1BTBV	29.21	0.4720	13.74	3.19	38.24	0.3391	18.74	5.86	27.60	0.1612	17.59	5.60
IMBV	18.23	0.4739	16.15	3.19	27.53	0.8695	18.57	6.86	36.04	0.2182	18.74	6.78
MTSP-ℓ_1	58.49	0.4668	18.36	**8.23**	39.82	0.8851	18.09	14.09	56.46	0.2054	19.33	**15.29**
MTSP-$\ell_{2,1}$	67.86	0.5635	19.55	6.65	**53.30**	0.8754	18.32	15.73	57.47	0.2105	19.54	15.17
MTSP-$\ell_{\infty,1}$	**71.13**	**0.6440**	**19.98**	6.48	40.06	**0.8863**	18.38	**16.79**	**59.62**	**0.6469**	**20.81**	13.21

TABLE III
COMPARISON OF COMPUTATIONAL TIME (S) FOR THE SELECTION OF TEN BANDS

	Salinas	Pavia University	Indian Pines
CEM-BCM/BDM	126.52	63.54	129.44
CEM-BCC/BDC	127.97	64.02	133.96
LCMV-BCM/BDM	25.07	18.82	1.73
LCMV-BCC/BDC	25.05	18.85	2.73
CBBS-MI	76.94	28.56	47.25
CBBS-KLD	759.50	287.54	205.40
MTSP-ℓ_1	43.87	40.41	42.73
MTSP-$\ell_{2,1}$	39.94	41.47	44.93
MTSP-$\ell_{\infty,1}$	40.02	43.23	41.90

TABLE IV
COMPARISON OF COMPUTATIONAL TIME (S) FOR HYPERSPECTRAL IMAGE VISUALIZATION

	Salinas	Pavia University	Indian Pines
1BTBV	17.29	5.07	2.33
IMBV	44.09	42.32	30.28
MTSP-ℓ_1	19.61	12.43	26.85
MTSP-$\ell_{2,1}$	21.56	15.51	23.64
MTSP-$\ell_{\infty,1}$	19.95	13.01	23.77

E. Comparison of Computational Time

In order to evaluate the efficiency of the implemented band selection methods, the computational time of each method are compared. Timings have been taken on an Intel Core i3-550 3.2-GHz central processing unit with 2-GB random access memory. Table III compares the computational time for the selection of ten bands during the classification experiments, and Table IV compares the time cost for each hyperspectral image visualization method. The employed CBBS-MI and CBBS-KLD are implemented in C++, whereas the others are implemented in MATLAB. As a result, in both the two different comparison groups, the proposed MTSP methods take computational times moderately among the competitive algorithms. Although our methods cannot outperform all the competitors in this aspect, they are computationally acceptable while guaranteeing the superior band selection results. In addition, an enhanced hardware configuration could further improve our methods to a faster speed, which might be adequate for real situations. Therefore, from an overall perspective, the proposed MTSP methods can claim their value in the practical applications.

F. Verification for the Compressive Descriptor

In practice, performing random projection has to preserve the neighborhood structure of the original descriptions. In order to verify whether the proposed compressive descriptor satisfies this requirement, the Kendall's rank correlation analysis [53] is employed. This analysis starts with the construction of two types of nearest neighbor ranks for each band. One rank is determined according to the similarities based on the raw image information, and the other one is constructed using the compressed descriptions. Then, for each pair of nearest neighbor ranks, a Kendall's rank correlation coefficient is calculated to reflect their consistency. Larger correlation indicates better compressed description. The analysis results are illustrated in Fig. 8, which plots the statistics of the Kendall's rank correlations versus the dimensionality of the compressive descriptor. It is manifested that the compressive descriptor can provide the nearly perfect preservation for the neighborhood structures when the dimensionality reaches 200. Therefore, in this paper, k is fixed to 200 in all our comparison experiments.

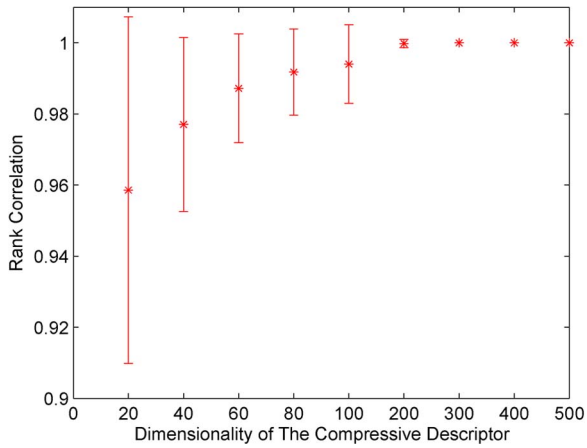

Fig. 8. Statistical comparison of the rank correlations for different dimensional compressive descriptors. The asterisks denote the mean values, and the distances above and below the asterisks reflect the standard deviations.

V. CONCLUSION

The large and growing volume of redundancy carried by hyperspectral image has caused increasing attention in both academia and industry. With regard to this issue, the traditional feature-extraction-based methods fail due to the weakness in preserving the physical characteristics of the original spectral channels. Differently, the problem addressed in this paper is the selection of most representative bands to preserve the original information of the data. This technique can reduce the redundancy of hyperspectral image without compromising and distorting the raw information in the selected bands.

The groupwise band selection framework proposed in this paper is fully unsupervised and computationally acceptable, which is based on the following functional components: 1) an

MTSP-based criterion for evaluating the representativeness of each band combination; 2) a compressive descriptor for efficiently and effectively describing the hyperspectral bands; and 3) a computational evolutionary strategy for quickly and reliably searching for the desired band combination from numerous possible solutions.

Experiments have been conducted on both hyperspectral classification and color visualization, which prove that, in these two applications, the proposed framework is more robust and reliable than the other competitors representing state of the art. Particularly, when the joint sparsity constraint is imposed into the proposed framework, i.e., when $\ell_{\infty,1}$ or $\ell_{2,1}$ is regularized, the performance is dominant at all the employed indicators in most cases. This result means that the proposed framework with joint sparsity constraint can select the representative bands more critical and appropriate for the validated applications. Compared with traditional pointwise-selection-based methods, the main drawback of the proposed framework is the computational time required by the global searching process, which may reach several tens of seconds. In the future, we will develop computationally more efficient implementations by resorting to parallel computer architectures.

REFERENCES

[1] F. Santini, L. Alberotanza, R. M. Cavalli, and S. Pignatti, "A two-step optimization procedure for assessing water constituent concentrations by hyperspectral remote sensing techniques: An application to the highly turbid Venice lagoon waters," *Remote Sens. Environ.*, vol. 114, no. 4, pp. 887–898, Apr. 2010.

[2] B. Luo, C. Yang, J. Chanussot, and L. Zhang, "Crop yield estimation based on unsupervised linear unmixing of multidate hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 162–173, Jan. 2013.

[3] H. Akbari, Y. Kosugi, K. Kojima, and N. Tanaka, "Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 2011–2017, Aug. 2010.

[4] A. Prieto, F. Bellas, R. J. Duro, and F. López-Peña, "An adaptive approach for the progressive integration of spatial and spectral features when training ground-based hyperspectral imaging classifiers," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 8, pp. 2083–2093, Aug. 2010.

[5] Q. Du, J. E. Fowler, and W. Zhu, "On the impact of atmospheric correction on lossy compression of multispectral and hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 130–132, Jan. 2009.

[6] H. Chen, Y. Zhang, J. Zhang, and Y. Chen, "A BOI-preserving-based compression method for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 3913–3923, Nov. 2010.

[7] O. Kuybeda, D. Malah, and M. Barzohar, "Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5579–5592, Dec. 2007.

[8] L. O. Jimenez and D. A. Landgrebe, "Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 28, no. 1, pp. 39–54, Feb. 1998.

[9] A. M. Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.

[10] S. Kumar, J. Ghosh, and M. M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1368–1379, Jul. 2001.

[11] S. B. Serpico and G. Moser, "Extraction of spectral channels from hyperspectral images for classification purposes," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 484–495, Feb. 2007.

[12] L. O. Jimenez-Rodriguez, E. Arzuaga-Cruz, and M. Velez-Reyes, "Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 469–483, Feb. 2007.

[13] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.

[14] C. Vaiphasa, A. K. Skidmore, W. F. de Boer, and T. Vaiphasa, "A hyperspectral band selector for plant species discrimination," *Photogramm. Remote Sens.*, vol. 62, no. 3, pp. 225–235, Aug. 2007.

[15] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. 110–122, Sep. 2009.

[16] W. Di, L. Zhang, D. Zhang, and Q. Pan, "Studies on hyperspectral face recognition in visible spectrum with feature band selection," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 6, pp. 1354–1361, Nov. 2010.

[17] Y. Zhao, L. Zhang, and S. G. Kong, "Band-subset-based clustering and fusion for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 747–756, Feb. 2011.

[18] B. Demir, A. Çelebi, and S. Ertürk, "A low-complexity approach for the color display of hyperspectral remote-sensing images using one-bit-transform-based band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 97–105, Jan. 2009.

[19] S. L. Moan, A. Mansouri, Y. Voisin, and J. Y. Hardeberg, "A constrained band selection method based on information measures for spectral image color visualization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5104–5115, Dec. 2011.

[20] T. Zhou and D. Tao, "Double shrinking sparse dimension reduction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 244–257, Jan. 2013.

[21] T. Zhou, D. Tao, and X. Wu, "Manifold elastic net: A unified framework for sparse dimension reduction," *Data Mining Knowl. Discov.*, vol. 22, no. 3, pp. 340–371, May 2011.

[22] T. Zhou and D. Tao, "GoDec: Randomized low rank & sparse matrix decomposition in noisy case," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 33–40.

[23] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.

[24] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.

[25] X. Jia, B.-C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, Mar. 2013.

[26] X. Chen, W. Pan, J. T. Kwok, and J. G. Carbonell, "Accelerated gradient method for multi-task sparse learning problem," in *Proc. Int. Conf. Data Mining*, 2009, pp. 746–751.

[27] P. Tseng, "Approximation accuracy, gradient methods, and error bound for structured convex optimization," *Math. Programm.*, vol. 125, no. 2, pp. 263–295, Oct. 2010.

[28] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, Jan. 2011.

[29] 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.

[30] 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.

[31] H. Su, H. Yang, Q. Du, and Y. Sheng, "Semisupervised band clustering for dimensionality reduction of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1135–1139, Nov. 2011.

[32] J. Yin, Y. Wang, and J. Hu, "A new dimensionality reduction algorithm for hyperspectral image using evolutionary strategy," *IEEE Trans. Ind. Informat.*, vol. 8, no. 4, pp. 935–943, Nov. 2012.

[33] Q. Du, N. Raksuntorn, S. Cai, and R. J. Moorhead, "Color display for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1858–1866, Jun. 2008.

[34] H. Su, Q. Du, and P. Du, "Hyperspectral image visualization using band selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, to be published.

[35] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[36] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[37] E. J. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, 2006, pp. 1433–1452.

[38] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[39] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[40] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construct. Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.
[41] D. Achlioptas, "Database-friendly random projections: Johnson–Lindenstrauss with binary coins," *J. Comput. Syst. Sci.*, vol. 66, no. 4, pp. 671–687, Jun. 2003.
[42] K. Zhang, L. Zhang, and M.-H. Yang, "Real-time compressive tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 864–877.
[43] P. Li, T. Hastie, and K. W. Church, "Very sparse random projections," in *Proc. ACM SIGKDD Int. Conf. Mining*, 2006, pp. 287–296.
[44] Q. Liu, X. Liao, H. Li, J. R. Stack, and L. Carin, "Semisupervised multitask learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1074–1086, Jun. 2009.
[45] G. Zhu, S. Yan, and Y. Ma, "Image tag refinement towards low-rank, content-tag prior and error sparsity," in *Proc. ACM Multimedia*, 2010, pp. 461–470.
[46] X. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2010, pp. 3493–3500.
[47] J. Li, Y. Tian, T. Huang, and W. Gao, "Multi-task rank learning for visual saliency estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 5, pp. 623–636, May 2011.
[48] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via structured multi-task sparse learning," *Int. J. Comput. Vis.*, vol. 101, no. 2, pp. 367–383, Jan. 2013.
[49] Q. Du, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral band selection using a collaborative sparse model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 3054–3057.
[50] L. N. de Castro and F. J. V. Zuben, "Learning and optimization using the clonal selection principle," *IEEE Trans. Evol. Comput.*, vol. 6, no. 3, pp. 239–251, Jun. 2002.
[51] K. Kotwal and S. Chaudhuri, "Visualization of hyperspectral images using bilateral filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2308–2316, May 2010.
[52] S. Cai, Q. Du, R. J. Moorhead, M. J. Mohammadi-Aragh, and D. Irby, "Noise-adjusted principle component analysis for hyperspectral remotely sensed imagery visualization," in *Proc. IEEE Visualization*, 2005, pp. 119–120.
[53] M. G. Kendall, *Rank Correlation Methods*. London, U.K.: Griffin, 1948.

Yuan Yuan (M'05–SM'09) is currently a Full Professor with the Chinese Academy of Sciences, Beijing, China. She has authored or coauthored over 150 papers, including about 100 in reputable journals such as IEEE Transactions and *Pattern Recognition*, as well as conference papers in CVPR, BMVC, ICIP, and ICASSP. Her current research interests include visual information processing and image/video content analysis.

Guokang Zhu is currently working toward the Ph.D. degree in the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.

His current research interests include computer vision and machine learning.

Qi Wang received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent system from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively.

He is currently an Associate Professor with the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His current research interests include computer vision and pattern recognition.