Pattern Recognition, Volume 77, May 2018, pp 316-328
DOI:10.1016/j.patcog.2017.10.008

Joint Bilateral Filtering and Spectral Similarity-Based Sparse Representation: A Generic Framework for Effective Feature Extraction and Data Classification in Hyperspectral Imaging

Tong Qiao[1,2], Zhijing Yang[1], Jinchang Ren[2], Peter Yuen[3], Huimin Zhao[4,5], Genyun Sun[6], Stephen Marshall[2] and Jon Atli Benediktsson[7]

[1] School of Electronic and Information Engineering, Guangdong University of Technology, Guangzhou, China
[2] Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1XW, UK.
[3] Electro-Optics, Image and Signal Processing, Cranfield University, Swindon, SN6 8LA, U.K.
[4] School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, 510665, China.
[5] The Guangzhou Key Laboratory of Digital Content Processing and Security Technologies, Guangzhou, China.
[6] School of Automations, Northwestern Polytechnical University, Xi'an, 710072, China.
[7] Faculty of Electrical and Computer Engineering, University of Iceland, 101, Reykjavik, Iceland.

Abstract

Classification of hyperspectral images (HSI) has been a challenging problem under active investigation for years especially due to the extremely high data dimensionality and limited number of samples available for training. It is found that hyperspectral image classification can be generally improved only if the feature extraction technique and the classifier are both addressed. In this paper, a novel classification framework for hyperspectral images based on the joint bilateral filter and sparse representation classification (SRC) is proposed. By employing the first principal component as the guidance image for the joint bilateral filter, spatial features can be extracted with minimum edge blurring thus improving the quality of the band-to-band images. For this reason, the performance of the joint bilateral filter has shown better than that of the conventional bilateral filter in this work. In addition, the spectral similarity-based joint SRC (SS-JSRC) is proposed to overcome the weakness of the traditional JSRC method. By combining the joint bilateral filtering and SS-JSRC together, the superiority of the proposed classification framework is demonstrated with respect to several state-of-the-art spectral-spatial classification approaches commonly employed in the HSI community, with better classification accuracy and Kappa coefficient achieved.

Index Terms

Hyperspectral imaging; joint bilateral filtering; sparse representation; feature extraction; data classification.

Corresponding authors: Dr Z. Yang (yzhj@gdut.edu.cn).

Published by Elsevier. This is the Author Accepted Manuscript issued with: Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0).
The final published version (version of record) is available online at DOI:10.1016/j.patcog.2017.10.008. Please refer to any applicable publisher terms of use.

I. INTRODUCTION

By utilizing hundreds of continuous spectral bands in a broad range, hyperspectral imaging (HSI) has been an important surveillance and reconnaissance technology for military [1] as well as other civil applications such as food quality assessment [2, 3], pharmaceutical analysis [4], security control [5], etc. Among all HSI-based applications, land cover analysis has been an active topic in remote sensing field. To achieve high classification accuracy in the high dimensional data such as hyperspectral images, sufficient training samples are needed to cope with the curse of dimensionality, i.e., Hughes phenomenon [6]. However, it is not feasible at times having enough of samples for training in real world. Besides, due to the sensor noise and material variability, the same class of material may exhibit somewhat different spectral characteristic and even so different categories of samples may have very similar spectral signatures [7]. These factors inevitably influence the classification accuracy significantly.

Efforts such as the extraction of more effective features in both spatial and spectral domains, and to construct suitable classification models for the raw/extracted feature, are suggested to overcome these classification issues and to improve accuracy. Previous work in spectral feature extraction techniques have been principal component analysis (PCA) [8–10], independent component analysis (ICA) [11], minimum noise fraction (MNF) [12], non-negative matrix factorisation (NMF) [13] and one-dimensional singular spectrum analysis (1D-SSA) [14, 15], where effective spectral features are achieved in the transform domain.

In addition to spectral feature extraction, spatial feature extraction is found more attractive through its effective exploitation of contextual information. For instance, two-dimensional empirical mode decomposition (2D-EMD) [16, 17] and two-dimensional singular spectrum analysis (2D-SSA) [18], which can decompose each band into a few components containing the spatial structural content, have been proven to be superior in extracting spatial features. However, the computational complexity of 2D-EMD on a single band image is high, let alone hundreds of bands in a hyperspectral image [7]. Besides, it is noticed that the extracted spatial information using 2D-SSA could over-smooth the data with blurred feature artefacts, leading to classification error especially around the edges of objects. Another popular approach for spatial feature extraction is based on the morphological profile (MP) constructed by opening/closing operators using a series of structure elements (SEs) with different sizes and it has been developed as valuable technique for extracting spatial geometrical details. As shown in [19] and [20], MP and morphological attribute profile (MAP) were applied to the first a few principle components achieved from a hyperspectral image, generating the extended MP (EMP) and the extended MAP (EMAP).

For data classification in hyperspectral imaging, various machine-learning techniques have been proposed, including maximum likelihood [21, 22] or Bayesian estimation methods [23, 24], decision trees [25, 26], artificial neural networks [27, 28], etc. Among all approaches, the support vector machine (SVM) has demonstrated to be more superior in supervised hyperspectral image

classification [3, 6, 10] as it can handle large input spaces and less sensitive to the Hughes phenomenon compared to that of artificial neural networks [29]. In recent years, sparse representation has received considerable attention in fields of signal/image processing, computer vision and pattern recognition. Sparse representation is an indispensable prerequisite of the compressed sensing (CS) framework [30– 32], which demonstrated that most natural signals can be precisely reconstructed by only a few measured values. Thus, sparse representation can be an outstanding tool for classification of high dimensional data such as [33] for face recognition where the test image can be sufficiently represented by samples extracted from the same category.

In [34], sparse representation classification (SRC) was applied for hyperspectral image classification and a joint sparse representation classification (JSRC) method was proposed by assuming all pixels around the centre test pixel in the neighbourhood window shared a common sparsity pattern. In this work the hyperspectral image classification problem was approached by a joint sparsity model (JSM) [35, 36] that utilizes spatial information and the classification accuracy was found greatly enhanced compared with that of the pixel-wise spectral SRC method. In the original JSRC method, all pixels around the centre test pixel in the neighbourhood window had been assigned equal when the JSM was constructed. This would induce error particularly when the region of interest consisted of multiple classes. To address this problem, Zhang et al. [37] proposed a nonlocal weighted joint sparse representation classification (NLW-JSRC) method to determine the weighted contribution of neighbouring pixels. In the NLW-JSRC method, the weight of each neighbouring pixel was calculated by comparing the structural similarity between an image patch around that pixel and the patch that surround the centre test pixel. In this way, the classification accuracy was found improved by considering both the spectral information between these two pixels and the geometrical details between these two nonlocal image patches.

Another approach for improving hyperspectral data classification has been post-processing of preliminarily classified results. In both [38] and [39], the authors proposed to use edge-preserving filter for smoothing the pixel-wise classification map under the smoothness of class boundary assumption. Results showed that the filtering-based post- processing could improve classification accuracies of some hyperspectral scenes significantly.

The edge-preserving filtering approach such as the use of bilateral filter has lately become a popular topic in the area of image processing and recently hyperspectral image analysis [40]. Most work in the area had been using bilateral filtering for enhancing hyperspectral image quality [41], however, the application of the technique for feature extraction in hyperspectral image is seldom. In this paper, a spectral-spatial classification framework based on bilateral filter and sparse representation is proposed for improving classification accuracy. It is demonstrated in this work that the proposed spatial/spectral feature extraction enhances classification over that using the contextual features extracted by the conventional JSRC method. The proposed method consists of spatial feature which is firstly extracted by a joint bilateral filter using the first principal component as the guidance image. It is observed that the quality of the extracted features have been effectively enhanced

particularly for those spatially low contrast bands. Although the NLW-JSRC could have achieved better classification than that of the JSRC, the choice of appropriate Gaussian kernels and the huge computational cost in NLW-JSRC would be a complex issue to consider. Hence, a spectral similarity-based JSRC (SS-JSRC) method that utilizes the statistics of neighborhood window to exclude irrelevant pixels has been proposed as an alternative method for enhanced classification in this paper.

The layout of this paper is as follows. The background principles of the bilateral filter and SRC are discussed in Section II. Section III explains the proposed feature extraction and classification framework, followed by experimental data sets and setup in Section IV. Classification results on two hyperspectral images are given in Section V, and conclusion is drawn in Section VI.

II. Background Principles

Bilateral filtering and SRC are the two main concepts adopted in the proposed classification framework. The background principles of these two techniques are briefly introduced below.

A. Bilateral Filtering

The bilateral filtering proposed by Tomasi and Manduchi [42] in 1998 is a non-linear filtering technique that combines domain and range filtering together to smooth images while preserving edges. This filtering technique replaces the pixel value at one location (test pixel) by the weighted sum of nearby pixel intensities based on their geometric closeness and photometric similarity with respected to the test pixel. Given an input image $f(x)$ and by adopting the same notation as that used in [42], the bilateral-filtered image $h(x)$ is achieved mathematically as follows:

$$h(x) = \frac{1}{\kappa(x)} \int_{-\infty}^{\infty} f(\xi) c(\xi, x) s(f(\xi), f(x)) d\xi \tag{1}$$

where $c(\xi, x)$ is a function that measures the geometric closeness (in the domain) between the neighborhood center x and a neighboring pixel ξ. The function $s(f(\xi), f(x))$ calculates the photometric similarity (in the range) between x and ξ, where the normalization factor is given by

$$\kappa(x) = \int_{-\infty}^{\infty} c(\xi, x) s(f(\xi), f(x)) d\xi \tag{2}$$

In this work both the closeness function $c(\xi, x)$ and the similarity function $s(f(\xi), f(x))$ in a bilateral filter are represented by Gaussian functions [42] for simplicity. The discrete representation of bilateral filtering commonly used for image processing with an input image I at the location (i, j) can be expressed as [40]:

$$I_{BF}(i,j)=\frac{1}{\kappa(i,j)}\sum_{(p,q)\in w}\{G_{\sigma_d}(i-p,j-q)\times G_{\sigma_r}(I(i,j)-I(p,q))I(p,q)\} \tag{3}$$

with the normalization factor

$$\kappa(i,j)=\sum_{(p,q)\in w}\{G_{\sigma_d}(i-p,j-q)\times G_{\sigma_r}(I(i,j)-I(p,q))\} \tag{4}$$

where (p,q) represents the location of pixels in the $(2\sigma_d+1)\times(2\sigma_d+1)$ local window w around the pixel at (i,j).

The kernels for domain and range filtering are defined as follows:

$$G_{\sigma_d}(i-p,j-q)=\exp\left(-\frac{(i-p)^2+(j-q)^2}{2\sigma_d^2}\right) \tag{5}$$

$$G_{\sigma_r}(I(i,j)-I(p,q))=\exp\left(-\frac{(I(i,j)-I(p,q))^2}{2\sigma_r^2}\right) \tag{6}$$

where the value of σ_d determines the neighborhood window under consideration and the value of σ_r defines the weight of contribution from neighboring pixels [42]. Although optimal selection of these two parameters seems image-dependent, for particular applications it could be possible to be determined automatically.

B. Sparse Representation Classification

As pixels in an image or a hypercube can be sparsely represented by a linear combination of a few training samples from M distinct classes, sparse representation in combination with supervised classification has been successfully applied for applications such as face recognition and land cover mapping [33, 34]. Assuming there are N_m training samples of the m^{th} class in a B-dimensional hyperspectral data cube, denoted as $\mathbf{A}^m=[a_1^m \quad a_2^m \quad \ldots \quad a_1^m]\in\Re^{B\times N_m}$, any test sample $\mathbf{x}\in\Re^B$ from the same class could approximately be represented by a linear combination of the training samples associated with class m:

$$\mathbf{x}=\sum_{i=1}^{N_m}\alpha_i^m\mathbf{a}_i^m=\mathbf{A}^m\cdot\boldsymbol{\alpha}^m \tag{7}$$

where $\boldsymbol{\alpha}^m$ is an N_m-dimensional coefficient vector with only a few non-zero entries in the sparsity model. Since normally the label for a test sample is unknown, a new matrix $\mathbf{A}$ should be defined by concatenating all training samples from all classes:

$$\mathbf{A}=[\mathbf{A}^1 \quad \mathbf{A}^2 \quad \ldots \quad \mathbf{A}^M]\in\Re^{B\times N} \tag{8}$$

where $N=\sum_{m=1}^{M}N_m$. Then, the test sample $\mathbf{x}$ can be approximated by

$$\mathbf{x} = \sum_{i=1}^{M} \mathbf{A}^i \boldsymbol{\alpha}^i = \mathbf{A} \cdot \boldsymbol{\alpha} \tag{9}$$

where $\boldsymbol{\alpha}$ is an N -dimensional sparse coefficient vector with most of entries as zero.

The task of hyperspectral image SRC is to achieve the sparse vector $\boldsymbol{\alpha}$ and to assign a label to the test sample according to $\boldsymbol{\alpha}$. The sparse solution of $\boldsymbol{\alpha}$ can be obtained by solving the following optimisation problem:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}\|_0 \qquad s.t. \quad \mathbf{A}\boldsymbol{\alpha} = \mathbf{x} \tag{10}$$

where $\|.\|_0$ is the ℓ^0 -norm representing the number of non-zero entries in a vector. The optimisation problem in (10) can be also modified by minimising the reconstruction error with a pre-defined sparsity level K_0 [34]:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\mathbf{A} \cdot \boldsymbol{\alpha} - \mathbf{x}\|_2 \; s.t. \|\boldsymbol{\alpha}\|_0 \le K_0 \tag{11}$$

The problem in (11) is a NP-hard problem that could be solved by greedy pursuit algorithms, such as orthogonal matching pursuit (OMP) [43] or subspace pursuit (SP) [44]. In this paper, OMP was used to solve the problem where the support set is augmented by one index for each iteration until K_0 non-zero entries are selected, and the computational complexity is in the order of $O(BNK_0)$. With the obtained sparse vector $\boldsymbol{\alpha}$, the label of $\mathbf{x}$ could be assigned to the class with lowest residue:

$$\text{Class}(\mathbf{x}) = \arg\min_{m \in [1,M]} (r^m(\mathbf{x})) \tag{12}$$

where the residue $r^m \mid m \in [1, M]$ is calculated for each class separately as follows:

$$r^m(\mathbf{x}) = \left\| \mathbf{x} - \mathbf{A}^m \hat{\boldsymbol{\alpha}}^m \right\|_2 \tag{13}$$

For a hypercube, the spectral profiles of neighbouring pixels in a small window are usually highly correlated. By assuming that pixels within the neighbourhood window lie in a common subspace, Chen *et al.* [34] proposed the JSRC method for enhancing hyperspectral image analysis. Considering that the investigated test sample $\mathbf{x}_1$ is located at the centre of the neighbourhood window, the sample matrix for the window with T pixels can be modelled as $\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \ldots \quad \mathbf{x}_T] \in \Re^{B \times T}$. Applying the joint sparsity model, the sample matrix $\mathbf{X}$ can be represented by:

$$\begin{aligned} \mathbf{X} &= [\mathbf{A}\boldsymbol{\alpha}_1 \quad \mathbf{A}\boldsymbol{\alpha}_2 \quad \ldots \quad \mathbf{A}\boldsymbol{\alpha}_T] \\ &= \mathbf{A}[\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \quad \ldots \quad \boldsymbol{\alpha}_T] = \mathbf{AS} \end{aligned} \tag{14}$$

where $\mathbf{S} \in \Re^{B \times T}$ is the set of sparse coefficient vectors for all pixels within the neighbourhood window. The sparsity support for each pixel is the same, but the associated coefficients are different. Similarly, the task of JSRC is to obtain the sparse matrix $\mathbf{S}$ by solving the following optimisation problem:

$$\hat{\mathbf{S}} = \arg\min \|\mathbf{S}\|_{row,0} \quad \text{s.t.} \quad \mathbf{AS} = \mathbf{X} \tag{15}$$

The symbol $\|\cdot\|_{row,0}$ stands for the number of non-zero rows in the denoted matrix. In the same way as (11), the optimisation problem in (15) can be rewritten as:

$$\hat{\mathbf{S}} = \arg\min \|\mathbf{AS} - \mathbf{X}\|_F \quad s.t. \quad \|\mathbf{S}\|_{row,0} \le K_0 \tag{16}$$

where $\|\cdot\|_F$ is the Frobenius norm. In this paper, a greedy pursuit algorithm named as simultaneous OMP (SOMP) [35] is used to solve the NP-hard problem in (16). Given the solved sparse matrix $\mathbf{S}$, the investigated test pixel centred at the neighbourhood window can be assigned to the class with the smallest residue, as shown below:

$$\begin{aligned} \text{Class}(\mathbf{x}_1) &= \arg\min_{m\in[1,M]}(r^m(\mathbf{X})) \\ &= \arg\min_{m\in[1,M]} \left\|\mathbf{X} - \mathbf{A}^m\hat{\mathbf{S}}^m\right\|_F \end{aligned} \tag{17}$$

III. The Proposed Approach

The spectral-spatial classification framework proposed in this paper combines feature extraction and classification processes together. The effective feature extraction step is achieved by joint bilateral filtering with the first principal component as the guidance image, and the filtered image is classified using the SS-JSRC method. The details of the proposed classification framework will be described in the following section.

The basic bilateral filtering method as introduced in Section II evaluates the Gaussian kernel for range filtering based on the image intensities of a small patch. If the pixel at (k,l) in the image patch has a similar intensity with the 'to-be-smoothed' pixel at (i,j), the weight of pixel (k,l) will be large. This is how the edge information in the image is preserved using bilateral filtering.

Hyperspectral image often consists of spatially low contrast images in a number of spectral bands. If the range filtering kernel was calculated using these low contrast images, the edge information would be further degraded after bilateral filtering was applied. Therefore, in this paper, it is proposed to use a joint bilateral filter by using the first principal component as the guidance image since the first principal component after PCA accounts for as much of the variability in the hyperspectral image as possible. Thus, Eq. (3) can be modified as follows:

$$\begin{aligned} I_{JBF}(i,j,b) = \frac{1}{k(i,j)} \sum_{(p,q)\in w} \{ & G_{\sigma_d}(i-p, j-q) \\ & \times G_{\sigma_r}(I_{PC}(i,j) - I_{PC}(p,q)) I(p,q,b)\} \end{aligned} \tag{18}$$

where I_{PC} represents the first principal component (PC), b is the spectral band index and I_{JBF} is the output of joint bilateral filtering (JBF). The normalisation factor is also changed as follows:

$$k(i,j) = \sum_{(p,q)\in w} \{G_{\sigma_d}(i-p, j-q) \times G_{\sigma_r}(I_{PC}(i,j) - I_{PC}(p,q))\} \tag{19}$$

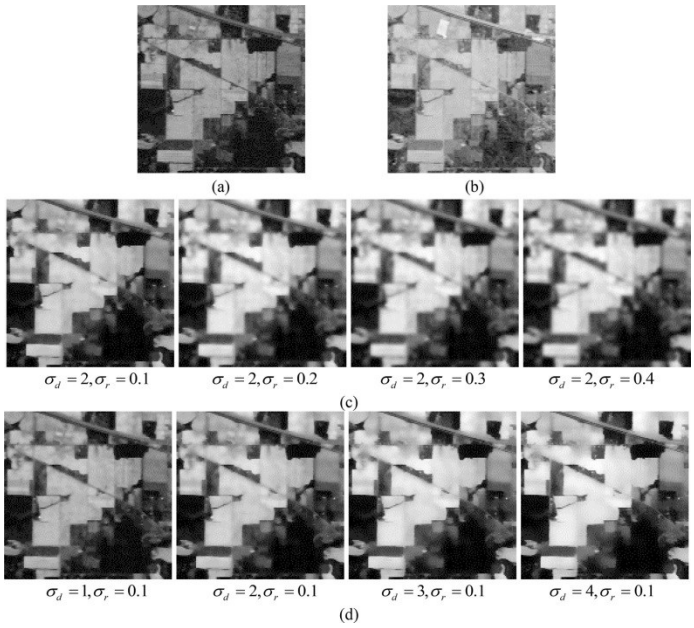

Fig. 1. The performance of joint bilateral filtering with different sets of parameters. (a) Input band image from the Indian Pines data set, (b) the first principal component, (c) filtered images with different range parameters, and (d) filtered images with different domain parameters.

The input band image and the first principal component of the Indian Pines data set [45] are shown in Fig.1 (a) and Fig.1 (b), respectively, while (c) and (d) show the JBF-filtered images with different sets of domain and range (σ_d, σ_r) parameters. Before applying the JBF, the radiances of the hyperspectral image is normalised to the range [0,1]. In Fig. 1(c), the kernel parameter of domain filter σ_d is set to 2 and the kernel parameter of range filter σ_r is changed from 0.1 to 0.4. It is seen that JBF increases Gaussian blurring with increasing σ_r. Hence, the range parameter σ_r is chosen to be small at 0.1 as the edge-preserving factor and the domain parameter σ_d is varied from 1 to 4 to compare the effects as shown in Fig. 1(d). By judging from the quality of the filtered images that presented in Fig. 1, it can be seen that a higher domain parameter leads to smoother features, therefore, the domain parameter σ_d has been chosen as 4 in the following experiments.

As mentioned above, JBF has been employed in this work instead of using the conventional bilateral filtering for improving image qualities particularly for those low-contrast image bands. Fig. 2(a) illustrates the image of a low contrast band of the Indian Pine with the first principal component presented in Fig. 2(b). The resulting images after the application of the normal bilateral filtering and JBF using the same parameter set ($\sigma_d = 4, \sigma_r = 0.1$) are shown in Fig. 2(c) and (d), respectively. It is noted that all the edge information has been kept and enhanced when JBF is applied.

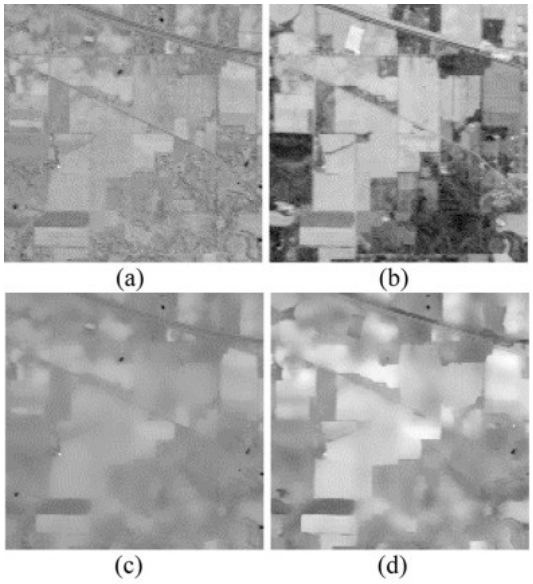

Fig. 2. Comparison of bilateral filtering and joint bilateral filtering on low-contrast image. (a) Input band image from the Indian Pines data set, (b) the first principal component, (c) output image using bilateral filtering, and (d) output image using joint bilateral filtering with the first principal component as the guidance image.

Using the first principal component for spatial filtering and feature extraction has been widely used by many others. These include extraction of morphological profiles [49], shape-adaptive local smoothing [52] and creating convolutional neural networks for classification [53]. To include more information for more effective feature extraction, it is feasible to fuse the first principal component with other principal components or even selected band images [49], which will be further investigated as our future work. In addition, it is worth noting that the smoothed image in Fig. 2(d) appears to be brighter than Fig. 2(c) is due to the higher intensity from the guidance image used in JBF, i.e. Fig. 2(b). As the proposed filter can successfully remove noise and potentially suppress the range of the band image, the detail from the resulting image can be enhanced even with increased brightness when the intensity is re-scaled to [0,255].

The hyperspectral data is ready for classification after spatial features are extracted using joint bilateral filtering. As mentioned in Section II, all pixels in the neighborhood window within the JSRC method have been conventionally considered to be equally the same. However, in very inhomogeneous scene the window may include class boundaries with pixels belong to different categories and a different sparsity pattern should be assigned to these different classes of pixels. Hence we have adopted the SS-JSRC methodology to screen out different classes of pixels within the neighborhood window and the Flowchart of the algorithm is schematically illustrated in Fig. 3.

The Indian Pines data set presented in Fig. 3 illustrates an example and it depicts the test sample $\mathbf{x}_1$ at the center of a small neighborhood window which consists of pixel samples in matrix $\mathbf{X}$ as in Eq.(14). In this example there are two classes of pixels labelled in blue and red colors respectively which are both contained in the window. The example illustrates with the center test

pixel belongs to class *red*, and it is necessary to exclude pixels from the class *blue* for the evaluation of the sparsity pattern of $\mathbf{x}_1$. This can be achieved by performing a Euclidean distance (ED) measures for every pixels in the window with respected to the center test pixel. The standard deviation (SD) of this matrix can then be evaluated and a threshold based on the standard deviation can be set: pixels with distances lower than the set threshold are considered to be spectrally similar to the center test pixel below:

$$W = \begin{cases} 0 & ED > n * SD \\ 1 & ED \le n * SD \end{cases} \tag{20}$$

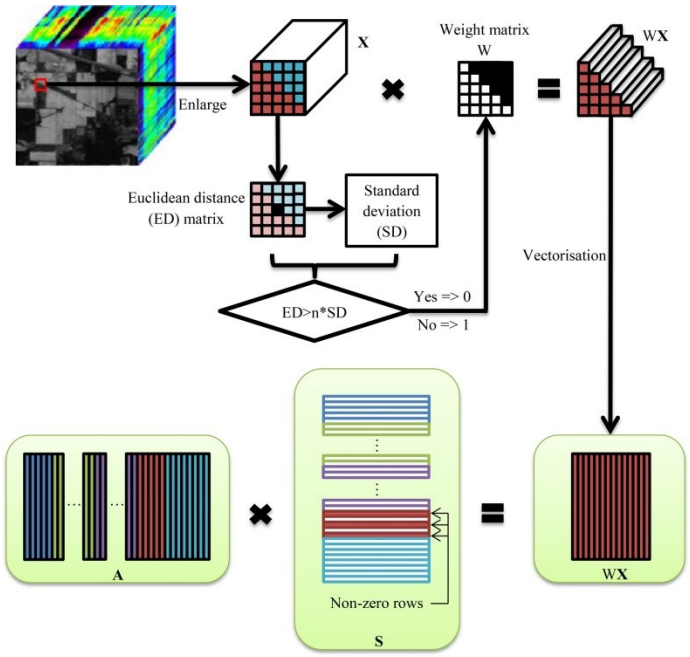

Fig. 3. Illustrates the flowchart of the proposed SS-JSRC method schematically.

The threshold parameter *n* is set by experiment in the present work, which is normally set within [1.0, 3.0]. Then the joint sparsity model can be applied with the reduced sample matrix $W\mathbf{X}$ and the classified label for the center test pixel can be achieved as follows:

$$\hat{\mathbf{S}} = \arg\min \left\| \mathbf{AS} - W\mathbf{X} \right\|_F \quad \text{s.t.} \quad \left\| \mathbf{S} \right\|_{row,0} \le K_0 \tag{21}$$

$$\begin{aligned} \text{Class}(\mathbf{x}_1) &= \arg\min_{m \in [1,M]} (r^m(W\mathbf{X})) \\ &= \arg\min_{m \in [1,M]} \left\| W\mathbf{X} - \mathbf{A}^m \hat{\mathbf{S}}^m \right\|_F \end{aligned} \tag{22}$$

For details of the SOMP algorithm which is used to solve (22), please refer to [34]. The sparsity level K_0 in this work is also set by experiment empirically.

IV. Data Sets and Experimental Setup

In this section, the experimental data sets as well as the model parameter setting for the proposed classification framework is presented.

Two publicly available and widely used HSI remote sensing data sets are employed here to evaluate the performance of the proposed classification framework: the natural scene Indian Pines collected by the airborne visible/infrared imaging spectrometer (AVIRIS) [45] as well as the urban scene Pavia University data collected by the reflective optics system imaging spectrometer (ROSIS) [46].

The Pavia University data set was collected during a flight campaign over the Pavia district in north Italy, with a spatial resolution of 1.3 m per pixel [49]. There are originally 115 bands with a spectral coverage ranging from 430 to 860 nm. However, 12 channels have been removed due to noise, leaving 103 bands with 610 × 340 pixels per band. Nine classes of interest are provided in the ground reference map, including urban, soil, vegetation features and so on. A small portion of the original data set with a size of 150×150 has been cropped in this work in order to save computing time. One of the image band randomly selected from the cropped Pavia University data set and its corresponding ground reference map containing 8 classes are shown in Fig. 4.

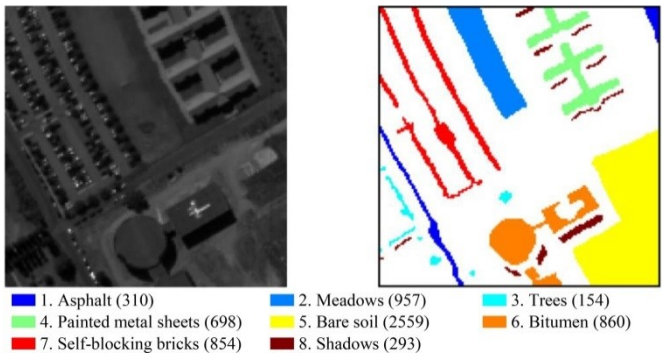

Fig. 4. The cropped Pavia University data set with a size of 150×150: band 70 out of 103 bands (left) and the 8-class ground reference map (right) with number of samples in each class shown at the bottom.

The Indian Pines data set was collected in the Indian Pines test site in north western Indiana, USA in June 12, 1992. The scene consists of two-thirds agriculture, and one-third forest or other natural perennial vegetation [47]. Acquired in band interleaved (BIL) format, it has 145 by 145 pixels with a spatial resolution of 18 m, and 220 continuous spectral channels ranging from 400 to 2500 nm covering the complete VIS-NIR-SWIR spectrum. The nominal spectral resolution is 10 nm and the radiometric resolution is 16 bits [48]. Due to atmospheric water absorption, bands 104-108, 150-163 and 220 do not contain useful information and are consequently removed to prevent from decreasing the classification accuracy, resulting in a reduced data set with 200 spectral bands. There are 16 classes in the ground reference map. One of the image band and the reduced ground reference map of the Indian Pines data set are shown in Fig. 5.

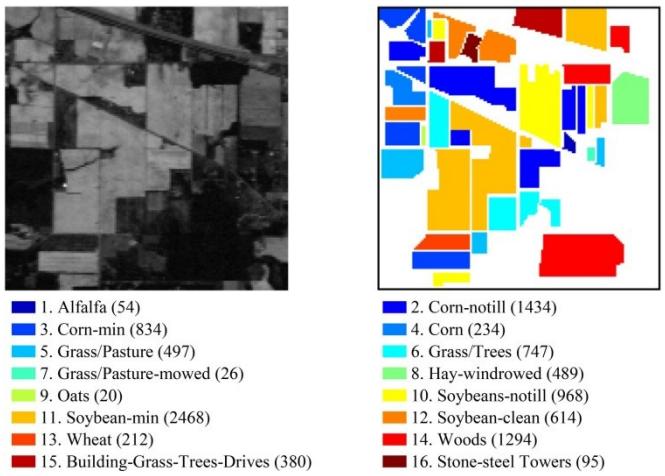

Fig. 5. The Indian Pines data set: band 150 out of 200 bands (left) and the 16-class ground reference map (right) with number of samples in each class shown at the bottom.

Experiments have been repeated ten times using randomly selected training and test samples hoping to obtain a better presentation of results. For Indian Pines, 10% of the labelled samples from each class has been chosen for training the classifier, while a smaller percentage, 5%, has been used for the Pavia University data set. The classification performance is evaluated using the overall accuracy (OA), the average accuracy (AA) as well as the kappa coefficient (κ). OA refers to the percentage of all pixels that are correctly labelled and AA stands for the average percentage of correctly labelled pixels for each class. The kappa coefficient, also named as the kappa statistics, provides a standard for overall classification performance by comparing the agreement against the one that occurs by chance. The value of κ ranges from -1 to 1, where -1 stands for complete disagreement and 1 stands for perfect agreement.

In order to verify the efficacy of the proposed SS-JSRC method, it is firstly compared with the classical classifier SVM with the radial basis function (RBF) kernel [6], SRC [34], JSRC [34] and NLW-JSRC [37]. Optimal parameters for SVM have been achieved using ten-fold cross-validation. It is suggested to choose a larger neighborhood window size for images containing large homogeneous regions in [34]. As can be noticed that there are large homogeneous regions in both data sets, the neighborhood window size is chosen as 9 × 9. Since the Gaussian kernel parameter σ used for the nonlocal weighting scheme was not provided in [37], we choose $\sigma = 2$ for our experiments, while keeping other parameters the same. The sparsity level K_0 for all sparsity representation-based methods are tested for K_0 from 5 to 40. In Fig. 6, the effect of the sparsity level K_0 and the threshold parameter n in the SS-JSRC method is revealed using both the Indian Pines dataset and the Pavia University dataset. As can be seen, the overall classification accuracies are generally high when the sparsity level K_0 is 5 and the threshold parameter n is close to 2. Therefore, the value of n is set to 2 for following experiments.

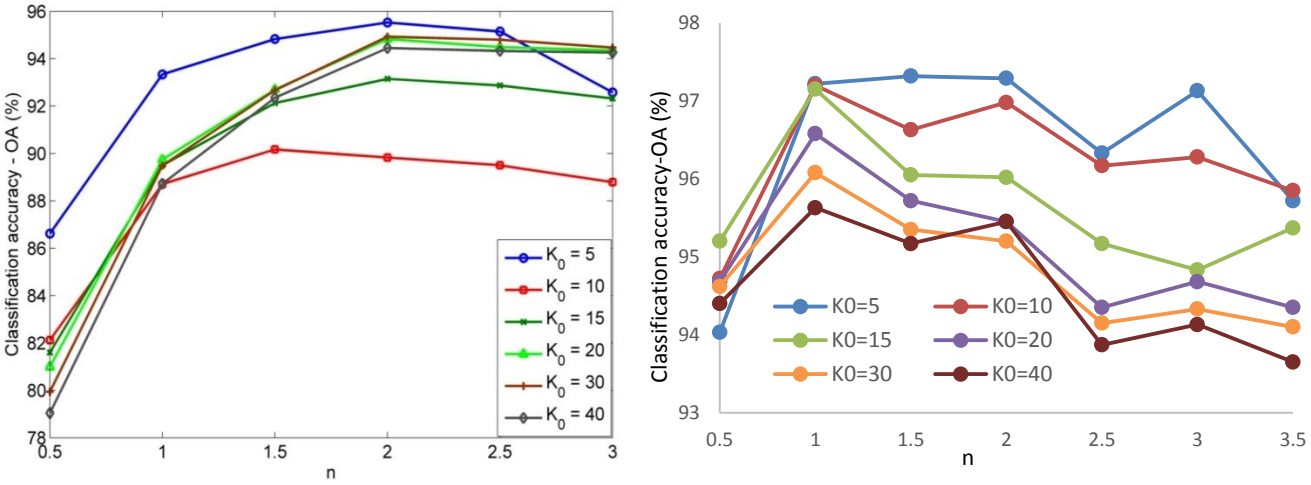

Fig. 6. Effect of the sparsity level K_0 and the threshold parameter n for Indian Pines (left) and Pavia University (right) with the proposed SS-JSRC method where the training ratios are 10% and 1%, respectively.

After the evaluation of classifiers, the proposed classification framework (JSR-SS-JSRC) using the joint bilateral filtering together with the SS-JSRC method is compared with a few state-of-the-art spectral-spatial classification frameworks. For the bilateral filter, the domain parameter σ_d and the range parameter σ_r are set as 4 and 0.1 respectively as discussed in Section III. The first comparison method is SSA-SVM [18], where 2D-SSA is applied on each band for spatial feature extraction, followed by SVM for classification. Another popular method EMAP- SVM [20], where spatial features are achieved using a few attributes constructed on components extracted by PCA and classified by SVM, is included for comparison as well. The third comparison method proposes a spatial-aware dictionary learning (SADL) approach and the learnt sparse coefficients are classified using SVM [50]. Instead of only comparing with the ‘feature extraction and classification’ family, the proposed framework is also compared with the ’classification and post-processing’ family. The method in [38], which utilizes an edge-preserving filter for post-processing the initial SVM-classified result, is adopted and represented as SVM-EPF. In order to show that the joint bilateral filtering is superior to the normal bilateral filtering, features learnt with these two filters are classified using SVM for comparison as well and these two approaches are abbreviated as BF-SVM and JBF-SVM respectively.

V. Results and Discussions

With the two data sets and the experimental setup discussed in Section IV, experiments have been carried out and results are shown in the following subsections.

A. Evaluation of Classifiers

First, results from the SS-JSRC method are compared with those using SVM, SRC, JSRC and NLW-JSRC. The mean value and

the standard deviation with ten experiments for Indian Pines and Pavia University are given in Table I and Table II, together with chosen value for the sparsity level K_0.

TABLE I

MEAN CLASS-BY-CLASS, AVERAGE, OVERALL ACCURACIES (%) AND THE KAPPA COEFFICIENT OF DIFFERENT CLASSIFIERS FOR THE INDIAN PINES DATA SET WITH 10% OF DATA USED FOR TRAINING, FOLLOWED BY THE STANDARD DEVIATION.

Class	Classifiers				
	SVM	SRC	JSRC	NLW-JSRC	SS-JSRC
		K_0=5	K_0=30	K_0=30	K_0=5
1	72.86 ± 17.64	66.33 ± 7.28	83.88 ± 5.65	81.02 ± 6.60	**91.75 ± 4.53**
2	85.78 ± 1.58	64.83 ± 1.98	**94.67 ± 1.40**	91.85 ± 2.88	94.65 ± 1.26
3	79.43 ± 1.95	62.85 ± 3.22	92.30 ± 2.45	85.38 ± 2.80	**92.39 ± 1.55**
4	74.93 ± 11.13	46.02 ± 7.26	**93.27 ± 3.21**	82.80 ± 6.20	83.46 ± 7.04
5	94.36 ± 2.27	90.16 ± 1.84	92.98 ± 2.29	91.48 ± 3.06	**94.59 ± 2.30**
6	97.59 ± 0.89	94.70 ± 1.20	99.17 ± 0.45	**99.51 ± 0.29**	96.91 ± 1.23
7	80.44 ± 11.27	84.78 ± 8.26	52.64 ± 21.54	68.70 ± 11.37	**86.96 ± 7.94**
8	98.34 ± 1.01	97.75 ± 1.32	99.96 ± 0.14	**100.0 ± 0.0**	99.41 ± 0.12
9	71.11 ± 19.91	52.78 ± 11.19	3.33 ± 4.68	44.45 ± 17.76	**77.78 ± 21.60**
10	83.37 ± 3.68	72.09 ± 2.59	86.83 ± 1.25	83.82 ± 4.04	**92.04 ± 2.06**
11	88.64 ± 1.08	76.92 ± 1.49	**97.48 ± 0.76**	96.83 ± 0.88	96.38 ± 1.16
12	84.27 ± 3.36	57.65 ± 2.69	86.62 ± 3.38	88.25 ± 3.06	**93.16 ± 3.65**
13	99.01 ± 0.87	97.91 ± 1.31	99.16 ± 0.62	99.32 ± 0.82	95.33 ± 3.40
14	95.41 ± 1.30	93.75 ± 1.79	99.30 ± 0.36	**99.61 ± 0.35**	99.44 ± 0.50
15	61.52 ± 3.84	44.91 ± 3.30	**97.69 ± 1.17**	94.36 ± 2.17	93.10 ± 3.77
16	90.12 ± 2.16	90.35 ± 4.92	94.94 ± 2.01	**96.82 ± 2.15**	96.47 ± 2.22
OA	87.76 ± 0.19	76.12 ± 0.49	94.85 ± 0.48	93.32 ± 0.64	**95.19 ± 0.44**
AA	84.82 ± 1.76	74.61 ± 0.99	85.89 ± 1.50	87.76 ± 1.36	**92.74 ± 1.37**
K	0.860 ± 0.002	0.727 ± 0.006	0.941 ± 0.005	0.924 ± 0.007	**0.945 ± 0.005**

TABLE II

MEAN CLASS-BY-CLASS, AVERAGE, OVERALL ACCURACIES (%) AND THE KAPPA COEFFICIENT OF DIFFERENT CLASSIFIERS FOR THE PAVIA UNIVERSITY DATA SET WITH 5% OF DATA USED FOR TRAINING, FOLLOWED BY THE STANDARD DEVIATION.

Class	Classifiers				
	SVM	SRC	JSRC	NLW-JSRC	SS-JSRC
		K_0=5	K_0=30	K_0=30	K_0=5
1	65.92 ± 4.68	48.20 ± 4.72	77.35 ± 11.32	64.56 ± 4.78	**80.31 ± 8.20**
2	96.16 ± 1.40	94.04 ± 1.80	**99.76 ± 0.34**	99.55 ± 0.42	98.29 ± 1.03
3	94.66 ± 4.76	97.26 ± 1.99	90.68 ± 6.41	**98.50 ± 0.78**	94.32 ± 2.31
4	99.44 ± 0.51	99.73 ± 0.18	**99.98 ± 0.05**	99.88 ± 0.15	99.73 ± 0.09
5	96.36 ± 1.46	91.51 ± 1.53	**99.54 ± 0.73**	97.68 ± 0.66	98.56 ± 0.45
6	91.71 ± 1.73	86.11 ± 1.39	97.53 ± 1.69	96.51 ± 1.08	**98.63 ± 0.69**
7	90.80 ± 1.95	81.53 ± 3.53	98.75 ± 1.27	96.67 ± 1.25	**98.98 ± 1.05**
8	**96.33 ± 1.82**	93.42 ± 2.27	71.19 ± 5.86	83.13 ± 3.59	91.84 ± 1.68
OA	93.90 ± 0.95	88.97 ± 0.47	96.79 ± 0.66	95.74 ± 0.32	**97.47 ± 0.51**
AA	91.42 ± 1.10	86.47 ± 0.81	91.85 ± 1.63	92.06 ± 0.82	**95.08 ± 1.23**
K	0.922 ± 0.012	0.859 ± 0.006	0.959 ± 0.009	0.946 ± 0.004	**0.968 ± 0.007**

It can be seen that in most cases, classifiers containing contextual information (JSRC, NLW-JSRC and SS-JSRC) consistently outperform pixel-wise classifiers (SVM and SRC), which demonstrates the robustness of the joint sparsity model for improving the classification performance. Although it is claimed that NLW-JSRC helps to improve the classification accuracies over JSRC [37], it is impossible to achieve the optimal results without the 'appropriate' parameterization for the Gaussian kernel, not to mention the

heavy computational complexity when calculating the weight parameters. In comparison with the proposed SS-JSRC method, the overall accuracy has increased from 94.58% to 95.19% for Indian Pines and 96.79% to 97.47% for Pavia University, which seems not a big improvement. However, it is noted that the improvement of the average accuracy in the proposed method is more noticeable since the results of those classes with limited training samples have been enhanced due to the simple pixel removal process in the SS-JSRC method. For example, class 9 in the Indian Pines data set is a narrow strip as shown in the ground reference map. With the JSRC method, pixels from both class 3 and class 6 are included in the 9 × 9 window, while irrelevant pixels can be successfully removed using the SS-JSRC method.

TABLE III

MEAN CLASS-BY-CLASS, AVERAGE, OVERALL ACCURACIES (%) AND THE KAPPA COEFFICIENT OF DIFFERENT CLASSIFICATION FRAMEWORKS FOR THE INDIAN PINES DATA SET WITH 10% OF DATA USED FOR TRAINING, FOLLOWED BY THE STANDARD DEVIATION.

Class	**Classifiers**						
	SSA-SVM	**EMAP-SVM**	**SADL-SVM**	**SVM-EPF**	**BF-SVM**	**JBF-SVM**	**JBF-SS-JSRC** $K_0 = 30$
1	78.96 ± 12.86	87.92 ± 4.79	88.33 ± 8.29	**100.0 ± 0.0**	86.04 ± 6.22	87.08 ± 8.61	98.54 ± 1.72
2	96.45 ± 0.82	90.30 ± 1.27	95.12 ± 1.23	97.40 ± 1.20	94.66 ± 0.78	**97.67 ± 0.65**	97.40 ± 0.71
3	93.91 ± 2.04	97.08 ± 1.35	96.33 ± 2.27	98.35 ± 0.57	93.99 ± 1.22	**98.55 ± 0.62**	97.72 ± 1.11
4	89.09 ± 2.07	92.14 ± 4.86	89.00 ± 4.08	92.99 ± 5.61	88.29 ± 2.32	95.24 ± 2.40	**95.67 ± 3.55**
5	93.74 ± 1.33	94.68 ± 2.00	97.18 ± 1.90	**99.40 ± 0.43**	96.40 ± 2.76	96.71 ± 2.31	97.52 ± 1.92
6	98.66 ± 0.79	98.82 ± 0.85	**99.09 ± 1.00**	98.62 ± 1.38	98.23 ± 0.77	97.75 ± 0.71	98.02 ± 0.40
7	72.17 ± 23.83	95.65 ± 2.90	**100.0 ± 0.0**	95.75 ± 13.44	96.09 ± 1.37	94.78 ± 3.43	95.22 ± 3.21
8	98.32 ± 0.99	99.46 ± 0.29	**100.0 ± 0.0**	99.34 ± 2.08	99.07 ± 0.23	99.61 ± 0.34	99.93 ± 0.15
9	77.78 ± 23.13	76.67 ± 15.00	**95.00 ± 10.62**	80.00 ± 42.16	75.00 ± 18.75	88.89 ± 15.93	88.89 ± 9.44
10	92.77 ± 1.91	89.69 ± 2.38	**97.16 ± 1.57**	93.39 ± 4.56	93.10 ± 1.81	95.68 ± 1.83	96.14 ± 1.14
11	95.41 ± 1.45	97.73 ± 0.51	**98.70 ± 0.93**	96.04 ± 1.15	94.38 ± 1.20	97.45 ± 1.03	98.65 ± 0.73
12	92.84 ± 3.61	89.97 ± 4.06	93.19 ± 3.05	**98.76 ± 0.53**	90.45 ± 2.21	95.38 ± 1.87	96.29 ± 1.85
13	98.58 ± 1.88	99.42 ± 0.16	98.74 ± 0.67	**100.0 ± 0.0**	98.21 ± 0.71	98.53 ± 1.05	99.11 ± 0.79
14	98.36 ± 0.62	99.32 ± 0.34	99.30 ± 0.79	96.65 ± 2.17	97.65 ± 1.24	98.93 ± 0.50	**99.93 ± 0.05**
15	91.35 ± 5.55	96.93 ± 1.55	97.95 ± 2.23	97.43 ± 3.04	92.34 ± 5.03	94.68 ± 2.71	**98.54 ± 0.77**
16	94.00 ± 6.71	89.77 ± 9.40	96.47 ± 4.37	**98.93 ± 0.38**	92.12 ± 6.86	95.76 ± 2.30	97.65 ± 0.55
OA	95.28 ± 0.64	95.36 ± 0.36	97.31 ± 0.40	96.93 ± 0.64	94.81 ± 0.39	97.29 ± 0.35	**98.05 ± 0.24**
AA	91.40 ± 2.20	93.47 ± 1.62	96.35 ± 0.84	96.44 ± 3.32	92.67 ± 1.85	95.79 ± 1.15	**97.20 ± 0.55**
κ	0.946 ± 0.007	0.947 ± 0.004	0.969 ± 0.005	0.9650 ± 0.007	0.941 ± 0.004	0.970 ± 0.004	**0.978 ± 0.003**

B. Comparison with State-Of-The-Art Spectral-Spatial Classification Frameworks

After the comparison of different classifiers, the proposed classification framework (JBF-SS-JSRC) is further compared with several other state-of-the-art classification frameworks under the same experimental setting used previously. The classification maps of various methods for Indian Pines and Pavia University are shown in Figure 7 and Figure 8. Detailed numerical results over ten repetitions are shown in Table III and Table IV, respectively, from which four main findings can be identified as follows.

1) By comparing results between BF-SVM and JBF-SVM, it is found that JBF with the first principal component as the guidance image yields better spatial feature extraction performance than conventional bilateral filtering, due mainly to the enhanced edge property in the proposed method. By taking the Indian Pines data set as shown in Figure 7(f) and 7(g) as an example, there are more misclassified pixels around edges with the normal bilateral filtering.

2) Better classification results have been achieved with the proposed JBF feature extraction techniques than others such as SSA

and SADL for the same classifier. It is also noted that classification results by EMAP occasionally are seen slightly better than those obtained by the proposed JBF method. However, the computation cost in EMAP is extremely high (see in Table V).

3) The proposed classification framework JBF-SS-JSRC outperforms the one using JBF-SVM, which proves the finding in Section V-A again that SS-JSRC is superior to the pixel-wise classifier using SVM.

4) For both data sets, the performance of the proposed JBF-SS-JSRC framework is slightly better than other state-of-the-art classification frameworks. For instance, as observed from Table III, the overall accuracy of JBF-SS-JSRC is 2.77%, 2.69%, 0.74% and 1.12% higher than that of SSA-SVM, EMAP-SVM, SADL-SVM and SVM-EPF, respectively.

TABLE IV

MEAN CLASS-BY-CLASS, AVERAGE, OVERALL ACCURACIES (%) AND THE KAPPA COEFFICIENT OF DIFFERENT CLASSIFICATION FRAMEWORKS FOR THE PAVIA UNIVERSITY DATA SET WITH 5% OF DATA USED FOR TRAINING, FOLLOWED BY THE STANDARD DEVIATION.

Class	**Classifiers**						
	SSA-SVM	**EMAP-SVM**	**SADL-SVM**	**SVM-EPF**	**BF-SVM**	**JBF-SVM**	**JBF-SS-JSRC** $K_0 = 5$
1	77.95 ± 4.92	**99.73 ± 0.50**	94.52 ± 3.37	99.45 ± 1.05	98.44 ± 1.88	97.18 ± 2.83	99.63 ± 0.95
2	91.57 ± 1.91	99.83 ± 0.08	98.84 ± 0.92	99.64 ± 0.96	99.83 ± 0.19	99.72 ± 0.33	**99.85 ± 0.49**
3	97.08 ± 2.46	89.66 ± 4.17	97.53 ± 3.53	**98.46 ± 2.80**	95.48 ± 2.61	98.36 ± 1.17	97.95 ± 2.35
4	99.78 ± 0.29	99.41 ± 0.29	99.74 ± 0.19	**100.0 ± 0.0**	99.47 ± 0.61	99.37 ± 0.39	99.91 ± 0.15
5	96.62 ± 1.09	**99.95 ± 0.06**	99.23 ± 0.78	99.94 ± 0.12	99.64 ± 0.36	99.60 ± 0.23	**99.93 ± 0.03**
6	94.85 ± 1.26	**99.93 ± 0.13**	97.77 ± 2.44	99.37 ± 1.27	98.00 ± 2.49	97.80 ± 2.09	99.89 ± 0.35
7	85.60 ± 2.59	99.06 ± 0.52	98.37 ± 0.55	96.38 ± 3.16	98.09 ± 1.40	98.22 ± 0.75	**99.62 ± 0.21**
8	98.70 ± 1.16	**100.0 ± 0.0**	99.93 ± 0.15	99.89 ± 0.17	95.76 ± 3.77	97.19 ± 2.55	96.58 ± 2.74
OA	93.79 ± 0.73	99.52 ± 0.13	98.70 ± 0.45	99.28 ± 0.60	98.92 ± 0.37	98.94 ± 0.44	**99.67 ± 0.06**
AA	92.77 ± 0.69	98.45 ± 0.56	98.24 ± 0.68	99.14 ± 0.64	98.09 ± 0.53	98.43 ± 0.69	**99.19 ± 0.17**
κ	0.921 ± 0.009	0.994 ± 0.002	0.984 ± 0.006	0.991 ± 0.008	0.986 ± 0.005	0.987 ± 0.006	**0.996 ± 0.001**

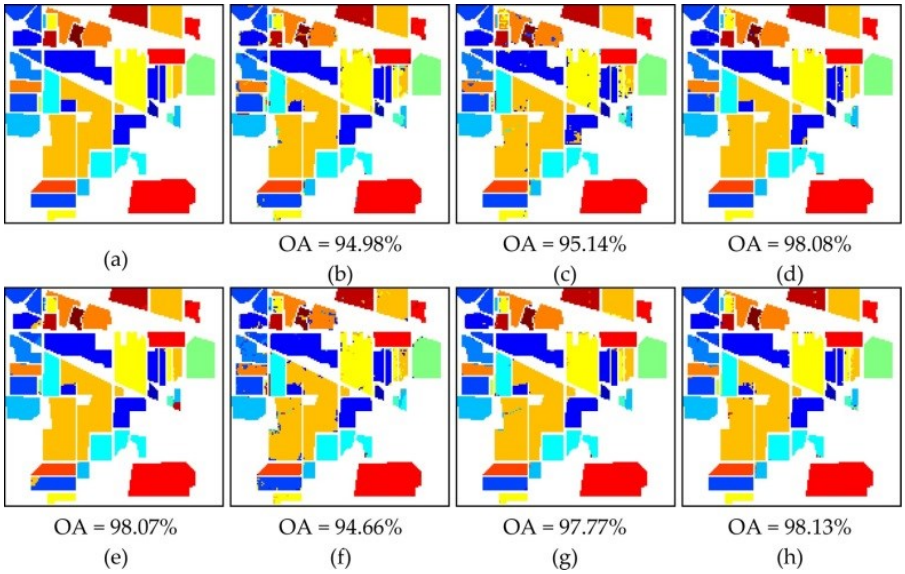

Fig. 7. The Indian Pines data set (a) ground reference and classification maps together with corresponding overall accuracy obtained by (b) SSA-SVM, (c) EMAP-SVM, (d) SADL-SVM, (e) SVM-EPF, (f) BF-SVM, (g) JBF-SVM and (h) JBF-SS-JSRC.

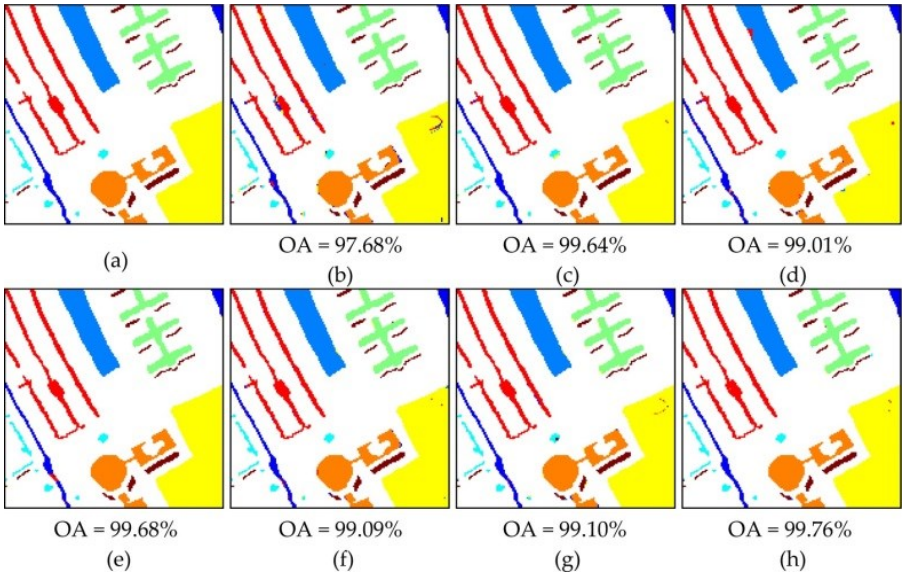

Fig. 8. The Pavia University data set (a) ground reference and classification maps together with corresponding overall accuracy obtained by (b) SSA-SVM, (c) EMAP-SVM, (d) SADL-SVM, (e) SVM-EPF, (f) BF-SVM, (g) JBF-SVM and (h) JBF-SS-JSRC.

The computing time of these feature extraction techniques is recorded on a personal computer with an Intel Core i5-2400 CPU at 3.10 GHz using Matlab 2014a (Mathworks) and results are presented in Table V. In addition, the computational complexity of the five algorithms in Table V are further listed in Table VI for comparison. Note the computational complexity of JBF is derived from fast implementation of PCA and BF, whose computational complexity can both be reduced to O(SB) [54-56]. It is apparent JBF is the most efficient one in the group. All in all, the obtained results confirm the strong superiority of joint bilateral filtering and SS-JSRC over other popular spectral-spatial classification frameworks.

TABLE V

COMPUTING TIME (IN SECONDS) FOR DIFFERENT FEATURE EXTRACTION TECHNIQUES.

Datasets	**Feature descriptor**				
	SSA	**EMAP**	**SADL**	**BF**	**JBF**
Indian Pines	22	637	77	66	4
Pavia University	12	423	9	36	3

TABLE VI

COMPUTING COMPLEXITY FOR DIFFERENT FEATURE EXTRACTION TECHNIQUES.

	Computational complexity	**Parameters**
SSA	$O(B(B-L)L^2)$	B is number of bands, L is window size for embedding
EMAP	O(S(G+4L+2LS))	S, G and L refer respectively to the numbers of pixels, graylevels and filtering levels
SADL	$O(k^2S+k^3)$	S is the number of pixels and k is the number of variables for regression analysis
BF	$O(BS^2)$	S is the number of pixels
JBF	O(SB)	Derived from fast implementation of BF and PCA, S is the number of pixels

As deep learning has been proposed as another important approach for feature extraction in hyperspectral imaging [57-58], we compare the results from our approach with those from deep learning techniques as follows. For Indiana Pines dataset, the OA for 16 classes with 10% training ratio is 95.97% when convolutional neural network (CNN) is combined with SVM [58]. This can be increased to 97.62% when SVM is replaced by the Extreme Learning Machine (ELM) in comparison to our result at 98.05%. In addition, CNN with SP has achieved an OA of 97.45% [59] or 97.56% with 3D CNN [60]. For Pavia University dataset in eight classes, the OA from our proposed approach under a training ratio of 5% is 99.67%, where the best results from [60] and [61] are 99.66% and 99.50%, respectively. As can be seen, the proposed approach has yielded comparable or even slightly better results than almost all these deep-learning based approaches, whilst the framework is much simpler than those using CNN as their performance can be severely affected by the chosen layers and architectures of the deep learning networks.

VI. Conclusions

In this paper, efforts have been focused on developing effective classification frameworks for hyperspectral images. The task is divided into two parts, which are extracting effective features and constructing suitable classifiers. To this end, joint bilateral filtering with the first principal component as the guidance image, is proposed to extract spatial features without blurring the edge information, where its superiority has been validated by comparing it against the normal bilateral filtering method. On the other hand, an alternative classifier SS-JSRC, which is an enhanced version of the JSRC method, is proposed for improving classification performances. With the SS-JSRC method, a simple yet powerful pixel removal process, based on the statistics within the neighboring window, is added before applying the joint sparsity model. Results have shown that both the classification accuracy and the kappa coefficient are seen to improve over those by the JSRC and NLW-JSRC methods, especially for classes located at small spatial areas. By combining joint bilateral filtering and the SS-JSRC method together, the classification performance of the proposed framework is shown to outperform several typical spectral-spatial classification approaches for both Indian Pines and Pavia University data sets. Future work will investigate the possibility of hyperspectral image classification using histograms of sparse codes, which has been successfully applied for object detection [51], where classification with limited training samples will be particularly focused [62]. In addition, saliency based segmentation and classification will also be investigated, using frameworks such as weakly supervised learning, deep learning and multi-instance learning [63-68].

Acknowledgment

Authors would thank Prof. D. Landgrebe and Prof. P. Gamba for providing the Indiana Pines and Pavia University datasets respectively online used in our experiments. The authors would also like to thank for the support from the National Nature Science Foundation of China (61471132, 61672008), the Training Program for Outstanding Young Teachers in Higher Education Institutions of Guangdong Province (YQ2015057), Guangdong Provincial Application-oriented Technical Research and

Development Special fund project (2016B010127006, 2015B010131017) and International Scientific and Technological Cooperation Projects of Education Department of Guangdong Province (2015KGJHZ021).

REFERENCES

[1]. C. Zhao, X. Li, J. Ren, and S. Marshall, "Improved sparse representation using adaptive spatial support for effective target detection in hyperspectral imagery," *Int. J. Remote Sens.*, vol. 34, no. 24, pp. 8669–8684, Oct. 2013.

[2]. D.-W. Sun, *Hyperspectral Imaging for Food Quality Analysis and Control*. Elsevier, 2010.

[3]. T. Kelman, J. Ren, and S. Marshall, "Effective classification of Chinese tea samples in hyperspectral imaging," *Artif. Intell. Res.*, vol. 2, no. 4, pp. 87–96, 2013.

[4]. Y. Roggo, A. Edmond, P. Chalus, and M. Ulmschneider, "Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms," *Anal. Chim. Acta*, vol. 535, no. 1-2, pp. 79–87, Apr. 2005.

[5]. K. Gill, J. Ren, S. Marshall, S. Karthick, and J. Gilchrist, "Quality-assured fingerprint image enhancement and extraction using hyperspectral imaging," in *Proc. 4th Int. Conf. on Image for Crime Detection and Prevention*, London, UK, 2011.

[6]. F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. and Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[7]. Z. He, Q. Wang, Y. Shen, and M. Sun, "Kernel sparse multitask learning for hyperspectral image classification with empirical mode decomposition and morphological wavelet-based features," *IEEE Trans. Geosci. and Remote Sens.*, vol. 52, no. 8, pp. 5150–5163, Aug. 2014.

[8]. J. Zabalza, J. Ren, J. Ren, Z. Liu, and S. Marshall, "Structured covariance principal component analysis for real-time onsite feature extraction and dimensionality reduction in hyperspectral imaging," *Appl. Optics*, vol. 53, no. 20, pp. 4440–4449, Jul. 2014.

[9]. J. Ren, J. Zabalza, S. Marshall, and J. Zheng, "Effective feature extraction and data reduction in remote sensing using hyperspectral imaging," *IEEE Signal Proc. Mag.*, vol. 31, no. 4, pp. 149–154, Jul. 2014.

[10]. J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, and J. Han, "Novel folded-pca for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, no. 7, pp. 112–122, Jul. 2014.

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

[12]. U. Amato, R. M. Cavalli, A. Palombo, S. Pignatti, and F. Santini, "Experimental approach to the selection of the components in the minimum noise fraction," *IEEE Trans. Geosci. and Remote Sens.*, vol. 47, no. 1, pp. 153–160, Jan. 2009.

[13]. X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[14]. J. Zabalza, J. Ren, W. Zheng, S. Marshall, and J. Wang, "Singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1886–1890, Nov. 2014.

[15]. J. Zabalza, J. Ren, Z. Wang, H. Zhao, J. Wang, and S. Marshall, "Fast implementation of singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE J. Sel. Topics Appl. Earth Observ. and Remote Sens.*, vol. 8, no. 6, pp. 2845–2853, Jun.

[16]. B. Demir and S. Ertürk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Trans. Geosci. and Remote Sens.*, vol. 48, no. 11, pp. 4071–4084, Nov. 2010.

[17]. B. Demir, S. Ertürk, and M. K. Güllü, "Hyperspectral image classification using denoising of intrinsic mode functions," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 220–224, Mar. 2011.

[18]. J. Zabalza, J. Ren, J. Zheng, J. Han, H. Zhao, S. Li, and S. Marshall, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 8, pp. 4418–4433, Aug. 2015.

[19]. M. Fauvel, J. A. Benediktsson, J. chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Trans. Geosci. and Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[20]. M. D. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.

[21]. J. A. Richards and X. Jia, "Using suitable neighbors to augment the training set in hyperspectral maximum likelihood classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 774–777, Oct. 2008.

[22]. S. Kuching, "The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis," *J. Comput. Sci.*, vol. 3, no. 6, pp. 419–423, 2007.

[23]. N. Bali and A. Mohammad-Djafari, "Bayesian approach with hidden Markov modeling and mea field approximation for hyperspectral data analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 217–255, Feb. 2008.

[24]. Q. Jackson, D. Landgrebe *et al.*, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. and Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.

[25]. M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sens. Environ.*, vol. 86, no. 4, pp. 554–565, Aug. 2003.

[26]. C.-C. Yang, S. O. Prasher, P. Enright, C. Madramootoo, M. Burgess, P. K. Goel, and I. Callum, "Application of decision tree technology for image classification using remote sensing data," *Agric. Syst.*, vol. 76, no. 3, pp. 1101–1117, Jun. 2003.

[27]. D. L. Civco, "Artificial neural networks for land-cover classification and mapping," *Int. J. Geogr. Inf. Sci.*, vol. 7, no. 2, pp. 173–186.

[28]. H. Bischof and A. Leonardis, "Finding optimal neural networks for land use classification," *IEEE Trans.Geosci. and Remote Sens.*, vol. 36, no. 1, pp. 337–341, Jan. 1998.

[29]. G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans.Geosci. and Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

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

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

[32]. E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[33]. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[34]. Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. and Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[35]. J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part i: Greedy pursuit," *Signal Process.–Special Issue on Sparse Approximations in Signal and Image Processing*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[36]. S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[37]. H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. and Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.

[38]. X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge- preserving filtering," *IEEE Trans. Geosci. and Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.

[39]. K. Schindler, "An overview and comparison of smooth labeling methods for land-cover classification," *IEEE Trans. Geosci. and Remote Sens.*, vol. 50, no. 11, pp. 4534–4545, Nov. 2012.

[40]. K. Kotwal and S. Chaudhuri, "Visualization of hyperspectral images using bilateral filtering," *IEEE Trans.Geosci. and Remote Sens.*, vol. 48, no. 5, pp. 2308–2316, May 2010.

[41]. H. Peng and R. Rao, "Hyperspectral image enhancement with vector bilateral filtering," in *Proc. 16th IEEE Int.l Conf. on Image Processing*. IEEE, 2009, pp. 3713–3716.

[42]. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. on Computer Vision*. IEEE, 1998, pp. 839–846.

[43]. J. Tropp, A. C. Gilbert *et al.*, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[44]. W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[45]. AVIRIS image Indian Pines. [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec

[46]. ROSIS image Pavia University. [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral Remote Sensing Scenes

[47]. G. Camps-Valls and L. Bruzzone, Eds., *Kernel Methods for Remote Sensing Data Analysis*. Chichester, UK: Wiley, 2009.

[48]. A. Chakrabarty, O. Choudhury, P. Sarkar, A. Paul, and D. Sarkar, "Hyperspectral image classification incorporating bacterial foraging-optimized spectral weighting," *Artif. Intell. Res.*, vol. 1, no. 1, pp. 63–83, 2012.

[49]. J. A. Benediktsson, J. A. Palmason and J. R. Sveinsson, "Classification of hyperspectral data from urban Areas based on extended morphological profiles," *IEEE Trans. Geosci. and Remote Sens.*, vol. 43, no. 3, pp. 480–491, March 2005.

[50]. A. Soltani-Farani, H. R. Rabiee, and S. A. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 1, pp. 527–541, Jan. 2015.

[51]. X. Ren and D. Ramanan, "Histograms of sparse codes for object detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2013, pp. 3246–3253.

[52]. W. Fu et al, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J. Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 9, no. 2, pp. 556-567, Feb. 2016.

[53]. Y. Chen, et al, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. and Remote Sens.*, vol. 54, no. 10, pp. 6232-6251, Oct. 2016.

[54]. S. Roweis, "EM algorithms for PCA and SPCA," in *Proc. NIPS*, pp. 626-632, 1997

[55]. S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," *International Journal of Computer Vision*, vol. 81, no. 1, pp. 24-52, 2009

[56]. Q. Yang, K.-H. Tan and N. Ahuja, "Real-time *O*(1) bilateral filtering," in *Proc. CVPR*, 2009

[57]. J. Zabalza, J. Ren et al, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing*, vol. 185, pp. 1-10, April 2016.

[58]. Y. Li, W. Xie and H. Li, "Hyperspectral image reconstruction by deep convolutional neural network for classification," *Pattern Recognition*, vol. 63, pp. 371-383, 2017

[59]. H. Liang and Q. Li, "Hyperspectral imagery classification using sparse representations of convolutional neural network features," *Remote Sensing*, vol. 8, no. 2, 2016.

[60]. Y. Chen, H. Jiang, C. Lim X. Jia and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. and Remote Sens.*, vol. 54, no. 10, pp. 6232-6251, Oct. 2016.

[61]. R. Kemker and C. Kanan, "Self-taught feature learning for hyperspectral image classification," *IEEE Trans. Geosci. and Remote Sens.*, vol. 55, no. 5, pp. 2693-2705, May 2017.

[62]. C.-Y. Zhao, X.-B. Gao, et al, "Multiple feature learning based hyperspectral image classification with limited training samples," *IEEE Trans. Geosci. and Remote Sens*, Vol.54, No.7, pp.4052-4062, 2016.

[63]. D. Zhang, J. Han, C. Li, J. Wang, and X. Li, "Detection of co-salient objects by looking deep and wide," *International Journal of Computer Vision*, vol. 120. no. 2, pp. 215-232, 2016.

[64]. D. Zhang, J. Han, J. Han, L. Shao, "Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1163-1176, 2016.

[65]. X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, "Semantic annotation of high-resolution satellite images via weakly supervised learning," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3660-3671, 2016.

[66]. D. Zhang, J. Han, L. Jiang, S. Ye, X. Chang, "Revealing event saliency in unconstrained video collection," *IEEE Trans. on Image Processing*, vol. 26, no. 4, pp. 1746-1758, 2017.

[67]. X. Yao, J. Han, D. Zhang and F. Nie, "Revisiting co-saliency detection: a novel approach based on two-stage multi-view spectral rotation co-clustering," *IEEE Trans. on Image Processing*, vol. 26, no. 7, pp. 3196-3209, 2017.

[68]. D. Zhang, D. Meng, J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 865-878, 2017.

https://dspace.org/

Joint bilateral filtering and spectral similarity-based sparse representation: A generic framework for effective feature extraction and data classification in hyperspectral imaging

Qiao, Tong; Yang, Zhijing; Ren, Jinchang; Yuen, Peter W. T.; Zhao, Huimin ...

2017-10-10
Elsevier
https://doi.org/10.1016/j.patcog.2017.10.008
Qiao T, Yang Z, Ren J, et al., Joint bilateral filtering and spectral similarity-based sparse representation: A generic framework for effective feature extraction and data classification in hyperspectral imaging. Pattern Recognition, Volume 77, May 2018, pp. 316-328

Downloaded from a DSpace repository. https://dspace.org/