Spatial-Spectral Kernel Sparse Representation for Hyperspectral Image Classification

Jianjun Liu, Zebin Wu, Zhihui Wei, Liang Xiao, and Le Sun

***Abstract*—Kernel sparse representation classification (KSRC), a nonlinear extension of sparse representation classification, shows its good performance for hyperspectral image classification. However, KSRC only considers the spectra of unordered pixels, without incorporating information on the spatially adjacent data. This paper proposes a neighboring filtering kernel to spatial-spectral kernel sparse representation for enhanced classification of hyperspectral images. The novelty of this work consists in: 1) presenting a framework of spatial-spectral KSRC; and 2) measuring the spatial similarity by means of neighborhood filtering in the kernel feature space. Experiments on several hyperspectral images demonstrate the effectiveness of the presented method, and the proposed neighboring filtering kernel outperforms the existing spatial-spectral kernels. In addition, the proposed spatial-spectral KSRC opens a wide field for future developments in which filtering methods can be easily incorporated.**

***Index Terms*—Classification, kernel sparse representation, hyperspectral image, spatial-spectral kernel.**

I. Introduction

In the hyperspectral case, the remote sensors capture digital images in hundreds of narrow and contiguous spectral bands spanning the visible to infrared spectrum [1]. The wealth of spectral information has opened new perspectives in many application domains, such as military [2], [3], agriculture [4], [5], and mineralogy [6]. A very important application of hyperspectral images is image classification where pixels are labeled to one of the classes. Various supervised and unsupervised methods have been developed for hyperspectral image classification (e.g., maximum-likelihood classifiers, neural networks, etc.) [7]–[10]. Kernel methods, such as support vector machines (SVM), have been a powerful tool to solve classification problems for high-dimensional data and have shown a good performance for hyperspectral image classification [11].

Recently, as an advanced technique, sparse representation has been widely used in various fields, such as image denoising [12], image super-resolution [13], face recognition [14], spectral unmixing [15] and hyperspectral image classification [16]. In these applications, the usage of sparsity prior often leads to state-of-the-art performance. Sparse representation classification (SRC) is a nonparametric learning method which does not need a training process but does need training data, and can directly assign a class label to a test sample. As an effective combination of machine learning and compressed sensing [17], SRC assumes that the features in the same class lie in the same low-dimensional subspace [14]. Therefore, an unknown sample can be sparsely represented by a linear combination of a few training samples from a structured dictionary. Although SRC gets good performance in hyperspectral image classification, it is hard to classify the data that is not linearly separable. To overcome this drawback, kernel SRC (KSRC) is proposed in [18] to capture the nonlinear similarity of features. In machine learning, the kernel trick is originally used to construct nonlinear SVM [19]. Similarly, the kernel trick in KSRC is applied to project the data into a feature space in which the data become linearly separable.

It is worth noting that hyperspectral image classification should not only focus on analyzing the spectral features but also incorporate information on the spatially adjacent pixels. That means we should treat hyperspectral data as images, not only a mere collection of independent and identically distributed pixels [20]. Previous approaches have shown a good performance by taking into account the contextual information in hyperspectral image classification. For instance, in [16], the spatial smoothness is considered as the prior information and two methods (the local smoothing constraint and the joint sparsity model) are introduced to SRC, and in [21], the joint sparsity model is introduced to KSRC.

Nevertheless, the original KSRC only treats hyperspectral data as unordered pixels without considering its image characteristics. According to the above introduction, the joint sparsity model can be used to measure the spatial similarity. However, this model is sensitive to the scale of training dictionary. In other words, sufficient training samples are necessary, while hyperspectral image classification faces a challenging problem of low number of training samples. To improve the classification accuracy, a straightforward way is to adopt kernel methods. It has been taken in SVM where variations of the kernels have been proposed and the good performance has been demonstrated. In [22], Camps-Valls *et al.* present a framework of composite kernels (CK) that simultaneously takes into account spectral, spatial, and local cross-information in hyperspectral images. In order to increase the reliability of the training phase when few training samples are available, Marconcini *et al.* in [23] exploit unlabeled data in a semi-supervised SVM by taking the

Manuscript received November 18, 2012; revised January 22, 2013; accepted March 04, 2013. Date of publication April 04, 2013; date of current version November 21, 2013. This work has been supported in part by the National Natural Science Foundation of China (Grant No.61101194, 61071146, 61171165), in part by the Jiangsu Provincial Natural Science Foundation of China (Grant No. BK2011701), in part by the Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20113219120024), and in part by the Project of China Geological Survey (Grant No. 1212011120227).

The authors are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: liuofficial@163.com; zebin.wu@gmail.com; gswei@mail.njust.edu.cn; xtxiaoliang@mail.njust.edu.cn; sunlecncom@163.com).

Digital Object Identifier 10.1109/JSTARS.2013.2252150

advantage of CK. Beyond the CK, other kernels that exploit the wealth of unlabeled samples have been proposed, such as cluster kernels [24], [25], mean map kernels [26]. And in [27], graph kernels are introduced to learn high order similarities between neighboring samples. In addition, another way to improve the classification accuracy has been proposed in [28], where Flamary *et al.* learn the spatial filters by an SVM maximizing the margin in the filtered domain.

In this paper, we propose a framework of spatial-spectral kernel sparse representation for hyperspectral image classification. The convexity of the objective function of spatial-spectral KSRC depends on the definition of spatial-spectral kernels. Thanks to the Mercer's condition [29], this convexity is guaranteed. As mentioned above, several kernels that follow the Mercer's condition have been proposed in SVM, and all these kernels can be applied to the spatial-spectral KSRC. However, as discussed in [27], the methodology of previous kernels still relies on performing a spatial feature extraction before kernel computation, typically limited to second-order statistics or morphological operators [30], [31]. Given this, we develop a neighboring filtering (NF) kernel to capture the spatial similarity between neighboring samples in the kernel feature space. Similar to the techniques in image processing, first, the neighborhood filtering [32] is investigated to generate the mapping of the pixels in the feature space, and then, a kernel is built on the estimated vectors. Unlike the CK methods, this method needs only one kernel where both spectral and spatial similarities are simultaneously computed. It is worth noting that the involved kernel extends the mean map kernel [26] and generalizes the CK framework approach by computing similarities in a proper feature space. In addition, by analyzing the relationship of KSRC and SRC, a general algorithm for the spatial-spectral KSRC is proposed.

The remainder of this paper is organized as follows. Section II briefly reviews the formulations of SRC and KSRC, and introduces the kernel methods. In Section III, we present the spatial-spectral KSRC, and give a general algorithm for this kind of methods. The effectiveness of the proposed method is demonstrated in Section IV by experiments on several real hyperspectral images. Finally, Section V concludes this paper.

II. Sparsity-Based Classification and Kernel Methods

In this section, we first introduce SRC and KSRC, and then discuss the relationships between them. Finally, the kernel methods are briefly reviewed.

A. Sparse Representation Classification

Like k-nearest neighbor [33] and nearest subspace [34], SRC is a nonparametric learning method, which directly assigns a class label to a test sample without a training process. That is to say, it does not need a set of hypothesis functions and to learn the weight vectors of the hypothesis functions. SRC can be treated as a learning machine in which the classification process is implemented by using the signal reconstruction methods. From this point of view, SRC is an effective combination of compressed sensing and machine learning [17].

SRC has been successfully applied to face recognition and hyperspectral image classification, which assumes that the features belonging to the same class approximately lie in the same low-dimensional subspace [14], [16]. Assume that we have an unknown test sample $\mathbf{x} \in \mathbb{R}^L$ (L is the number of spectral bands), then it can be written as a sparse linear combination of all the training pixels as

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{L\times J}$ denotes a structured dictionary whose columns $\{\mathbf{a}_i\}_{i=1,2,\ldots,J}$ are composed of the J training samples from all classes, and $\mathbf{s} \in \mathbb{R}^J$ is an unknown sparse vector. Suppose the dictionary $\mathbf{A}$ is given, the vector $\mathbf{s}$ can be obtained by solving the following optimization problem

$$\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_1, \tag{2}$$

where $\lambda > 0$ is used to control the sparsity of $\mathbf{s}$. The class label of $\mathbf{x}$ is determined by the minimal residual between $\mathbf{x}$ and its approximation from each class sub-dictionary

$$\text{class}(\mathbf{x}) = \arg\min_{i=1,\ldots,c} \|\mathbf{x} - \mathbf{A}\delta_i(\mathbf{s})\|_2, \tag{3}$$

where c is the number of classes, and $\delta_i(\cdot)$ is the characteristic function that selects coefficients related with the ith class and makes the rest to zero.

B. Kernel Sparse Representation Classification

If the data set is not linearly separable, the assumption of SRC is not established. In this case, the performance of SRC may decrease. As an extension of SRC, KSRC uses the kernel trick to project the data into a feature space, in which the data become linearly separable [18], [21].

Suppose there exists a feature mapping function $\phi : \mathbb{R}^L \to \mathbb{R}^K$, ($L \ll K$). It maps the feature (i.e., a test sample) and dictionary to the high dimensional feature space: $\mathbf{x} \to \phi(\mathbf{x})$, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_J] \to \Phi(\mathbf{A}) = [\phi(\mathbf{a}_1), \phi(\mathbf{a}_2), \ldots, \phi(\mathbf{a}_J)]$. After substituting the mapped features and dictionary to (2), we arrive at KSRC:

$$\min_{\mathbf{s}} \frac{1}{2}\|\phi(\mathbf{x}) - \Phi(\mathbf{A})\mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_1. \tag{4}$$

Similarly, the class label of $\mathbf{x}$ is determined as

$$\text{class}(\mathbf{x}) = \arg\min_{i=1,\ldots,c} \|\phi(\mathbf{x}) - \Phi(\mathbf{A})\delta_i(\mathbf{s})\|_2. \tag{5}$$

It is worth noting that all ϕ mappings used in KSRC occur in the form of inner products. This allows us to define a kernel function $\mathbf{k}$

$$\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \langle\phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle, \tag{6}$$

for any samples $\mathbf{x}_i \in \mathbb{R}^L$ and then KSRC can be constructed using only the kernel function, without having to consider the mapping ϕ explicitly. Then, by introducing (6) into (4), the optimization problem can be rewritten as:

$$\min_{\mathbf{s}} \frac{1}{2}\mathbf{s}^T\mathbf{Q}\mathbf{s} - \mathbf{s}^T\mathbf{p} + \lambda\|\mathbf{s}\|_1 + C, \tag{7}$$

where $C = (1/2)\mathbf{k}(\mathbf{x}, \mathbf{x})$ is a constant, $\mathbf{Q}$ is a $J \times J$ matrix with $\mathbf{Q}_{ij} = \mathbf{k}(\mathbf{a}_i, \mathbf{a}_j)$, and $\mathbf{p}$ is a $J \times 1$ vector with $\mathbf{p}_i = \mathbf{k}(\mathbf{a}_i, \mathbf{x})$. Review the formulation of (2), we may note that the objective of KSRC is the same as that of SRC except for the definition of $\mathbf{Q}$ and $\mathbf{p}$, while in SRC they are replaced by $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\mathbf{x}$. Analogously, the classification criterion (5) can be rewritten as

$$\text{class}(\mathbf{x}) = \arg \min_{i=1,\ldots,c} \delta_i^T(\mathbf{s})\mathbf{Q}\delta_i(\mathbf{s}) - 2\delta_i^T(\mathbf{s})\mathbf{p}. \quad (8)$$

C. Kernel Methods

In the context of KSRC in particular and kernel methods in general, one can use some proper kernel functions that accurately reflect the similarity among samples. However, not all metric distances are suitable. In fact, valid kernels are only those satisfying the Mercer's condition [26], [29]. Some commonly used kernels in kernel methods are the linear kernel ($\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$), polynomial kernel ($\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + 1)^d, d \in \mathbb{Z}^+$), and Gaussian radial basis function (RBF) kernel ($\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|_2^2), \gamma \in \mathbb{R}^+$).

The Mercer's condition constitutes the key requirement to obtain a unique global solution when solving KSRC. In other words, a valid kernel means that it is positive definite, at least as positive semidefinite. And in this case, KSRC is a convex optimization problem. In addition, some relevant properties for Mercer's kernels can be derived from the fact that they are positive semidefinite matrix, as follows.

Let $\mathbf{k}_1$ and $\mathbf{k}_2$ be valid Mercer's kernels over $\mathcal{X} \times \mathcal{X}$, with $\mathbf{x}_i \in \mathcal{X} \subseteq \mathbb{R}^L$, and $\nu > 0$. Then, the direct sum $\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{k}_1(\mathbf{x}_i, \mathbf{x}_j) + \mathbf{k}_2(\mathbf{x}_i, \mathbf{x}_j)$, tensor product $\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{k}_1(\mathbf{x}_i, \mathbf{x}_j) \cdot \mathbf{k}_2(\mathbf{x}_i, \mathbf{x}_j)$, or scaling $\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \nu\mathbf{k}_1(\mathbf{x}_i, \mathbf{x}_j)$ are valid Mercer's kernels [22].

III. Spatial-Spectral Kernel Sparse Representation

The kernel matrices can be precomputed at the very beginning of the minimization procedure, and thus, one usually works with the transformed input data $\mathbf{Q}$ and $\mathbf{p}$ rather than the original input space dictionary $\mathbf{A}$ and samples $\mathbf{x}$. This fact allows us to design kernels taking advantage of the properties of Mercer's kernels. And the performance of these kernels strongly depends on the adequate definition of the kernel structural form. Though the good performance offered by the typical RBF kernel, the underlying data structure is obviated. A suitable kernel is a kernel whose structure reflects data relations. To properly define such a suitable kernel, unlabeled information and geometrical relationships between labeled and unlabeled samples may be useful [24]. In this section, the spatial-spectral kernel sparse representation is proposed. First, we pay attention to the design of a suitable kernel, and then present the NF kernel. Finally, an optimization algorithm for this kind of methods is given.

A. Neighboring Filtering Kernel

A full family of CK for the combination of spectral and contextual information has been presented in SVM [22]–[26]. These kernels are valid, and they are all suitable for KSRC. Though one can improve the performance of KSRC, several limitations should be noted. As the discussion in [27], the kernel should learn all high order similarities between neighboring samples directly, and reflect the data lie in complex manifolds. For these, we propose the NF kernel, which computes the spatial similarity between neighboring samples in the feature space.

Given $\mathbf{x}^m \in \Omega, m = 1, 2, \ldots, \omega^2$, with Ω being the spatial window ω around pixel $\mathbf{x}$. Let $\phi(\mathbf{x}^m)$ be the image of $\mathbf{x}^m$ under the map ϕ. In order to describe $\phi(\mathbf{x})$, a straightforward way is to use the average of the spatially neighboring pixels in the kernel space. This method is the same as the mean filtering in image denosing [35], and the estimated vector is

$$\text{MF}(\phi(\mathbf{x})) = \frac{1}{\omega^2} \sum_{m=1}^{\omega^2} \phi(\mathbf{x}^m). \quad (9)$$

However, the mean filtering rarely reflects relative contributions (which treats every neighboring pixels equally). Given this, we consider the neighborhood filtering, i.e.,

$$\text{NF}(\phi(\mathbf{x})) = \frac{1}{\sum_m \mathbf{w}^m} \sum_{m=1}^{\omega^2} \mathbf{w}^m \phi(\mathbf{x}^m), \quad (10)$$

where $\mathbf{w}^m = \exp(-\gamma_0\|\mathbf{x} - \mathbf{x}^m\|_2^2)$ and the parameter $\gamma_0 > 0$ acts as a degree of filtering.

Let us now consider two different pixels $\mathbf{x}_i$ and $\mathbf{x}_j$. We are interested in defining a similarity function that estimates the proximity between them in a sufficiently rich feature space. A straightforward kernel function reflecting the similarity between them is obtained by evaluating the kernel function between the estimated vectors

$$\begin{aligned} \mathbf{k}_{\text{NF}}(\mathbf{x}_i, \mathbf{x}_j) &= \langle \text{NF}(\phi(\mathbf{x}_i)), \text{NF}(\phi(\mathbf{x}_j)) \rangle \\ &= \left\langle \frac{\sum_{m=1}^{\omega^2} \mathbf{w}_i^m \phi(\mathbf{x}_i^m)}{\sum_m \mathbf{w}_i^m}, \frac{\sum_{n=1}^{\omega^2} \mathbf{w}_j^n \phi(\mathbf{x}_j^n)}{\sum_n \mathbf{w}_j^n} \right\rangle \\ &= \frac{\sum_{m=1}^{\omega^2} \sum_{n=1}^{\omega^2} \mathbf{w}_i^m \mathbf{w}_j^n \mathbf{k}(\mathbf{x}_i^m, \mathbf{x}_j^n)}{\sum_m \mathbf{w}_i^m \sum_n \mathbf{w}_j^n}, \end{aligned} \quad (11)$$

which we call the NF kernel. Similarly, we can get the mean filtering (MF) kernel for the mean filtering as follows

$$\begin{aligned} \mathbf{k}_{\text{MF}}(\mathbf{x}_i, \mathbf{x}_j) &= \langle \text{MF}(\phi(\mathbf{x}_i)), \text{MF}(\phi(\mathbf{x}_j)) \rangle \\ &= \left\langle \frac{1}{\omega^2} \sum_{m=1}^{\omega^2} \phi(\mathbf{x}_i^m), \frac{1}{\omega^2} \sum_{n=1}^{\omega^2} \phi(\mathbf{x}_j^n) \right\rangle \\ &= \frac{1}{\omega^4} \sum_{m=1}^{\omega^2} \sum_{n=1}^{\omega^2} \mathbf{k}(\mathbf{x}_i^m, \mathbf{x}_j^n), \end{aligned} \quad (12)$$

which computes the spatial similarity between neighboring samples, while the cluster similarity is computed in the mean map kernel [26].

B. Optimization Algorithm

Since $\mathbf{Q}$ is a valid kernel, the objective of (7) is convex, which is the same as the objective of (2) except for the definition of $\mathbf{Q}$ and $\mathbf{p}$. So the alternating direction method of multipliers (ADMM) [36], [37] used for (2) can be applied to solve (7).

After introducing a new variable $\mathbf{u} \in \mathbb{R}^J$ to (7), the objective can be rewritten as

$$\begin{aligned} \min_{\mathbf{s},\mathbf{u}} \quad & \frac{1}{2}\mathbf{s}^T\mathbf{Q}\mathbf{s} - \mathbf{s}^T\mathbf{p} + \lambda\|\mathbf{u}\|_1 \\ \text{s.t.} \quad & \mathbf{u} = \mathbf{s}. \end{aligned} \tag{13}$$

The constraint $\mathbf{u} = \mathbf{s}$ is enforced using ADMM defined as

$$\begin{cases} (\mathbf{s}^{(t+1)}, \mathbf{u}^{(t+1)}) = \arg\min_{\mathbf{s},\mathbf{u}} \frac{1}{2}\mathbf{s}^T\mathbf{Q}\mathbf{s} - \mathbf{s}^T\mathbf{p} + \lambda\|\mathbf{u}\|_1 \\ \qquad\qquad + \frac{\mu}{2}\|\mathbf{s} - \mathbf{u} - \mathbf{d}^{(t)}\|_2^2 \\ \mathbf{d}^{(t+1)} = \mathbf{d}^{(t)} - (\mathbf{s}^{(t+1)} - \mathbf{u}^{(t+1)}) \end{cases} \tag{14}$$

where $t \geq 0$ and $\mu > 0$. The minimizing solution $\mathbf{s}^{(t+1)}$ is simply:

$$\mathbf{s}^{(t+1)} \leftarrow (\mathbf{Q} + \mu\mathbf{I})^{-1}\left(\mathbf{p} + \mu(\mathbf{u}^{(t)} + \mathbf{d}^{(t)})\right), \tag{15}$$

where $\mathbf{I}$ is the identity matrix. The minimizing solution $\mathbf{u}^{(t+1)}$ is the well-known soft threshold [38]:

$$\mathbf{u}^{(t+1)} \leftarrow \text{ soft } (\mathbf{s}^{(t+1)} - \mathbf{d}^{(t)}, \lambda/\mu), \tag{16}$$

where $\text{soft}(\cdot, \tau)$ denotes the component-wise application of the soft-threshold function $y \leftarrow \text{ sign } (y)\max\{|y| - \tau, 0\}$.

Algorithm 1 Spatial-Spectral Kernel Sparse Representation Classification Algorithm

1: **Input**: A training dictionary $\mathbf{A} \in \mathbb{R}^{L\times J}$, and a test sample $\mathbf{x} \in \mathbb{R}^L$.
2: Select the Mercer kernel $\mathbf{k}_{NF}$ (or others) and its parameters.
3: Compute the matrix $\mathbf{Q}$, and the vector $\mathbf{p}$.
4: Set $t = 0$, choose $\mu > 0$, $\mathbf{s}^{(0)}$,$\mathbf{u}^{(0)}$,$\mathbf{d}^{(0)}$.
5: **repeat**
6: Compute $\mathbf{s}^{(t+1)}$, $\mathbf{u}^{(t+1)}$, and $\mathbf{d}^{(t+1)}$ using (14)
7: $t \leftarrow t + 1$
8: **until** some stopping criterion is satisfied.
9: Compute the c residuals $r_i(\mathbf{x}) = \delta_i^T(\mathbf{s})\mathbf{Q}\delta_i(\mathbf{s}) - 2\delta_i^T(\mathbf{s})\mathbf{p}$, $i = 1, \ldots, c$.
10: **Output**: The estimated label of $\mathbf{x}$ according to (8).

The complete classification procedure of the proposed spatial-spectral KSRC is shown in Algorithm 1. As well as [27], the main problem is that the computational cost increases with the window size in $\mathcal{O}(\omega^4)$, which can be unaffordable for large window sizes. However, one can build the largest kernel matrix possible firstly, and then derive the lower scales from it through properly indexing kernel entries, without needing to recompute the whole kernel matrix. As for the huge hyperspectral images, one can divide them into several blocks by columns, and then build the corresponding large matrices. And it is also the key step for a successful implementation of the advanced parallel technique. Once the kernels are computed, the spatial-spectral KSRC is the same as the standard KSRC. Unlike the general machine learning methods, such as SVM, KSRC does not involve a training stage. It solves a convex optimization problem for each test pixel. Therefore, KSRC is more computationally intensive than SVM.

IV. Experiential Results and Analysis

In this section, we demonstrate the effectiveness of the proposed spatial-spectral KSRC on three hyperspectral scenes. For each image, KSRC and SVM trained with four kernels (RBF, CK, MF and NF) are compared visually and quantitatively to each other, and the results are denoted by KSRC/SVM, KSRC-CK/SVM-CK, KSRC-MF/SVM-MF, and KSRC-NF/SVM-NF respectively. The proposed method KSRC-NF and the other three sparse representation based classifiers are solved by Algorithm 1, where the stopping criterion is defined as the difference between the recent two results $\epsilon = \|\mathbf{s}^{t+1} - \mathbf{s}^t\|_2/\|\mathbf{s}^t\|_2$. The optimization problem of SVM in the remaining four methods is solved by using LIBSVM software package [39]. In all methods, the free parameters are obtained by cross-validation. For the four sparse representation based classifiers, we empirically set $\mu = 10^{-3}$, $\lambda = 10^{-4}$ and $\epsilon = 0.001$. The regularization parameter of SVM is varied in the range $C \in \{10^{-1}, \ldots, 10^7\}$. A polynomial kernel is used for the spectral features of CK, and the kernel parameter is varied in the range $d = \{1, \ldots, 10\}$. RBF kernel is used for the others, and the kernel parameter is varied in the range $\gamma \in \{e^{-3}/L, \ldots, e^{12}/L\}$. For the proposed NF kernel, the degree of filtering γ_0 is chosen between $[e^{-3}/L, e^{12}/L]$. The weight ν of CK is tuned between [0.1, 0.9]. In addition, we build the spatial kernels of CK by using the mean and standard deviation of the neighborhood pixels in a window per spectral channel. Other details of CK are explained in [22].

Before the following experiments, each original data is scaled between [0, 1]. The classification accuracy is assessed with the overall accuracy (OA) which is the number of accurately classified samples divided by the number of test samples, the kappa coefficient of agreement (κ) which is the percentage of agreement corrected by the amount of agreement that could be expected due to chance alone, and the class-specific accuracy [40]. To investigate the challenging classification problem with high dimensionality of images and low number of training samples, we randomly choose different numbers of the labeled samples for training ($n \in \{5, 10, 20, 40, 60, 80, 100\}$ samples per class), which are smaller than the numbers of spectral bands in each image, and use the rest for testing. In addition, each experiment is repeated ten times to avoid any bias induced by random sampling, and the mean value and standard deviation are reported.

A. Indian Pines

The first hyperspectral image in our experiment is the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) image taken over northwest Indiana's Indian Pine test site in June 1992 [41], called Indian Pines. There are 220 bands in the image, covering the wavelength range of 0.4–2.5 μm. The spectral and spatial resolution are 10 nm and 17 m, respectively. This image consists of 145×145 pixels and 16 ground-truth classes ranging from 20 to 2468 pixels in size. The false color composite image is shown in Fig. 1(a). From the 16 different land-cover classes available in the original ground-truth, seven

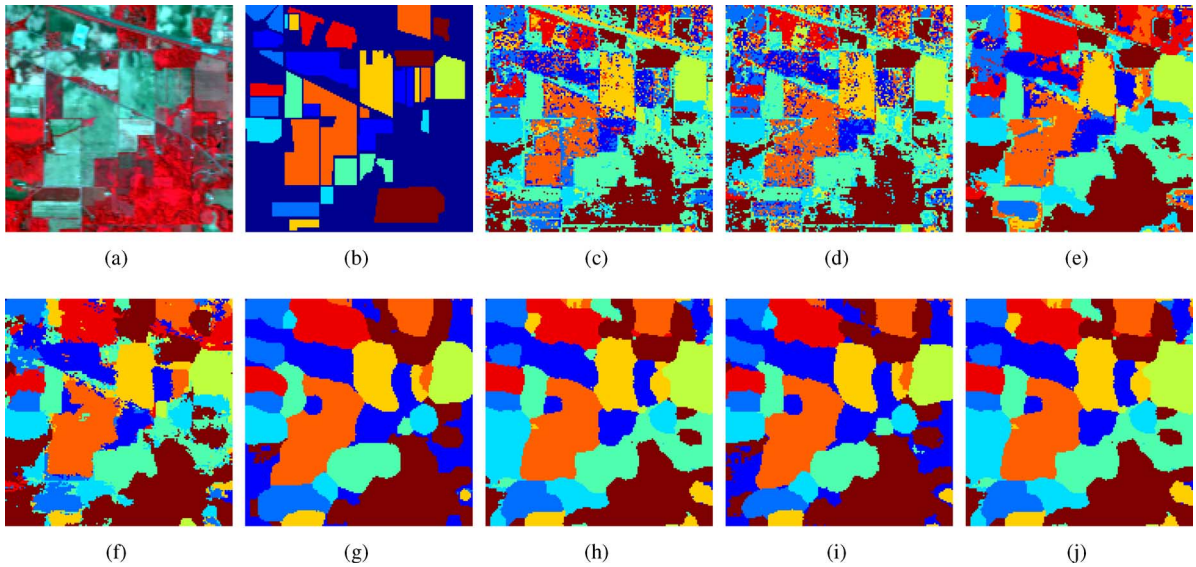


Fig. 1. For Indian Pines image. (a) RGB composite image of three bands. (b) Ground-truth map. (c) SVM. (d) KSRC. (e) SVM-CK. (f) KSRC-CK. (g) SVM-MF. (h) KSRC-MF. (i) SVM-NF. (j) KSRC-NF.

TABLE I
NINE GROUND TRUTH CLASSES IN AVIRIS INDIAN PINES AND THE LABELED NUMBER OF EACH CLASS

NO	Name	Samples
1	Corn-notill	1434
2	Corn-min	834
3	Grass/Pasture	497
4	Grass/Trees	747
5	Hay-windrowed	489
6	Soybeans-notill	968
7	Soybeans-min	2468
8	Soybean-clean	614
9	Woods	1294
Total		9345

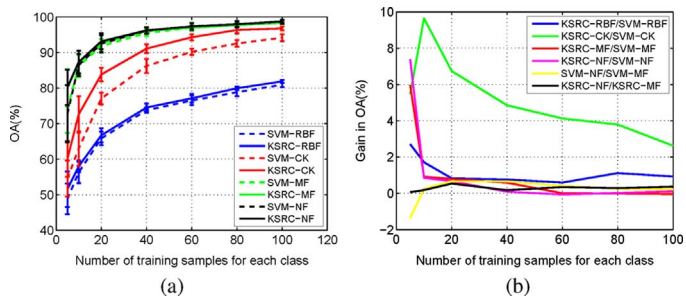


Fig. 2. The classification accuracy of Indian Pines data set. (a) OA. (b) Performance gain in OA.

were discarded since they have an insufficient number of training samples [42]. Nine ground-truth classes are considered in this experiment, as seen in Table I, and the spatial distribution of them is visually shown in Fig. 1(b). We remove 20 noisy bands covering the region of water absorption, and work with 200 spectral bands [20]. Since the Indian Pines image consists of large homogenous regions, a large window size $\omega = 11$ is used.

Results for different numbers of training samples are shown in Fig. 2. In this figure and the following, error bars indicate the standard deviation for ten random sampling, and suggest that not only numerical but also statistical differences can be observed. The performance gains are obtained by measuring the difference between the given pairs. For example, 'KSRC-RBF/SVM-RBF' means

$$\mathrm{OA} = \mathrm{OA}_{\text{KSRC-RBF}} - \mathrm{OA}_{\text{SVM-RBF}}.$$

From this figure, several conclusions can be extracted. First, there is a significant performance improvement after introducing the NF kernel to the standard SVM and KSRC. Second, with the same kernel, KSRC outperforms SVM. Third, the performance of the proposed NF kernel is better than the CK and MF kernels. When the number of training samples is low, the performance of these methods is poor, mainly due to more training samples are needed to capture the distribution of testing samples. When the number of training samples is higher, the effect of kernels is visible.

The classification maps on labeled pixels obtained from the various methods are shown in Fig. 1(c)–(j), and the classification accuracy of only 40 randomly selected training samples per class is reported in Table II. The numerical comparisons shown in Table II are confirmed by inspecting these classification maps. It is clear that the classification maps of the methods based on spatial-spectral kernels are much smoother than the classical kernels. Compared with the CK, the maps of NF and MF kernels look much more spatially homogeneous, and the misclassified pixels concentrate in the boundary. Among all these methods, KSRC-NF gets the best results. For SVM-CK, the misclassification areas concentrate in 'Corn-notill', and 'Soybeans-min'. For 'Grass/Trees' and 'Hay-windrowed', all these methods get smooth maps. It is especially hard for the classical kernels to capture the distribution of some mixtures, such as 'Corn-notill', 'Corn-min', 'Soybeans-notill', 'Soybeans-min', and 'Soybean-clean'.

To study the computational complexity of the proposed method, we show the CPU times of these spatial-spectral

TABLE II
THE CLASSIFICATION ACCURACY FOR AVIRIS INDIAN PINES USING DIFFERENT CLASSIFIERS

Class	SVM	KSRC	SVM-CK	KSRC-CK	SVM-MF	KSRC-MF	SVM-NF	KSRC-NF
1	61.05	63.62	78.34	89.31	93.40	96.98	93.62	97.34
2	73.90	71.15	88.92	89.16	97.48	85.64	98.99	87.27
3	92.12	88.84	97.59	99.56	97.81	99.78	99.56	99.78
4	98.16	94.20	98.16	99.57	98.87	99.00	100	99.85
5	98.22	99.10	100	100	100	99.33	100	99.55
6	70.47	67.45	85.56	93.31	92.03	94.07	92.56	94.39
7	55.72	60.33	74.51	82.00	91.47	96.25	93.08	97.03
8	77.35	77.00	94.08	94.59	99.48	96.51	100	96.51
9	93.22	96.25	97.61	98.88	99.84	99.04	100	99.92
OA(%)	73.60	74.66	86.29	91.27	95.37	96.16	96.26	96.80
κ	0.695	0.707	0.840	89.80	0.945	0.954	0.956	0.962

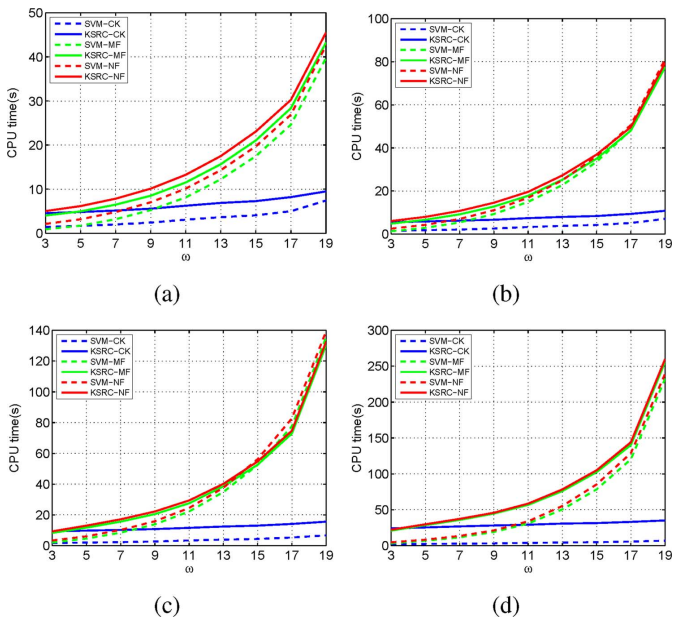


Fig. 3. CPU time (in seconds). (a) 10 training samples for each class. (b) 20 training samples for each class. (c) 40 training samples for each class. (d) 80 training samples for each class.

methods for different numbers of training samples and window sizes in Fig. 3. Every test is carried out in a 64-b quad-core Intel Xeon CPU 3.33-GHz processor under Windows 7. It can be seen that CK is the most efficient one of the three kernels. MF and NF kernels are sensitive to the numbers of training samples and window sizes, and MF kernel is a little faster than NF kernel. We may note that training an SVM with the same kernel is more efficient than KSRC. That is to say, KSRC is sensitive to the scale of training samples. However, with the complex MF and NF kernels, this gap is relatively small.

B. University of Pavia

The second hyperspectral image used in our experiments, University of Pavia, is an urban image acquired by the Reflective Optics System Imaging Spectrometer (ROSIS), with spectral coverage ranging from 0.43 to 0.86 μm [20]. The ROSIS sensor has spatial resolution of 1.3 m per pixel with 115 spectral bands. The University of Pavia image, with size of 610×340 pixels, contains 102 spectral bands after removal of noisy bands (see Fig. 5(a) for a color composite). There are nine ground-truth classes of interests, as seen in Table III, and Fig. 5(b) shows the

TABLE III
NINE CLASSES IN UNIVERSITY OF PAVIA AND THE LABELED NUMBER OF EACH CLASS

Class	Name	Samples
1	Asphalt	6631
2	Meadow	18649
3	Gravel	2099
4	Trees	3064
5	Metal sheets	1345
6	Bare soil	5029
7	Bitumen	1330
8	Bricks	3682
9	Shadows	947
Total		42776

spatial distribution of them. The window size ω for each spatial-spectral kernel is set to 11.

Fig. 4 illustrates the results for different numbers of training samples. It is clear that the spatial-spectral kernels outperform the traditional kernels in most cases. Though the same kernels are used, the results obtained from KSRC based methods are better than that of SVM based methods in most situations, especially when the number of training samples is low.

Fig. 5(c)–(j) show the classification maps obtained from the various methods, and all of them correspond to the tests using 40 labeled samples per class. Numerical results reported in Table IV are confirmed by the visual inspection of the classification maps. It can be seen that the spatial-spectral kernels get much smoother maps than the RBF kernel. Among all these methods, the classification map of KSRC-NF is much closer to the ground truth. All these methods get accurate maps of certain areas, such as 'Trees', 'Metal sheets', and 'Shadows'.

C. Kennedy Space Center

The third hyperspectral image is collected by AVIRIS over the Kennedy Space Center (KSC), Florida, USA, on March, 1996 [43]. This image contains 224 bands whose wavelength covers the range of 0.4–2.5 μm, and the spectral and spatial resolution are 10 nm and 18m, respectively. This image is acquired over an area of 512×614 pixels, and the false color composite image is shown in Fig. 6(a). After removing water absorption and low signal-to-noise bands, a total of 176 bands remained for classification. Discrimination of land cover for this environment is difficult due to the similarity of spectral signatures for certain vegetation types [43]. As seen in Table V, 13 ground-truth classes ranging from 105 to 927 pixels in size that occur in this environment is defined for this image. The ground truth map is

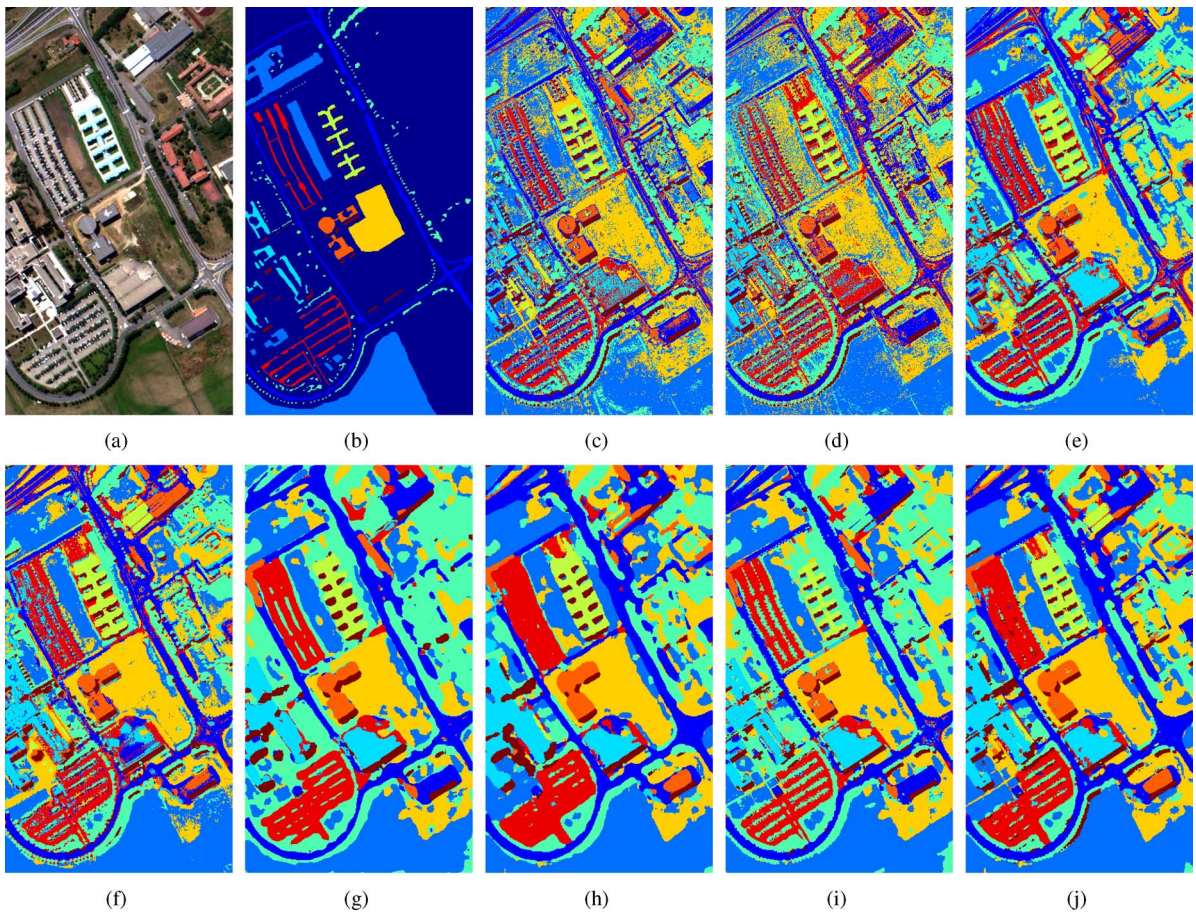


Fig. 4. For University of Pavia image. (a) RGB composite image of three bands. (b) Ground-truth map. (c) SVM. (d) KSRC. (e) SVM-CK. (f) KSRC-CK. (g) SVM-MF. (h) KSRC-MF. (i) SVM-NF. (j) KSRC-NF.

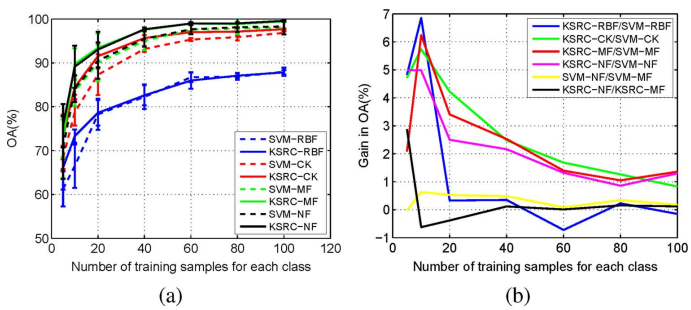


Fig. 5. The classification accuracy of University of Pavia data set. (a) OA. (b) Performance gain in OA.

shown in Fig. 6(b). For the spatial-spectral kernels, the window size ω is set to 11.

Since some classes have insufficient numbers of training samples, we only present the results of which the number of training samples of each class is less than or equal to 40. Fig. 7 reports the numerical results. As in the pervious experiments, some conclusions can also be extracted. First, the NF and MF kernels outperform CK. Second, for the spatial-spectral kernels, the performance of KSRC is better than SVM. Third, after introducing the spatial-spectral kernels to the classical KSRC and SVM, the performance is improved significantly.

The classification maps of 10 labeled samples per class are shown in Fig. 6(c)–(j), and the corresponding numerical results are shown in Table VI. These maps confirm the interest of introducing spatial-spectral kernels to hyperspectral image classification and confirm the validity of the proposed method. Among all these methods, KSRC-NF yields the highest accuracy. For 'Water', all these methods get accurate maps. The classical kernels are hard to capture the distribution of some objects, such as 'Scrub', 'Cabbage palm/oak hammock', 'Slash pine' and 'Oak/broadleaf hammock'.

D. Impact of Window Sizes

Since the MF and NF kernels are sensitive to the window sizes, the window size ω is an important parameter to be selected. Now we show the classification accuracy of the two kernels for different window sizes in Fig. 8. Forty labeled pixels per class are chosen as training samples for Indian Pines and University of Pavia, and ten for Kennedy Space Center. The remaining is used for testing. It can be seen that the classification performance increases almost monotonically as ω increases, but the results typically saturate for $\omega \approx 11$. For larger values of ω, a decrease in performance is expected as neighboring relations would be oversmoothed. In addition, KSRC-NF outperforms the other methods, and it exhibits a wider optimal region

TABLE IV
THE CLASSIFICATION ACCURACY FOR UNIVERSITY OF PAVIA USING DIFFERENT CLASSIFIERS

Class	SVM	KSRC	SVM-CK	KSRC-CK	SVM-MF	KSRC-MF	SVM-NF	KSRC-NF
1	72.83	74.63	77.06	86.89	91.12	94.44	90.11	96.35
2	78.97	81.30	96.97	97.69	94.37	99.42	94.51	99.45
3	83.05	82.80	87.23	95.09	95.53	98.98	97.09	99.70
4	94.54	95.86	98.68	98.31	95.54	91.36	99.01	94.97
5	99.23	99.54	100	99.84	99.46	99.61	99.62	100
6	88.74	86.24	95.63	98.73	99.98	99.85	99.94	99.93
7	93.26	84.10	95.81	99.68	100	99.76	100	100
8	81.71	77.70	91.21	91.62	98.22	97.58	99.29	94.56
9	100	100	100	100	98.68	98.67	99.89	100
OA(%)	82.21	82.52	92.99	95.71	95.41	97.94	95.75	98.34
κ	0.772	0.775	0.907	0.943	0.939	0.972	0.944	0.978

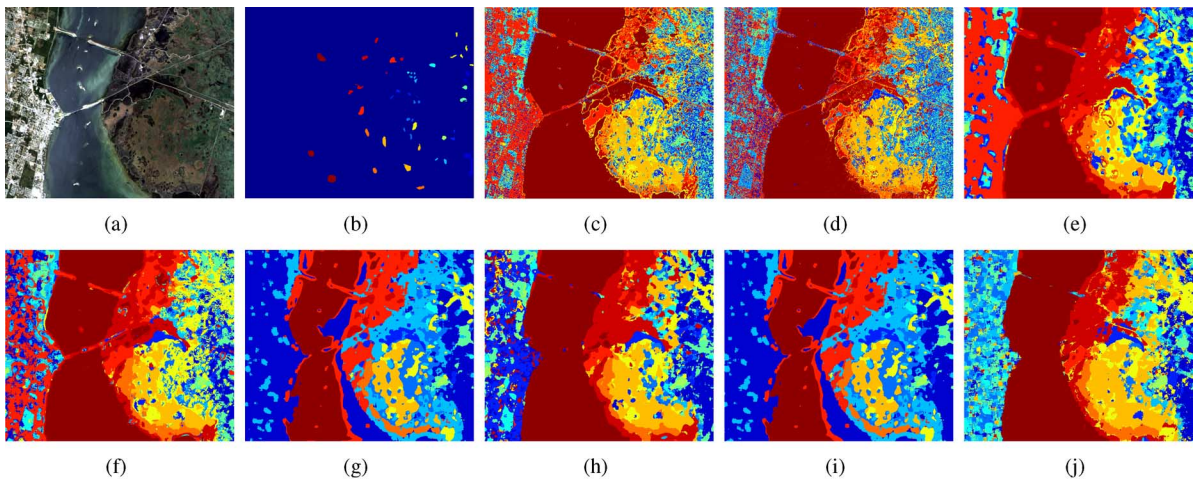


Fig. 6. For Kennedy Space Center image. (a) RGB composite image of three bands. (b) Ground-truth map. (c) SVM. (d) KSRC. (e) SVM-CK. (f) KSRC-CK. (g) SVM-MF. (h) KSRC-MF. (i) SVM-NF. (j) KSRC-NF.

TABLE V
THIRTEEN CLASSES IN KENNEDY SPACE CENTER AND THE LABELED NUMBER OF EACH CLASS

Class	Name	Samples
1	Scrub	761
2	Willow swamp	243
3	Cabbage palm hammock	256
4	Cabbage palm/oak hammock	252
5	Slash pine	161
6	Oak/broadleaf hammock	229
7	Hardwood swamp	105
8	Graminoid marsh	431
9	Spartina marsh	520
10	Cattail marsh	404
11	Salt marsh	419
12	Mud flats	503
13	Water	927
Total		5211

so it is less sensitive than the other methods to the choice of a suitable window size.

V. CONCLUSION

In this paper, we present a spatial-spectral kernel sparse representation for hyperspectral image classification. And the classical KSRC is extended to the spatial-spectral KSRC by incorporating information of the spatial adjacent pixels. To capture the spatial similarity between neighboring samples in the feature space, we propose the NF kernel, which can be interpreted as the neighborhood filtering in the feature space. Experimental results on three different hyperspectral scenes suggest good robustness and accuracy of the proposed spatial-spectral KSRC, and the validity of the NF kernel.

It is worth noting that the methodology opens a wide field for introducing filtering methods to spatial-spectral KSRC. In addition, the existing kernels used in SVM are all suitable for KSRC. Another issue to be addressed in the future is to devise automatic ways of selecting the window size, relating it to the spatial resolution of the image, and the degree of filtering, related to the spatial relations of classes.

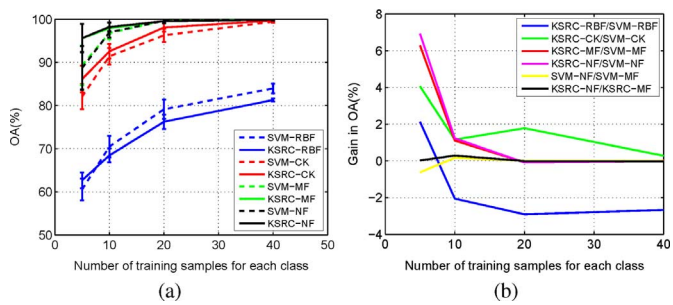


Fig. 7. The classification accuracy of Kennedy Space Center data set. (a) OA. (b) Performance gain in OA.

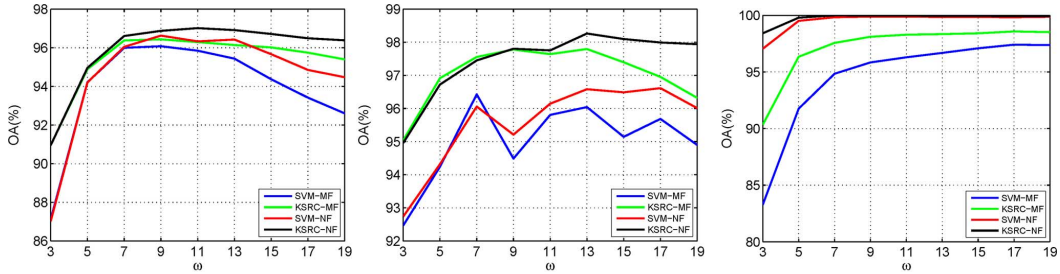


Fig. 8. The classification accuracy for different window sizes. (a) Indian Pines. (b) University of pavia. (c) Kennedy Space Center.

TABLE VI
THE CLASSIFICATION ACCURACY FOR KENNEDY SPACE CENTER USING DIFFERENT CLASSIFIERS

Class	SVM	KSRC	SVM-CK	KSRC-CK	SVM-MF	KSRC-MF	SVM-NF	KSRC-NF
1	36.35	30.09	81.23	70.70	100	94.94	100	100
2	89.70	83.26	87.55	98.28	100	100	100	100
3	56.10	32.92	87.40	97.96	100	90.24	100	100
4	20.25	37.60	90.91	75.20	96.69	89.25	97.11	91.73
5	39.74	19.86	83.44	88.07	100	100	100	80.79
6	41.55	37.89	92.69	94.97	100	99.08	100	100
7	73.68	42.10	100	100	100	100	100	100
8	68.17	67.45	88.84	97.62	99.76	94.06	100	100
9	90.98	84.90	83.14	96.86	94.71	98.23	84.31	100
10	67.77	82.74	100	99.74	100	100	100	100
11	97.80	78.48	94.38	100	100	96.33	100	99.02
12	71.40	82.15	97.57	96.95	89.05	99.59	100	100
13	99.89	100	100	100	100	100	97.82	100
OA(%)	70.37	67.52	91.49	92.91	97.22	97.22	97.89	98.95
κ	0.672	0.641	0.905	0.921	0.969	0.969	0.976	0.988

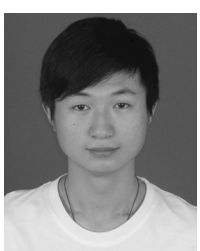
ACKNOWLEDGMENT

The authors would like to thank Prof. P. Gamba from the University of Pavia, Italy, for kindly providing the Pavia data set.

REFERENCES

[1] M. Borengasser, W. Hungate, and R. Watkins, *Hyperspectral Remote Sensing: Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2007.
[2] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, 2002.
[3] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, 2002.
[4] N. Patel, C. Patnaik, S. Dutta, A. Shekh, and A. Dave, "Study of crop growth parameters using airborne imaging spectrometer data," *Int. J. Remote Sens.*, vol. 22, no. 12, pp. 2401–2411, 2001.
[5] B. Datt, T. McVicar, T. Van Niel, D. Jupp, and J. Pearlman, "Preprocessing eo-1 hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1246–1259, 2003.
[6] B. Hörig, F. Kühn, F. Oschütz, and F. Lehmann, "Hymap hyperspectral remote sensing to detect hydrocarbons," *Int. J. Remote Sens.*, vol. 22, no. 8, pp. 1413–1422, 2001.
[7] F. Melgani and S. Serpico, "A statistical approach to the fusion of spectral and spatio-temporal contextual information for the classification of remote-sensing images," *Pattern Recognit. Lett.*, vol. 23, no. 9, pp. 1053–1061, 2002.
[8] A. Bardossy and L. Samaniego, "Fuzzy rule-based classification of remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 2, pp. 362–374, 2002.
[9] L. Bruzzone and R. Cossu, "A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 1984–1996, 2002.
[10] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004.
[11] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, 2005.
[12] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.
[13] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.
[14] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.
[15] M. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, 2011.
[16] Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, 2011.
[17] L. Zhang, W. D. Zhou, P. C. Chang, J. Liu, Z. Yan, T. Wang, and F. Z. Li, "Kernel sparse representation-based classifier," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1684–1695, 2012.
[18] S. Gao, I. Tsang, and L. Chia, "Kernel sparse representation for image classification and face recognition," in *Proc. ECCV*, Crete, Greece, Sep. 2010, pp. 1–14, Springer.
[19] A. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.
[20] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, 2009.
[21] Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification via kernel sparse representation," in *Proc. ICIP*, Sep. 2011, pp. 1233–1236.
[22] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Mari ́, J. Vila-France ́s, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image-classification," *IEEE Geosci. Remote Sens.*, vol. 3, no. 1, pp. 93–97, 2006.
[23] M. Marconcini, G. Camps-Valls, and L. Bruzzone, "A composite semisupervised svm for classification of hyperspectral images," *IEEE Geosci. Remote Sens.*, vol. 6, no. 2, pp. 234–238, 2009.
[24] D. Tuia and G. Camps-Valls, "Semisupervised remote sensing image classification with cluster kernels," *IEEE Geosci. Remote Sens.*, vol. 6, no. 2, pp. 224–228, 2009.

[25] D. Tuia and G. Camps-Valls, "Urban image classification with semisupervised multiscale cluster kernels," *IEEE J. Sel. Topics Quantum Electron.*, vol. 4, no. 1, pp. 65–74, 2011.
[26] L. Gómez-Chova, G. Camps-Valls, L. Bruzzone, and J. Calpe-Maravilla, "Mean map kernel methods for semisupervised cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 207–220, 2010.
[27] G. Camps-Valls, N. Shervashidze, and K. Borgwardt, "Spatio-spectral remote sensing image classification with graph kernels," *IEEE Geosci. Remote Sens.*, vol. 7, no. 4, pp. 741–745, 2010.
[28] R. Flamary, D. Tuia, B. Labbé, G. Camps-Valls, and A. Rakotomamonjy, "Large margin filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 648–659, 2012.
[29] A. Aizerman, E. Braverman, and L. Rozoner, "Theoretical foundations of the potential function method in pattern recognition learning," *Autom. Remote Contr.*, vol. 25, pp. 821–837, 1964.
[30] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. Rojo-Alvarez, and M. Martínez-Ramón, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, 2008.
[31] D. Tuia, F. Ratle, A. Pozdnoukhov, and G. Camps-Valls, "Multisource composite kernels for urban-image classification," *IEEE Geosci. Remote Sens.*, vol. 7, no. 1, pp. 88–92, 2010.
[32] A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising," in *Proc. CVPR*, San Diego, CA, USA, Jun. 2005, vol. 2, pp. 60–65, IEEE.
[33] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2000.
[34] S. Z. Li, "Face recognition based on nearest linear combinations," in *Proc CVPR*, Santa Barbara, CA, USA, Jun. 1998, pp. 839–844.
[35] S. Umbaugh, *Computer Imaging: Digital Image Analysis and Processing*. Boca Raton, FL, USA: CRC Press, 2005.
[36] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. WHISPERS*, Reykjavik, Iceland, Jun. 2010, pp. 1–4, IEEE.
[37] J. Yang and Y. Zhang, "Alternating direction algorithms for l_1-problems in compressive sensing," Arxiv Preprint arXiv:0912.1185 2009.
[38] P. Combettes *et al.*, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling Simul.*, vol. 4, no. 4, pp. 1168–1200, 2006.
[39] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Tran. Intell. Syst. Technol.* vol. 2, pp. 27:1–27:27, 2011 [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm
[40] M. Fauvel, J. Chanussot, and J. Benediktsson, "A spatial-spectral kernelbased approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, 2012.
[41] N. AVIRIS, "Indianas indian pines 1992 data set," [Online]. Available: [Online]. Available: http://cobweb.ecn.purdue.edu/biehl/MultiSpec/documentation.html
[42] G. Camps-Valls, T. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, 2007.
[43] J. Ham, Y. Chen, M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, 2005.



JianJun Liu was born in Jiangsu, China, in 1987. He received the B.Sc. degree in applied mathematics from Nanjing University of Science and Technology (NUST), Nanjing, Jiangsu, China, in 2009, where he is currently working toward the Ph.D. degree in computer science.

His research interests are in the area of spectral unmixing, hyperspectral image classification, image processing, sparse representation, and compressive sensing.

ZeBin Wu was born in Zhejiang, China, in 1981. He received the B.Sc. and Ph.D. degree in computer science from Nanjing University of Science and Technology (NUST), Nanjing, Jiangsu, China, in 2003 and 2008, respectively.

He is currently an Associate Professor at the School of Computer Science of NUST. His research interests include virtual reality and system simulation, remote sensing data analysis, and distributed computing.

ZhiHui Wei was born in Jiangsu, China, in 1963. He received the B.Sc., M.Sc., and Ph.D. degree from South East University, Nanjing, Jiangsu, China, in 1983, 1986, and 2003, respectively.

He is currently a Professor and Doctoral Supervisor at Nanjing University of Science and Technology (NUST), Nanjing, Jiangsu, China. His main research areas include partial differential equations, mathematical image processing, image modeling, multiscale analysis, video and image coding, sparse representation and compressive sensing. His current research is focused on the theory of image sampling, multiscale geometrical analysis, sparse representation, and compressive sensing.

Liang Xiao was born in Hunan, China, in 1976. He received the B.Sc. degree in applied mathematics and his Ph.D. degree in computer science from Nanjing University of Science and Technology (NUST), Nanjing, Jiangsu, China, in 1999 and 2004, respectively.

From 2006 to 2008, he was a post-doctoral Research Fellow at the Pattern Recognition Laboratory of the NUST. From 2009 to 2010, he was a post-doctoral Fellow at Rensselaer Polytechnic Institute (RPI), USA. He is currently an Associate Professor at the School of Computer Science of NUST. His main research areas include inverse problems in image processing, scientific computing, data mining, and pattern recognition.

Le Sun was born in Jiangsu, China, in 1987. He received the B.Sc. degree in applied mathematics from NUST, Nanjing, Jiangsu, China, in 2009, where he is currently working toward the Ph.D. degree in computer science.

His research interests include spectral unmixing, hyperspectral image classification, sparse representation, and compressive sensing.